

Reakcja platform społecznościowych na zgłoszenia w 2025 roku



Ośrodek Analizy
Dezinformacji
NASK

Reakcja platform społecznościowych na zgłoszenia w 2025 roku

ZESPÓŁ SZYBKIEGO REAGOWANIA I WYKRYWANIA DEZINFORMACJI

Agnieszka Kaczmarczyk

Bartosz Popadiak

Ewelina Włodarczyk

Katarzyna Lipka

Martyna Rataj

Mikołaj Bogusz



Ośrodek Analizy
Dezinformacji
NASK



Ministerstwo
Cyfryzacji



PROJEKT FINANSOWANY ZE ŚRODKÓW
MINISTERSTWA CYFRYZACJI

Tytuł

Reakcja platform społecznościowych na zgłoszenia w 2025 roku

Autorzy

Agnieszka Kaczmarczyk, Bartosz Popadiak, Ewelina Włodarczyk,
Katarzyna Lipka, Martyna Rataj, Mikołaj Bogusz

Redakcja i opieka merytoryczna

Izabela Jarka

Korekta

Łukasz Szczęsny

Opracowanie graficzne

NASK – PIB

Publikacja jest rozpowszechniana na zasadach licencji Creative Commons.
Uznanie autorstwa – Użycie niekomercyjne (CC BY-NC) 4.0 Międzynarodowe.

ISBN: 978-83-68356-48-9

2026

NASK – Państwowy Instytut Badawczy
ul. Kolska 12, 01-045 Warszawa
www.nask.pl

dezinformacja@nask.pl

Publikacja wyraża jedynie poglądy autorów i nie może być utożsamiana
z oficjalnym stanowiskiem Ministra Cyfryzacji

Spis treści

Wstęp	11
METODOLOGIA	12
Metodologia obsługi zgłoszeń	
Ośrodek Analizy Dezinformacji NASK	13
Platformy Meta – Facebook, Instagram i Threads	15
X.com	15
TikTok	18
Google (YouTube i Google Ads)	18
LinkedIn	18
Discord	18
Telegram	19
Bluesky	21
PODSUMOWANIE 2025	22
Kluczowe wnioski	23
Zasadne zgłoszenia przekazane do platform społecznościowych	24
Zgłoszenia zewnętrzne	26
Zgłoszenia wewnętrzne	28
Podsumowanie 2025 w podziale na platformy społecznościowe	30

I KWARTAŁ 2025	32
Reakcja platform na wszystkie zgłoszenia OAD NASK	36
Meta	39
Facebook	39
Instagram	40
Threads	40
X.com	41
TikTok	41
Google (Youtube i Google Ads)	42
LinkedIn	42
Discrod	42
Telegram	42
Zgłoszenia wewnętrzne	43
Meta	44
Facebook	45
Instagram	45
X.com	46
TikTok	46
Google (YouTube i Google Ads)	47
Discrod	47
Zgłoszenia zewnętrzne	48
Meta	50
Facebook	50
Instagram	50
Threads	51
X.com	51
TikTok	51
Google (YouTube i Google Ads)	52
LinkedIn	52
Discrod	52
Telegram	52

Case studies	53
Zgłoszenia dotyczące kampanii wyborczej 2025	53
Przykłady zgłoszonej dezinformacji wyborczej	54
Reklamy z wizerunkiem osób publicznych	58
Szkodliwe treści medyczne	60
Oferta sprzedaży fałszywej aplikacji mObywatel	65
Kanały aktywnie publikujące treści dezinformacyjne	67
Podszywanie pod osobę publiczną / instytucję	70

II KWARTAŁ 2025 **73**

Reakcja platform na wszystkie zgłoszenia OAD NASK	77
Meta	80
Facebook	81
Instagram	81
X.com	82
TikTok	82
Google (YouTube i Google Ads)	83
LinkedIn	83
Telegram	83
Bluesky	83
Pozostałe platformy	84
Zgłoszenia wewnętrzne	85
Meta	86
Facebook	87
Instagram	87
Threads	87
X.com	88
TikTok	88
Google (YouTube i Google ds)	89
Telegram	89
Bluesky	89
Pozostałe platformy	89

Zgłoszenia zewnętrzne	90
Meta	91
Facebook	92
Instagram	92
Threads	93
X.com	93
TikTok	94
Google	94
LinkedIn	95
Telegram	95
Bluesky	95
Pozostałe platformy	95
Case studies	96
Dezinformacja na temat zabójstwa na terenie kampusu UW	96
Dezinformacja na temat ekshumacji ofiar rzezi wołyńskiej	99
Dezinformacja klimatyczna	102
Dezinformacja medyczna	103
III KWARTAŁ 2025	114
Reakcja platform na wszystkie zgłoszenia OAD NASK	118
Meta	121
Facebook	121
Instagram	122
X.com	123
TikTok	123
Google (YouTube i Google Ads)	124
Telegram	124
Discrod	124
Pozostałe platformy	125

Zgłoszenia wewnętrzne	126
Meta	127
Facebook	128
Instagram	128
X.com	129
TikTok	129
Google (Youtube i Google Ads)	130
Pozostałe platformy	130
Zgłoszenia zewnętrzne	131
Meta	132
Facebook	133
Instagram	133
X.com	134
TikTok	134
Google (YouTube i Google Ads)	135
Telegram	135
Discrod	135
Pozostałe platformy	136
Case studies	137
Dezinformacja na temat incydentu z dronami	137
Działalność konta „Polak na Donbasie” na platformie TikTok	144
Treści powiązane z internetowym trendem „szon patrol”	145
Kanały w serwisie youtube publikujące treści polityczne wygenerowane przez AI	146
Kampania propagandowa izraelskiego rządu („Lapam”)	147
Nagranie deepfake z rosyjskim reporterem Nikołajem Dołgaczowem	157
Reklamy na platformie facebook podszywające się pod ministerstwa	158
Reklamy w serwisie facebook wykorzystujące wizerunki znanych lekarzy	160
Operacje Overload i Undercut	166

IV KWARTAŁ 2025	168
Reakcja platform na wszystkie zgłoszenia OAD NASK	171
Meta	174
Facebook	174
Instagram	175
Threads	175
X.com	176
TikTok	176
Google (YouTube i Google Ads)	177
Telegram	177
Pozostałe platformy	178
Zgłoszenia wewnętrzne	179
Meta	180
Facebook	181
Instagram	181
X.com	181
TikTok	182
Google (YouTube i Google Ads)	182
Pozostałe platformy	183
Zgłoszenia zewnętrzne	184
Meta	185
Facebook	186
Instagram	186
Threads	187
X.com	187
TikTok	188
Google (Youtube i Google Ads)	188
Telegram	189
Pozostałe platformy	189

Case studies	190
Dezinformacja na temat aktu dywersji na tory kolejowe	190
Dezinformacja na temat bezpieczeństwa szczepionek	196
Pacjent zmarł, bo lekarz nie znał języka polskiego	198
Reklamy na platformie youtube z wykorzystaniem deepfake'a z prezydentem karolem nawrockim	201
Zabójstwo 11-latki w jeleniej górze	203
Konto wieczna_krucjata (prawilne_polki)	205
Podszywanie pod organizację rządową	207
Podsumowanie: współpraca z platformami – trudności i wyzwania	209
Platformy Meta – Facebook, Instagram i Threads	209
X.com	214
TikTok	218
Google (YouTube i Google Ads)	221
LinkedIn	225
Discrod	227
Telegram	228
Bluesky	229

Wstęp

Raport przedstawia analizę działań podejmowanych przez platformy społecznościowe w zakresie moderacji treści mogących naruszać regulamin serwisu bądź inne regulacje prawne w okresie od 1 stycznia do 31 grudnia 2025 roku w odpowiedzi na zgłoszenia Ośrodka Analizy Dezinformacji NASK.

Naruszenia związane z incydentami cyberbezpieczeństwa w obszarze dezinformacji na platformach społecznościowych są poważnym zagrożeniem dla społeczeństwa. Mogą mieć realny wpływ na zmianę postrzegania rzeczywistości przez odbiorców, między innymi wskutek zamykania się w bańkach informacyjnych czy prowadzonych przez aktorów dezinformacji operacji celowego wywierania wpływu na poglądy obywateli i obywaterek. Poprzez rozpowszechnianie fałszywych przekazów w przestrzeni cyfrowej odbywa się próba wpływania na działania i funkcjonowanie różnych grup społecznych, do których kierowane są treści dezinformacyjne.

Metodologia

M.

Metodologia obsługi zgłoszeń Ośrodka Analizy Dezinformacji NASK

Celem raportu jest przedstawienie reakcji platform społecznościowych na zgłoszone incydenty, a także identyfikacja głównych problemów i wyzwań w tym obszarze. W raporcie uwzględniono zgłoszenia wewnętrzne, pochodzące z codziennego monitoringu mediów społecznościowych, prowadzonego przez OAD NASK przez siedem dni w tygodniu, a także zgłoszenia zewnętrzne, przesłane przez obywateli, instytucje bądź służby za pomocą zewnętrznych formularzy na stronach: nask.pl, bezpiecnewybory.pl¹ oraz na skrzynkę e-mail: dezinformacja@nask.pl.

Analiza obejmuje dane ilościowe dotyczące liczby zgłoszeń, liczby treści usuniętych, liczby materiałów poddanych moderacji oraz liczby treści odrzuconych przez platformy społecznościowe.

W raporcie uwzględniono także szczegółową kategoryzację zgłoszeń z uwzględnieniem rodzaju naruszenia zidentyfikowanego na podstawie regulaminów platform społecznościowych bądź innych regulacji prawnych. Raport zawiera również case studies, które skupiały się między innymi na przypadkach:

- dezinformacji i manipulacji na temat procesu wyborczego oraz kandydatów w wyborach prezydenckich;
- dezinformacji zdrowotnej;
- ofert sprzedaży fałszywej aplikacji mObywatel;
- reklam bezprawnie wykorzystujących wizerunek osób publicznych;
- podszywania się pod osoby oraz instytucje publiczne;
- zabójstwa na terenie kampusu Uniwersytetu Warszawskiego;
- ekshumacji ofiar rzezi wołyńskiej;
- zagadnień związanych z klimatem;
- incydentu z rosyjskimi dronami;
- trendu młodzieżowego „szon patrol”;

¹ Zgłoszenia zbierane w ramach projektu „Osłona Kampanii Wyborczej 2025” były do 06.06.2025 r.

- kampanii reklamowej izraelskiego rządu;
- masowego generowania zmanipulowanych treści AI;
- przekazów dezinformacyjnych na temat aktu dywersji na kolei;
- przekazów dezinformacyjnych związanych z zabójstwem 11-letniej uczennicy.

Ilustrują one specyfikę problemu szkodliwych treści w obrębie dezinformacji oraz prezentują reakcję platform na zgłoszenia w raportowanym okresie.

Na potrzeby raportu należy wyjaśnić przyjętą definicję zgłoszenia, rozumianego w tym przypadku również jako incydent dezinformacyjny. To wszystkie materiały zidentyfikowane przez Ośrodek Analizy Dezinformacji, które noszą znamiona naruszenia zapisów regulaminów platform społecznościowych lub łamią inne regulacje prawne, na przykład krajowe bądź unijne. Tylko tego rodzaju materiały uznawane są jako zasadne i przekazywane przez NASK do punktów kontaktowych platform społecznościowych lub za pomocą dedykowanych narzędzi zgłoszeniowych oferowanych OAD NASK przez serwisy. Zgłoszenia dzielone są ze względu na ich pochodzenie na: wewnętrzne i zewnętrzne, zgodnie z wcześniej wyjaśnioną ścieżką.

Głównym obszarem obsługi są incydenty dezinformacyjne, których treść jest stworzona w języku polskim. Wyjątek stanowią zgłoszenia dotyczące szerszych kampanii dezinformacyjnych, takich jak Undercut czy Overload.

W zależności od reakcji platform społecznościowych zgłoszone incydenty dzielimy na:

- **Odrzucone**, czyli odrzucone przez zespół moderacji bądź takie, które w ciągu dwóch tygodni nie uzyskały żadnej odpowiedzi. W przypadku zgłoszeń, które zawierają treść zagrażającą bezpieczeństwu państwa, zawsze dokonujemy odwołania przez dedykowany punkt kontaktowy, za pomocą dedykowanego narzędzia lub z poziomu użytkownika, jeśli to możliwe.
- **Poddane moderacji**, czyli otagowane w jakikolwiek sposób jako treści dezinformacyjne/manipulacyjne, na przykład community notes bądź stosowną, informacyjną nakładkę na materiale lub częściowe usunięcie fragmentu wideo. Do tej grupy należą również incydenty, w przypadku których konta sprawców zostały zawieszone na platformie.
- **Usunięte**, czyli takie które zostały w pełni obsłużone, a wskazana treść lub całe konto/kanal zostały trwale usunięte.

Co istotne, platformy społecznościowe nie informują o czasie rozpatrzenia zgłoszenia. Obejmuje to również te serwisy, które udostępniają specjalne narzędzia przeznaczone do obsługi zgłoszeń. Z tego względu OAD NASK nie

jest w stanie w krótkim czasie udzielić odpowiedzi obywatelom bądź instytucjom, ponieważ nie zna realnego czasu obsługi zgłoszenia ze strony platform. Każde zgłoszenie musi być śledzone indywidualnie przez analityka. W przypadku braku reakcji ze strony platformy zgłoszenie musi być ponawiane. W raporcie omówiono także wyzwania wynikające bezpośrednio ze współpracy z platformami społecznościowymi w zakresie obsługi zgłoszeń, takie jak brak przejrzystych procedur, długi czas reakcji czy niekonsekwentne podejście do moderacji treści.

Poniżej szczegółowo przedstawiono proces obsługi zgłoszeń przez OAD NASK do poszczególnych platform społecznościowych wraz z rodzajami naruszeń regulaminów serwisów:

Platformy Meta – Facebook, Instagram i Threads

Proces zgłaszania treści

Platforma Meta udostępnia pracownikom OAD NASK dedykowane narzędzie umożliwiające zgłaszanie szkodliwych treści pochodzących z serwisów Facebook, Instagram oraz Threads. Platforma pozwala na śledzenie bieżącego statusu zgłoszeń (bez dostępu do treści zgłoszenia) oraz złożenie odwołania w przypadku odrzucenia zgłoszenia.

W przypadku zgłoszeń o wysokim priorytecie pracownicy OAD NASK kontaktują się także z wyznaczonym punktem kontaktowym bezpośrednio.

X.com

Proces zgłaszania treści

- A. W przypadku platformy X brakuje kategorii dotyczącej dezinformacji/manipulacji. Z poziomu standardowego użytkownika w serwisie dostępne są jedynie poniższe kategorie zgłoszeń:
- **Nienawiść**
Obelgi, rasistowskie lub seksistowskie stereotypy, dehumanizacja, wzbudzanie strachu lub podżeganie do dyskryminacji, nawiązania nacechowane nienawiścią, symbole i logo nacechowane nienawiścią.
 - **Nadużycia i nękanie**
Obelgi, niechciane treści o zabarwieniu erotycznym i drastyczne treści sugerujące uprzedmiotowienie innych osób; nieodpowiednie lub drastyczne treści; negowanie aktów przemocy; nękanie konkretnych osób; podżeganie do nękania.

- **Mowa nienawiści**
Groźby karalne, życzenia obrażeń ciała, gloryfikacja przemocy, podżeganie do przemocy, zawołane podżeganie do przemocy.
- **Bezpieczeństwo dzieci**
Wykorzystywanie seksualne dzieci, grooming, przemoc fizyczna wobec dzieci, zbyt młody użytkownik.
- **Prywatność**
Udostępnianie informacji osobistych, zagrożenie udostępnieniem/ujawnieniem informacji osobistych, udostępnianie obrazów przedstawiających sytuacje intymne, na które nie wyrażono zgody; udostępnianie przedstawiających swój wizerunek obrazów, których nie życzą sobie na platformie.
- **Spam**
Fałszywe interakcje, oszustwa; fałszywe konta, złośliwe linki.
- **Samobójstwo lub samookaleczenie**
Zachęcanie do dokonywania aktów samookaleczenia, rozpowszechnianie takich zachowań lub udzielanie porad lub instrukcji ich dotyczących.
- **Nieodpowiednie lub niepokojące multimedia**
Drastyczne treści, nieuzasadnione okrucieństwo; nagość i zachowania seksualne osób dorosłych; brutalne zachowania na tle seksualnym; zoofilia i nekrofilia; multimedia przedstawiające zmarłą osobę.
- **Naśladownictwo**
Udawanie, że jest się kimś innym; obejmuje to także (w teorii) niezgodne konta parodystyczne/fanowskie.
- **Elementy związane z przemocą i nienawiścią**
Skrajny ekstremizm i terroryzm; grupy oraz sieci propagujące nienawiść.

Dostępna jest także dodatkowa klasyfikacja zgłoszeń z poziomu użytkownika związana z regulacjami Unii Europejskiej zawartymi w Akcie o usługach cyfrowych (Digital Services Act)². W jej ramach można zgłaszać treści nielegalne w UE oraz poszczególnych krajach członkowskich.

Znajdują się tam takie kategorie jak:

- Dobrostan zwierząt
- Naruszenia ochrony danych i prywatności
- Zniesławienie/pomówienie
- Nielegalne lub szkodliwe wypowiedzi

² <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> (dostęp: 26.01.2026 godz. 12:40)

- Negatywny wpływ na debatę obywatelską lub wybory
 - Zachowanie bez zgody
 - Pornografia lub treści o charakterze seksualnym
 - Ochrona osób małoletnich
 - Ryzyko dla bezpieczeństwa publicznego
 - Oszustwa i fałszerstwa
 - Zakres usług platformy
 - Samookaleczenie
 - Produkty niebezpieczne i nielegalne
 - Przemoc
- B.** Kolejny krok w procesie zgłaszania treści stanowi przekazanie ich przez OAD NASK do ustalonego z platformą punktu kontaktowego. Zgodnie z uzgodnioną procedurą materiały są zgłaszane do punktu po ich uprzednim zgłoszeniu do platformy z poziomu użytkownika. Warto w tym miejscu nadmienić, że w okresie poprzedzającym wybory prezydenckie w Polsce w 2025 roku punkt został rozszerzony z 2 do 3 opiekunów, z czego jeden miał w całości odpowiadać za zgłoszenia związane z dezinformacją wyborczą.
- C.** Od III kwartału 2025 roku dla OAD NASK posiada również dostęp do narzędzia zgłoszeniowego „Legal request submissions” przeznaczonego między innymi dla funkcjonariuszy organów ścigania oraz urzędników państwowych. Umożliwia on złożenie wniosku o wstrzymanie publikacji treści na podstawie żądania prawnego, również w przypadku wystąpienia treści dezinformacyjnej.

Zgodnie z ustaleniami z punktem kontaktowym X, wymienione poniżej incydenty nie wymagają zgłoszenia za pośrednictwem narzędzia:

- treści typu deepfake oraz innego rodzaju manipulacje materiałami audiowizualnymi;
- siatki kont-botów oraz pozostałe skoordynowane nieautentyczne działania (CIB).

Informacja o tych incydentach jest zgłaszana poprzez przesłanie wiadomości e-mail do punktu kontaktowego wyznaczonego przez platformę X.

TikTok

Proces zgłaszania treści

TikTok również oferuje dedykowane narzędzie zgłoszeniowe. Umożliwia ono dodawanie nowych zgłoszeń oraz przeglądanie już istniejących. W tym przypadku nie przewidziano jednak funkcji odwołania się od negatywnie rozpatrzonej sprawy. Platforma działa w języku angielskim, lecz w przypadku szczególnie ważnych zgłoszeń do dyspozycji pracowników jest polskojęzyczny opiekun w ramach punktu kontaktowego.

Google (YouTube i Google Ads)

Proces zgłaszania treści

W przypadku platformy Google zespół OAD NASK koncentruje swój monitoring głównie na jej dwóch serwisach: YouTube oraz Google Ads.

W porozumieniu z punktem kontaktowym przyjęto, że wszystkie incydenty (z wykluczeniem serwisu Google Play) będą zgłaszane drogą mailową bezpośrednio do polskiego opiekuna. Zgłoszenia incydentów zaobserwowanych w sklepie Play należy przysyłać za pomocą dedykowanego formularza. Zespół posiada również możliwość zgłaszania treści bezpośrednio z poziomu użytkownika.

LinkedIn

Proces zgłaszania treści

W przypadku sieci społecznościowej LinkedIn OAD NASK nie ma utworzonego punktu kontaktowego. W związku z tym jedyną opcją przekazywania treści dezinformacyjnych do platformy jest zgłaszanie ich z poziomu użytkownika.

Discrod

Proces zgłaszania treści

OAD NASK nie posiada na platformie komunikacyjnej Discrod punktu kontaktowego do zgłaszania treści dezinformacyjnych. Z tego powodu zgłaszanie materiałów możliwe jest jedynie poprzez ogólny formularz w Centrum Pomocy Discrod³ oraz z poziomu użytkownika w samej aplikacji.

³ <https://support.Discrod.com/> (dostęp: 28.01.2026, godz. 15:00)

Ogólny formularz zgłoszeń zawiera następujące kategorie kontaktowe:

- Pomoc i wsparcie
- Zwroty, anulowanie i płatności
- Zhakowane konto
- Odwołania i prośby o aktualizację wieku
- Zgłaszanie błędów
- Błędy w tłumaczeniu
- Żądanie usunięcia treści zgodnie z ustawą DMCA
- Złożenie roszczenia wzajemnego DMCA
- Złożenie skargi dotyczącej znaków towarowych
- Pomoc związana z monetyzacją serwera
- Odwołania i pytania właściciela serwera NSFW
- Kontakt z zespołem ds. prywatności Discroda
- Zespół wsparcia deweloperów ds. zgodności
- Zespół wsparcia dla deweloperów

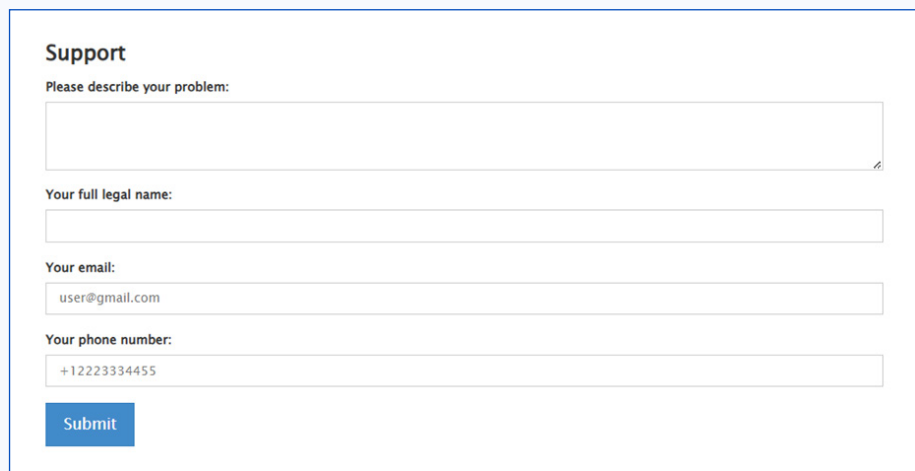
Jeśli chodzi o zgłaszanie z poziomu użytkownika w aplikacji Discrod, to dostępne są, takie kategorie jak:

- Nie lubię tego
- Spam
- Nadużycie lub nękanie
- Szkodliwa dezinformacja lub gloryfikowanie przemocy
- Ujawnianie prywatnych danych identyfikacyjnych
- Coś innego
- Zgłoś nielegalne treści na mocy aktu o usługach cyfrowych

Telegram

Proces zgłaszania treści

OAD NASK nie posiada punktu kontaktowego na platformie Telegram. Jedyną opcją w tym wypadku jest bezpośrednie zgłoszenie dezinformacyjnych materiałów do platformy poprzez ogólny formularz na jej stronie⁴.

A screenshot of the Telegram support form. The form is titled "Support" and contains the following fields: "Please describe your problem:" (a large text area), "Your full legal name:" (a text field), "Your email:" (a text field with the placeholder "user@gmail.com"), and "Your phone number:" (a text field with the placeholder "+12223334455"). A blue "Submit" button is located at the bottom of the form.

OBRAZ 1. Formularz zgłoszeniowy na platformie Telegram

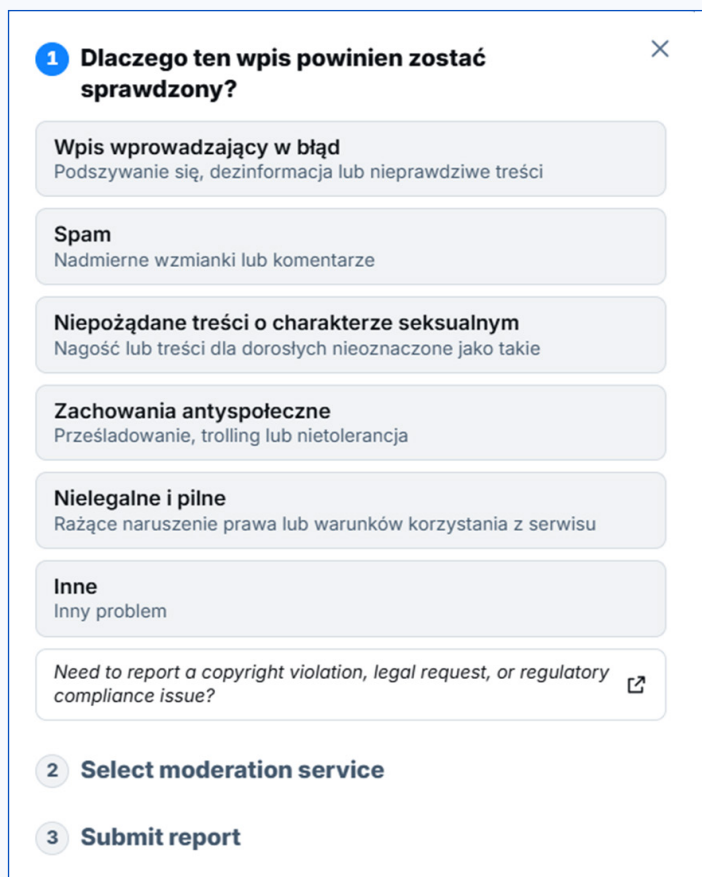
ŹRÓDŁO: [TELEGRAM.ORG BLUESKY](https://telegram.org/bluesky)

⁴ <https://Telegram.org/support> (dostęp: 28.01.2026, godz. 15:00)

Bluesky

Proces zgłaszania treści (formularz z poziomu użytkownika)

OAD NASK nie posiada punktu kontaktowego na platformie Telegram. Jedyną opcją w tym wypadku jest bezpośrednie zgłoszenie dezinformacyjnych materiałów do platformy poprzez ogólny formularz na jej stronie.



The image shows a mobile app interface for reporting content on Bluesky. It consists of a white box with a blue border and a close button (X) in the top right corner. The interface is divided into three numbered steps:

- 1 Dlaczego ten wpis powinien zostać sprawdzony?**
 - Wpis wprowadzający w błąd**
Podszywanie się, dezinformacja lub nieprawdziwe treści
 - Spam**
Nadmierne wzmianki lub komentarze
 - Niepożądane treści o charakterze seksualnym**
Nagość lub treści dla dorosłych nieoznaczone jako takie
 - Zachowania antyspołeczne**
Prześladowanie, trolling lub nietolerancja
 - Nielegalne i pilne**
Rażące naruszenie prawa lub warunków korzystania z serwisu
 - Inne**
Inny problem
 - Need to report a copyright violation, legal request, or regulatory compliance issue?* (with an external link icon)
- 2 Select moderation service**
- 3 Submit report**

OBRAZ 2.

Kategorie naruszeń dostępne w formularzu na platformie Bluesky

ŹRÓDŁO: [BLUESKY.COM](https://bluesky.com)

Podsumowanie 2025

P.

Kluczowe wnioski

W 2025 roku zgłoszono łącznie 46 511 materiałów. Platformy usunęły 12 proc. zgłoszonych treści, czyli 5572 materiały, a moderacji poddano 68 proc. zgłoszonych treści, a więc 31 638 materiałów. Platformy odrzuciły jednak 20 proc. zgłoszeń, czyli 9301 przypadków.

Przeprowadzona analiza ujawnia różnice w reakcji platform społecznościowych na zgłaszane incydenty w kolejnych kwartałach.

Na przestrzeni 2025 roku znacząco wzrósł udział zgłoszeń poddanych moderacji – w I kwartale treści zakwalifikowane do tej kategorii stanowiły 4 proc. wszystkich zgłoszeń, natomiast w IV kwartale moderacji zostało poddanych już 71 proc. zgłoszonych materiałów. Na prezentowane dane wpływ miały zgłoszenia profili zaangażowanych w operacje Undercut oraz Overload, których aktywność nasiliła się w drugiej połowie 2025 roku. W odpowiedzi na zgłoszenia platformy decydowały o zawieszeniu działalności tych kont. Takie działanie jest klasyfikowane według przyjętej metodologii jako moderacja treści.

W I kwartale 2025 roku odnotowano stosunkowo wysoki odsetek treści usuniętych przez platformy społecznościowe (35 proc.). Wśród zgłoszonych incydentów sporą grupę stanowiły oszustwa udające reklamy nieistniejących inwestycji i przypadki podszywania pod osoby publiczne. Od II do IV kwartału 2025 roku udział treści usuniętych przez platformy społecznościowe pozostawał względnie stały, wahając się w przedziale od 8 do 14 proc. zgłaszanych materiałów.

Uwagę zwraca malejący odsetek zgłoszeń odrzuconych przez platformy społecznościowe. W I kwartale 2025 roku kategoria ta stanowiła 61 proc. wszystkich zgłoszeń. Dane dotyczące reakcji platform na zgłoszenia przesłane w ostatnim kwartale roku wskazują, że udział odrzuconych zgłoszeń spadł do 16 proc.

Zasadne zgłoszenia przekazane do platform społecznościowych

W I kwartale 2025 roku do administracji platform społecznościowych zgłoszono 2366 materiałów. Platformy odrzuciły 61 proc. zgłoszeń, czyli 1452 przypadków. Moderacji poddano 4 proc. zgłoszonych treści, czyli 95 materiałów. Platformy usunęły 35 proc. zgłoszonych treści, czyli 819 materiałów.

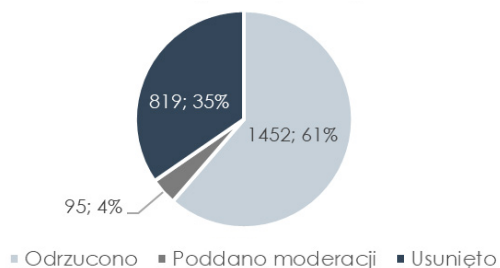
W II kwartale 2025 roku zgłoszono 22 642 materiały. Platformy odrzuciły 20 proc. zgłoszeń, czyli 4536 przypadków. Moderacji poddano 72 proc. zgłoszonych treści, czyli 16 302 materiały. Platformy usunęły 8 proc. zgłoszonych treści, czyli 1804 materiały.

W III kwartale 2025 roku zgłoszono 14 005 materiałów. Platformy odrzuciły 15 proc. zgłoszeń, czyli 2114 przypadków. Moderacji poddano 71 proc. zgłoszonych treści, czyli 9918 materiałów. Platformy usunęły 14 proc. zgłoszonych treści, czyli 1973 materiały.

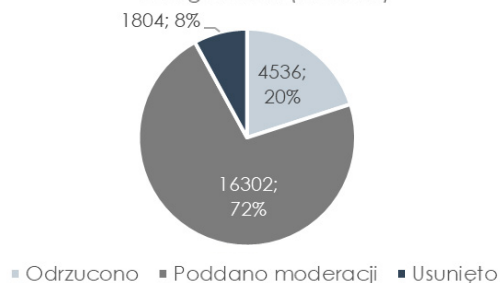
W IV kwartale 2025 roku zgłoszono 7498 materiałów. Platformy odrzuciły 16 proc. zgłoszeń, czyli 1199 przypadków. Moderacji poddano 71 proc. zgłoszonych treści, czyli 5323 materiały. Platformy usunęły 13 proc. zgłoszonych treści, czyli 976 materiałów.

W 2025 roku zgłoszono razem 46 511 treści. Platformy odrzuciły 20 proc. zgłoszeń, czyli 9301 przypadków. Moderacji poddano 68 proc. zgłoszonych treści, a więc 31 638 materiałów. Platformy zdecydowały się usunąć 12 proc. zgłoszonych treści, czyli 5572 materiały.

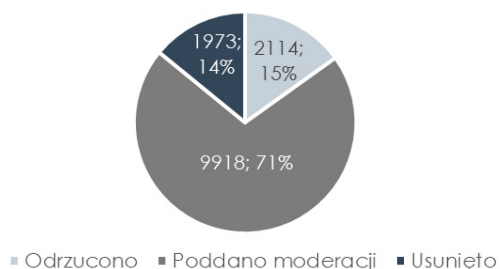
Wykres 1. Reakcja platform społecznościowych na zgłoszenia (Q1 2025)



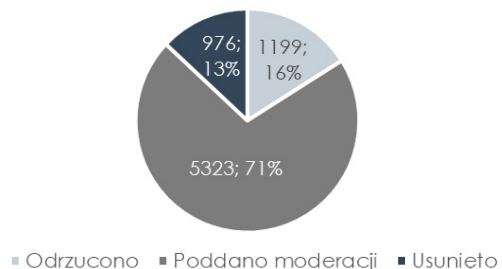
Wykres 2. Reakcja platform społecznościowych na zgłoszenia (Q2 2025)



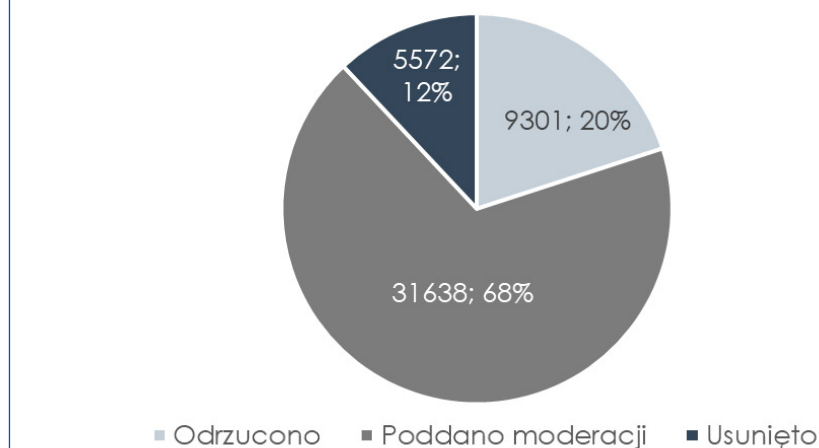
Wykres 3. Reakcja platform społecznościowych na zgłoszenia (Q3 2025)



Wykres 4. Reakcja platform społecznościowych na zgłoszenia (Q4 2025)

**WYKRESY 1-4.** Reakcja platform społecznościowych na zgłoszenia (Q1-Q4 2025)

Wykres 5. Reakcja platform społecznościowych na zgłoszenia (2025)

**WYKRES 5.** Reakcja platform społecznościowych na zgłoszenia (2025)

ŹRÓDŁO: OPRACOWANIE WŁASNE

Zgłoszenia zewnętrzne

W I kwartale 2025 roku do administracji platform społecznościowych przekazano 1411 zgłoszeń pochodzących od obywateli i instytucji. Platformy odrzuciły 66 proc. zgłoszeń, czyli 929 przypadków. Moderacji poddano 4 proc. zgłoszonych treści, czyli 58 materiałów. Platformy usunęły 30 proc. zgłoszonych treści, czyli 424 materiały.

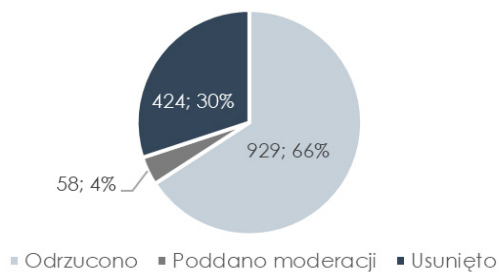
W II kwartale 2025 roku przekazano 3955 zgłoszeń pochodzących od obywateli i instytucji. Platformy odrzuciły 81 proc. zgłoszeń, czyli 3222 przypadki. Moderacji poddano 3 proc. zgłoszonych treści, czyli 119 materiałów. Platformy usunęły 16 proc. zgłoszonych treści, czyli 614 materiałów.

W III kwartale 2025 roku przekazano 2466 zgłoszeń pochodzących od obywateli i instytucji. Platformy odrzuciły 69 proc. zgłoszeń, czyli 1695 przypadków. Moderacji poddano 1 proc. zgłoszonych treści, czyli 22 materiały. Platformy usunęły 30 proc. zgłoszonych treści, czyli 749 materiałów.

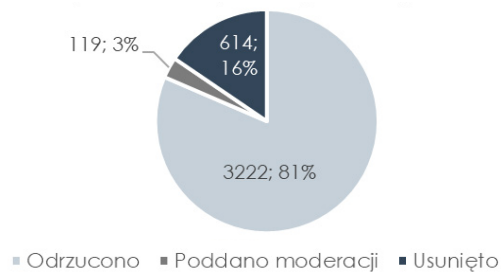
W IV kwartale 2025 roku przekazano 1663 zgłoszeń pochodzących od obywateli i instytucji. Platformy odrzuciły 57 proc. zgłoszeń, czyli 943 przypadki. Moderacji poddano 3 proc. zgłoszonych treści, czyli 47 materiałów. Platformy usunęły 40 proc. zgłoszonych treści, czyli 673 materiały⁵.

⁵ Zgłoszenia zewnętrzne to treści przesłane za pomocą formularzy ze stron nask.pl oraz bezpiecznewybory.pl, a także wszelkie materiały przekazane na skrzynkę dezinformacja@nask.pl. W podsumowaniu uwzględniono wyłącznie zgłoszenia zasadne, czyli takie, które naruszają regulamin platform społecznościowych bądź inne, istniejące regulacje prawne.

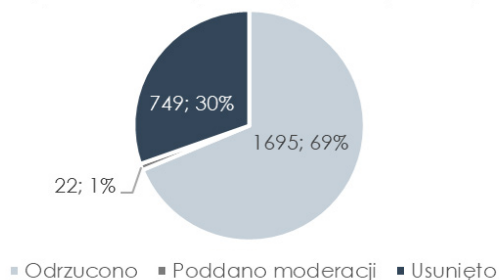
Wykres 6. Reakcja platform społecznościowych na zgłoszenia (Q1 2025)



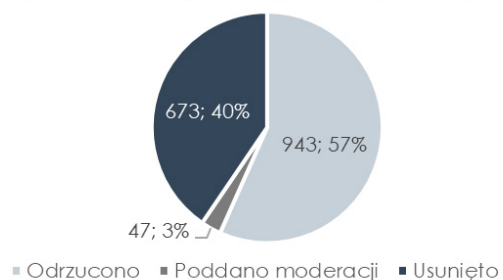
Wykres 7. Reakcja platform społecznościowych na zgłoszenia (Q2 2025)



Wykres 8. Reakcja platform społecznościowych na zgłoszenia (Q3 2025)



Wykres 9. Reakcja platform społecznościowych na zgłoszenia (Q4 2025)



WYKRESY 6-9. Reakcja platform społecznościowych na zgłoszenia zewnętrzne (Q1-Q4 2025)

ŹRÓDŁO: OPRACOWANIE WŁASNE

Zgłoszenia wewnętrzne

W I kwartale 2025 roku do platform społecznościowych zgłoszono 955 incydentów zidentyfikowanych w wyniku bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej prowadzonej przez OAD NASK. Platformy odrzuciły 55 proc. zgłoszeń, czyli 523 przypadki. Moderacji poddano 4 proc. zgłoszonych treści, czyli 37 materiałów. Platformy usunęły 41 proc. zgłoszonych treści, czyli 395 materiałów.

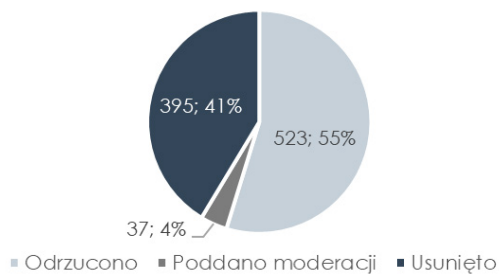
W II kwartale 2025 roku zgłoszono 18 687 incydentów zidentyfikowanych w wyniku bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej prowadzonej przez OAD NASK. Platformy odrzuciły 7 proc. zgłoszeń, czyli 1314 przypadków. Moderacji poddano 87 proc. zgłoszonych treści, czyli 16 183 materiały. Platformy usunęły 6 proc. zgłoszonych treści, czyli 1190 materiałów.

W III kwartale 2025 roku zgłoszono 11 539 incydentów zidentyfikowanych w wyniku bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej prowadzonej przez OAD NASK. Platformy odrzuciły 4 proc. zgłoszeń, czyli 419 przypadków. Moderacji poddano 86 proc. zgłoszonych treści, czyli 9896 materiałów. Platformy usunęły 10 proc. zgłoszonych treści, czyli 1224 materiały.

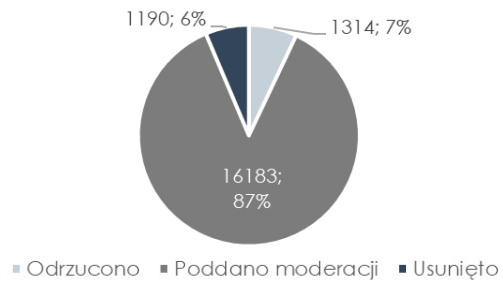
W IV kwartale 2025 roku zgłoszono 5835 incydentów zidentyfikowanych w wyniku bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej prowadzonej przez OAD NASK. Platformy odrzuciły 5 proc. zgłoszeń, czyli 303 przypadki. Moderacji poddano 91 proc. zgłoszonych treści, czyli 5276 materiałów. Platformy usunęły 4 proc. zgłoszonych treści, czyli 256 materiałów⁶.

⁶ Zgłoszenia wewnętrzne pochodzą z bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej, który jest prowadzony przez 7 dni w tygodniu.

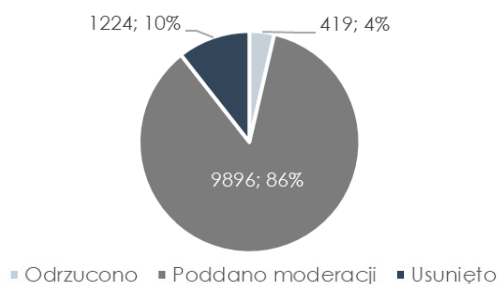
Wykres 10. Reakcja platform społecznościowych na zgłoszenia (Q1 2025)



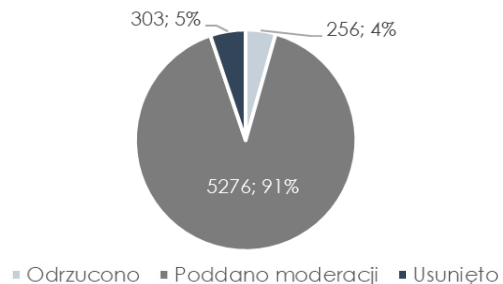
Wykres 11. Reakcja platform społecznościowych na zgłoszenia (Q2 2025)



Wykres 12. Reakcja platform społecznościowych na zgłoszenia (Q3 2025)



Wykres 13. Reakcja platform społecznościowych na zgłoszenia (Q4 2025)



WYKRESY 10-13. Reakcja platform społecznościowych na zgłoszenia wewnętrzne (Q1-Q4 2025)

ŹRÓDŁO: OPRACOWANIE WŁASNE

Podsumowanie 2025 w podziale na platformy społecznościowe

Wykres przedstawia reakcje platform społecznościowych na incydenty zgłoszone przez analityków OAD NASK. W 2025 roku materiały przekazano do takich serwisów jak: X, TikTok, Threads, Facebook, Instagram, Google (Ads i YouTube), Telegram, LinkedIn, Discrod i Bluesky. Grafika ilustruje również reakcje serwisów oznaczonych jako „Inne”. W tej grupie znalazły się między innymi serwisy Reddit i Twitch.

Administracja platformy Threads (Meta) odrzuciła 91 proc. zgłoszeń, czyli 202 przypadki, i usunęła 9 proc. zgłoszonych treści, czyli 20 materiałów. Platforma Facebook (Meta) odrzuciła 50 proc. zgłoszeń, czyli 4276 przypadków. Moderacji poddano 10 proc. zgłoszonych treści, czyli 840 materiałów. Platforma usunęła 40 proc. zgłoszonych treści, czyli 3399 materiałów. Platforma Instagram (Meta) odrzuciła 71,5 proc. zgłoszeń, czyli 271 przypadków. Moderacji poddano 1,6 proc. zgłoszonych treści, czyli 6 materiałów. Platforma usunęła 26,9 proc. zgłoszonych treści, czyli 102 materiały.

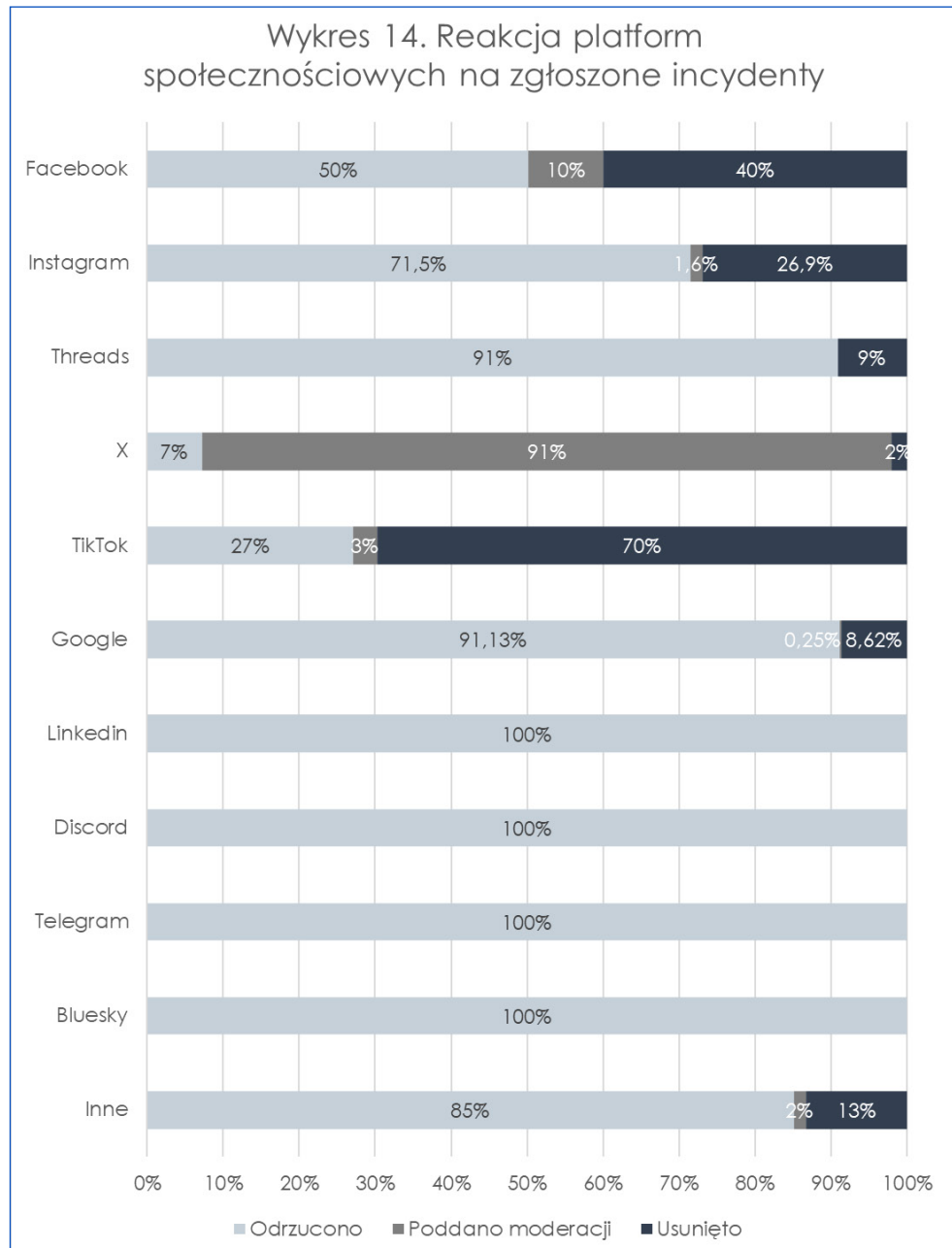
Platforma X odrzuciła 7 proc. zgłoszeń, czyli 2441 przypadków. Moderacji poddano 91 proc. zgłoszonych treści, czyli 30 731 materiałów. Platforma usunęła 2 proc. zgłoszonych treści, czyli 662 materiały.

Platforma TikTok odrzuciła 27 proc. zgłoszeń, czyli 480 przypadków. Moderacji poddano 3 proc. zgłoszonych treści, czyli 55 materiałów. Platforma usunęła 70 proc. zgłoszonych treści, czyli 1233 materiały.

Platformy Google (Ads i YouTube) odrzuciły 91,13 proc. zgłoszeń, czyli 1480 przypadków. Moderacji poddano mniej niż 0,25 proc. zgłoszonych treści, czyli 4 materiały. Platformy usunęły 8,62 proc. zgłoszonych treści, czyli 140 materiałów.

Platformy Telegram, LinkedIn, Discrod oraz Bluesky odrzuciły wszystkie przesłane zgłoszenia. Było to odpowiednio 12 przypadków na platformie Telegram, 18 na platformie LinkedIn, 15 na platformie Discrod i 3 na platformie Bluesky.

Inne platformy odrzuciły 85 proc. zgłoszeń, czyli 103 przypadki. Moderacji poddano 2 proc. zgłoszonych materiałów, czyli 2 materiały. Platformy z tej grupy usunęły 13 proc. zgłoszonych treści, czyli 16 materiałów.



WYKRES 14.

Reakcja platform społecznościowych na zgłoszone incydenty

ŹRÓDŁO: OPRACOWANIE WŁASNE

Szczegółowe dane statystyczne, przykładowe case studies oraz metodologia obsługi zgłoszeń realizowanych przez OAD NASK znajdują się we właściwych raportach kwartalnych.

I kwartał 2025

Q1.

I kwartał 2025

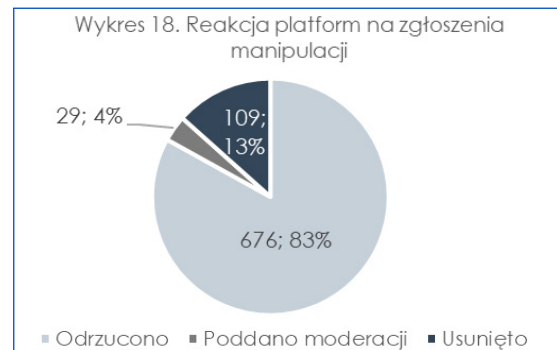
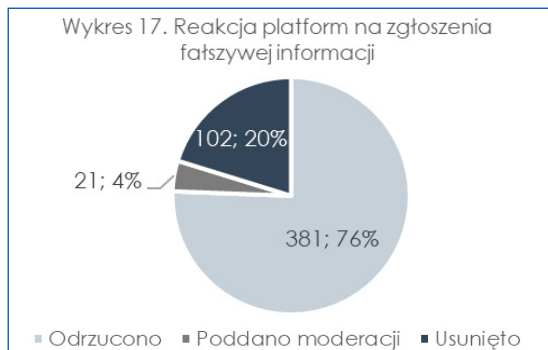
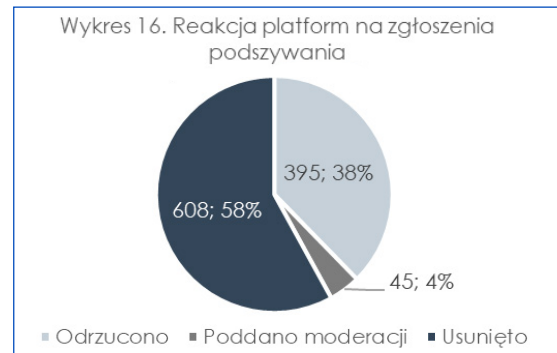
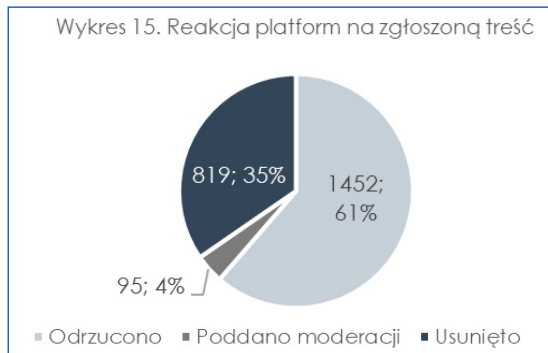
W I kwartale 2025 roku OAD NASK zgłosił do administracji platform społecznościowych 2366 przypadków treści naruszających regulacje. Usuniętych zostało 35 proc. zgłoszonych materiałów, czyli 819 publikacji. Dalsze 4 proc., czyli 95 wpisów, zostało poddanych moderacji, a 61 proc. zgłoszeń, czyli 1452 przypadki, nie spotkało się z żadną zauważalną reakcją ze strony moderacji.

- Spośród wszystkich zgłoszonych treści 504 materiały stanowiły fałszywą informację, 814 zawierało manipulacje, a 1048 związane były z podszywaniem się pod osobę lub instytucję publiczną.
- Platformy najczęściej podejmowały działania w reakcji na incydenty związane z podszywaniem się – w analizowanym okresie usunięto lub poddano moderacji 58 proc. zgłoszonych treści, czyli 608 incydentów podszywania się. Oznacza to jednak, że 42 proc. fałszywych stron i profili, czyli 440 zgłoszonych przypadków, nadal pozostaje aktywnych.
- Moderacje serwisów często różnie rozpatrują podobne przypadki podszywania się, podejmując działanie w odpowiedzi na niektóre zgłoszenia, podczas gdy inne (czasem dotyczące tej samej instytucji), odrzucają.
- W przypadku fałszywych informacji usunięto lub poddano moderacji 24 proc. zgłoszonych treści, czyli 123 materiały, a odrzucono 76 proc. zgłoszeń, czyli 381 przypadków.
- Platformy wykazały najmniejszy poziom zaangażowania w reakcji na zgłoszenia treści manipulacyjnych. W tej kategorii usunięto 13 proc. zgłoszonych treści, czyli 109 materiałów, a 4 proc. zgłoszonych treści, czyli 29 materiałów, poddano moderacji. Bez reakcji pozostawiono 83 proc. zgłoszeń, czyli 676 przypadków.
- Zgłoszenia treści ocenionych przez OAD NASK jako naruszające regulaminy platform lub mogące łamać inne regulacje prawne w wielu przypadkach zostały odrzucone bez podania powodu lub z adnotacją, że nie naruszają regulaminu danej platformy społecznościowej.
- Mimo deklaracji platform o priorytetowym traktowaniu kwestii dezinformacji związanej z okresem kampanii wyborczej zespół OAD NASK zaobserwował niski poziom reakcji w rozpatrywaniu zgłoszeń dotyczących procesu wyborczego.

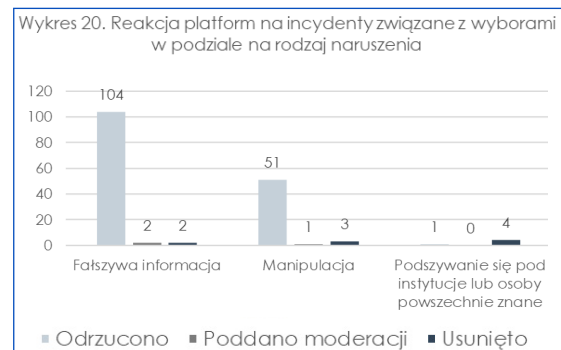
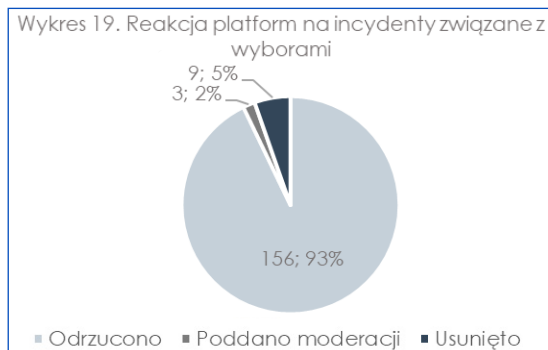
- Platformom przekazano 168 zgłoszeń dotyczących tematów wyborczych, z których 12 (7 proc.) zostało usuniętych lub poddanych moderacji. Nadal dostępnych jest 156 zgłoszonych materiałów (93 proc.), które mogły mieć negatywny wpływ na dyskurs wyborczy w Polsce.
- Odnotowano ograniczone działania platform w odniesieniu do fałszywych, łatwych do zweryfikowania informacji dotyczących procesu wyborczego oraz kandydatów. Tylko 4 ze 108 przypadków (4 proc.) spotkały się z reakcją ze strony serwisów (2 z nich usunięto a 2 poddano moderacji).
- Niski był również poziom zaangażowania platform w zakresie usuwania treści wygenerowanych przez AI, które miały ośmieszać kandydatów bądź dezinformować na temat ich skali poparcia.
- Reklamy o charakterze manipulacyjnym⁷ odnotowano między innymi w serwisach firmy Meta. Zdarzało się, że zgłoszenia tego rodzaju kończyły się zablokowaniem pojedynczej reklamy, a nie całego profilu lub strony, co pozwalało sprawcom na dalszą szkodliwą aktywność.
- Dodatkowym utrudnieniem związanym z kwestią zgłaszania szkodliwych reklam pojawiających się na platformie Facebook jest to, że ogólnodostępna biblioteka reklam jest aktualizowana z opóźnieniem. Ograniczona jest także możliwość wyszukiwania i filtrowania zamieszczanych przez reklamodawców treści.
- Przy okazji pojawienia się na platformie Discrod ofert sprzedaży fałszywej aplikacji mObywatel administracja serwisu nie rozpatrzyła pozytywnie żadnego z 13 wysłanych do niej zgłoszeń (mimo że proceder był adresowany do osób nieletnich).
- OAD NASK posiada wyznaczone punkty kontaktowe z platformami TikTok, X i serwisami firm Meta oraz Google. Spośród nich tylko platformy Meta oraz TikTok udostępniały dedykowane narzędzie/platformę zgłoszeniową OAD NASK. Google posiada między innymi standardowy formularz dla użytkowników, a dodatkowo także *The Misinformation Priority Flagger Program*⁸, w którego opisie zaznaczono, że firma nie jest zobowiązana do podejmowania jakichkolwiek działań na podstawie przekazanych zgłoszeń.

⁷ Wprowadzających w błąd treścią lub wykorzystujących deepfake'i, mające na celu wyłudzenie danych lub środków bądź wywieranie wpływu na poglądy grupy docelowej.

⁸ <https://support.google.com/youtube/answer/7554338?hl=en> (dostęp: 26.01.2026, godz. 12:00)



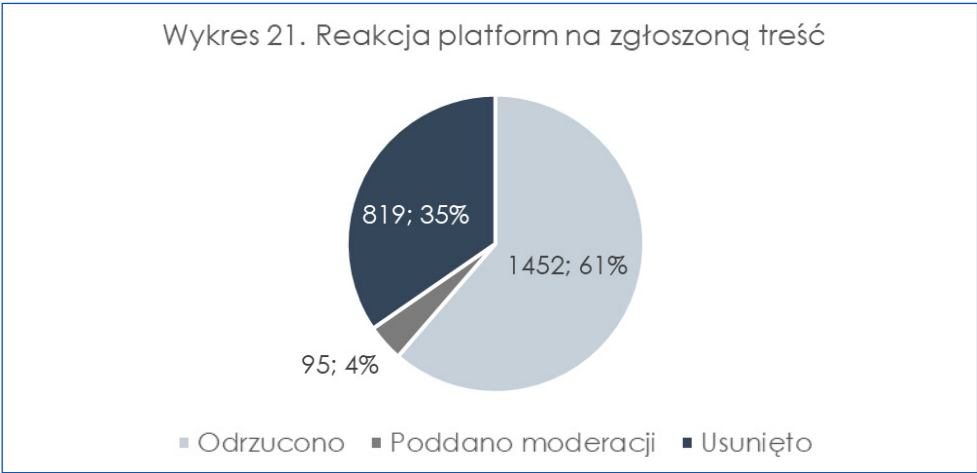
WYKRESY 15-18. Reakcja platform na zgłoszenia: treść (15); podszywanie (16); fałszywa informacja (17); manipulacja (18)



WYKRESY 19-20. Reakcja platform na incydenty związane z wyborami (19); w podziale na rodzaj naruszenia (20)

ŹRÓDŁO: OPRACOWANIE WŁASNE

Reakcja platform na wszystkie zgłoszenia OAD NASK



WYKRES 21. Reakcja platform na zgłoszoną treść

ŹRÓDŁO: OPRACOWANIE WŁASNE

W I kwartale 2025 roku z platform społecznościowych usunięto 35 proc. zgłoszonych przez OAD NASK treści, czyli 819 wpisów. Dalszych 4 proc. zgłoszonych treści, czyli 95 wpisów, zostało poddanych moderacji, natomiast 61 proc. zgłoszeń, czyli 1452 przypadki, nie spotkało się z żadną widoczną reakcją ze strony moderacji.

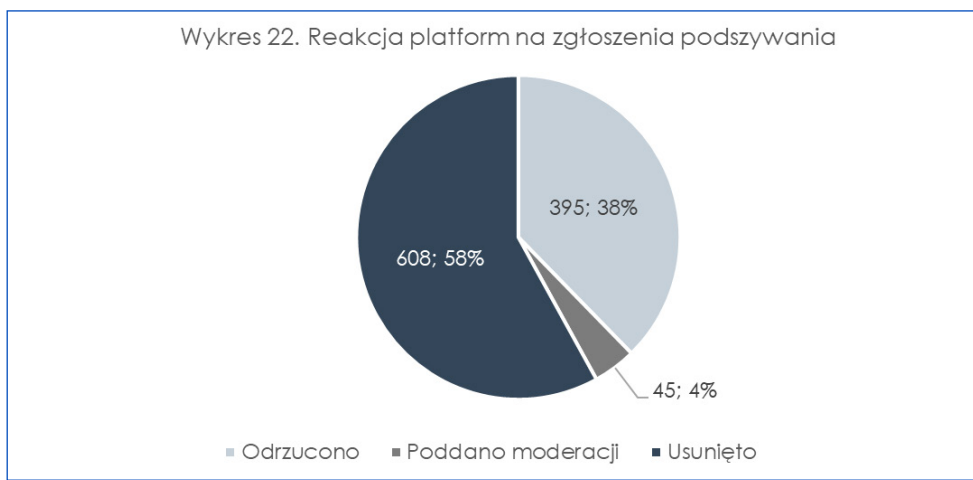
Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/institucję
Liczba zgłoszeń odrzuconych	381	676	395
Liczba zgłoszeń poddanych moderacji	21	29	45
Liczba zgłoszeń usuniętych	102	109	608
Całkowita liczba zgłoszeń	504	814	1048

TABELA 1. Zgłoszenia z podziałem na rodzaj naruszeń

ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń. Wśród wszystkich zgłoszeń zidentyfikowano:

- **21,3 proc. (504 przypadki)** wpisów zawierających fałszywe informacje;
- **34,4 proc. (814 przypadków)** wpisów zawierających manipulacje;
- **44,3 proc. (1048 przypadków)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

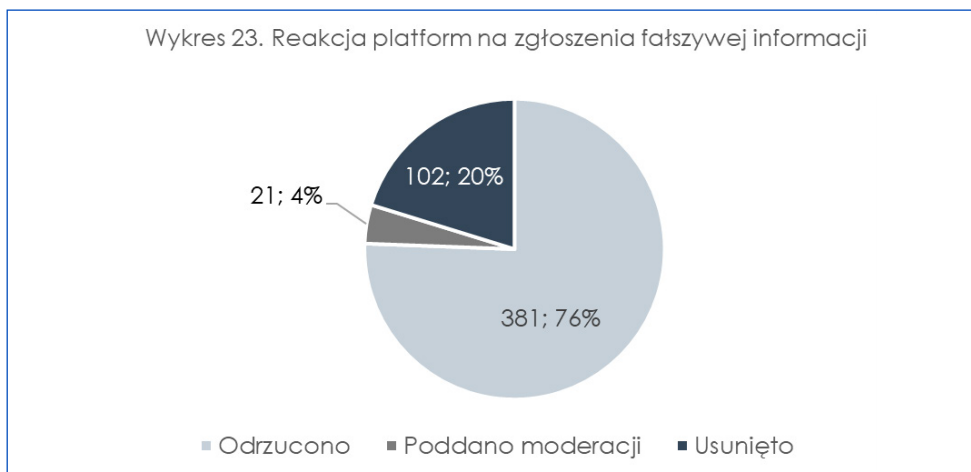


WYKRES 22.

Reakcja platform na zgłoszenia podszywania

ŹRÓDŁO: OPRACOWANIE WŁASNE

Analiza zgłoszonych treści wykazała, że platformy społecznościowe najczęściej reagowały na incydenty polegające na podszywaniu się pod osobę lub instytucję – usunięto 58 proc. zgłoszonych treści z tej kategorii, czyli 608 incydentów podszywania się. W badanym okresie platformy społecznościowe decydowały także czasem o zawieszeniu zgłoszonych kont. Takie przypadki zostały w tym raporcie zakwalifikowane jako incydenty „poddane moderacji”. W sumie zawieszonych zostało 4 proc. zgłoszonych profili, czyli 45 przypadków. Oznacza to jednak, że z 38 proc. zgłoszonych fałszywych stron lub profili, czyli 395 przypadków, nadal pozostaje aktywne.

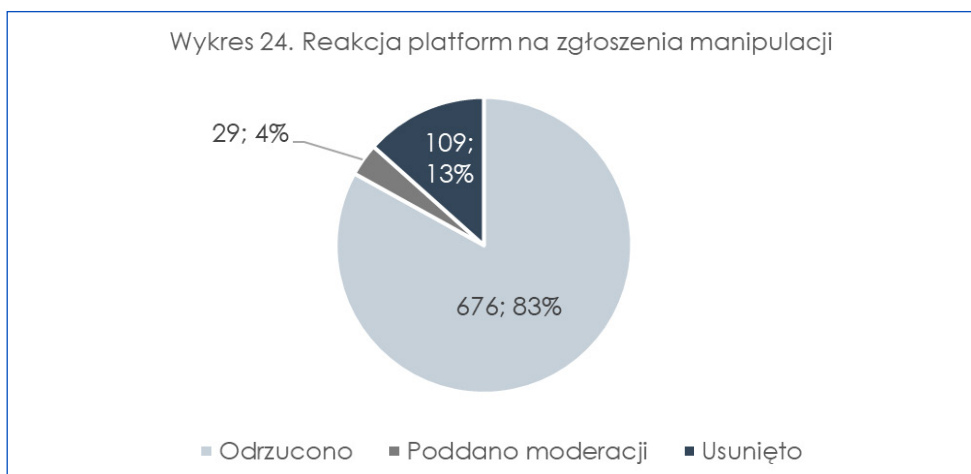


WYKRES 23.

Reakcja platform na zgłoszenia fałszywej informacji

ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku materiałów zawierających fałszywe informacje usunięciu lub moderacji poddano 24 proc. treści (123 materiały), a odrzucono 76 proc. zgłoszeń (381 przypadków).



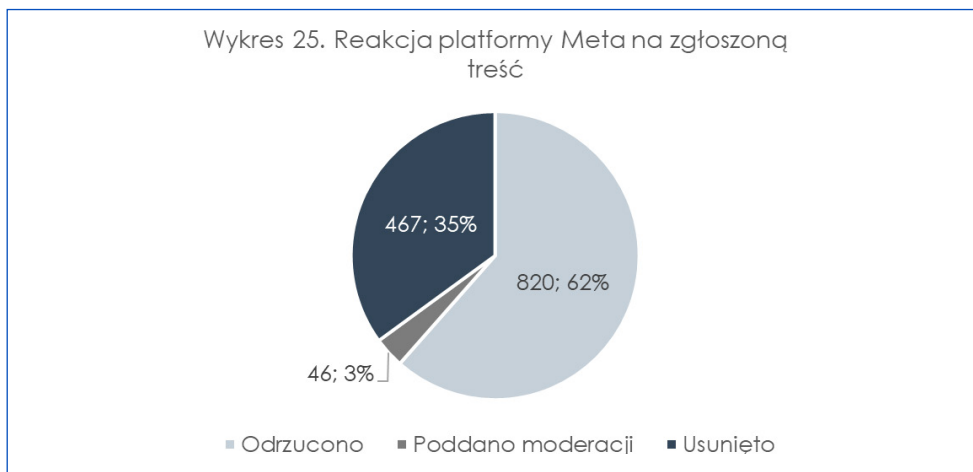
WYKRES 24.

Reakcja platform na zgłoszenia manipulacji

ŹRÓDŁO: OPRACOWANIE WŁASNE

Należy zwrócić uwagę, że platformy rzadziej podejmowały działania polegające na usuwaniu treści manipulacyjnych, szczególnie tych, które wymagały analizy z uwzględnieniem lokalnego kontekstu. Decyzję o usunięciu podjęto w przypadku 13 proc. zgłoszonych treści (109 materiałów), 4 proc. (29 materiałów) poddano moderacji, a odrzuconych zostało 83 proc. zgłoszeń (676 przypadków).

Meta

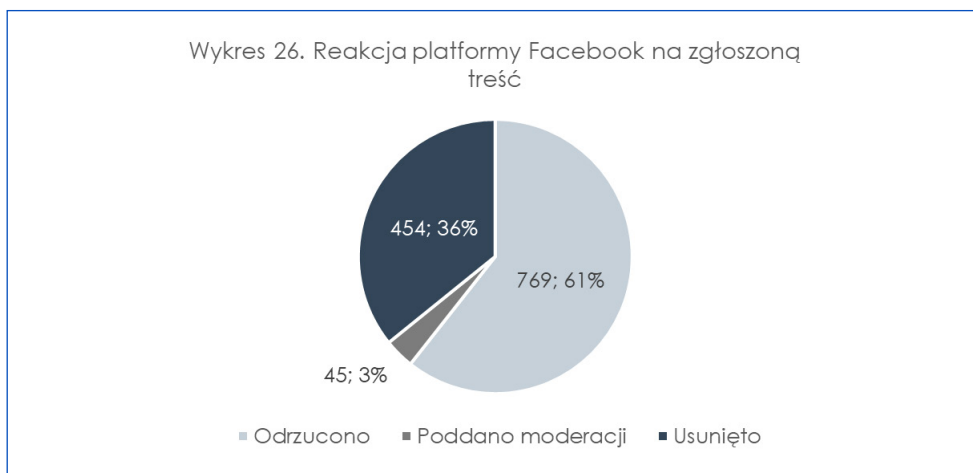


WYKRES 25. Reakcja platformy Meta na zgłoszoną treść

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszonych do administracji platform firmy Meta treści pochodzących z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 35 proc. zgłoszonych treści, czyli 467 materiałów, 3 proc. zgłoszonych treści, czyli 46 materiałów, poddano moderacji, a 62 proc. zgłoszeń, czyli 820 przypadków, zostało odrzuconych.

Facebook



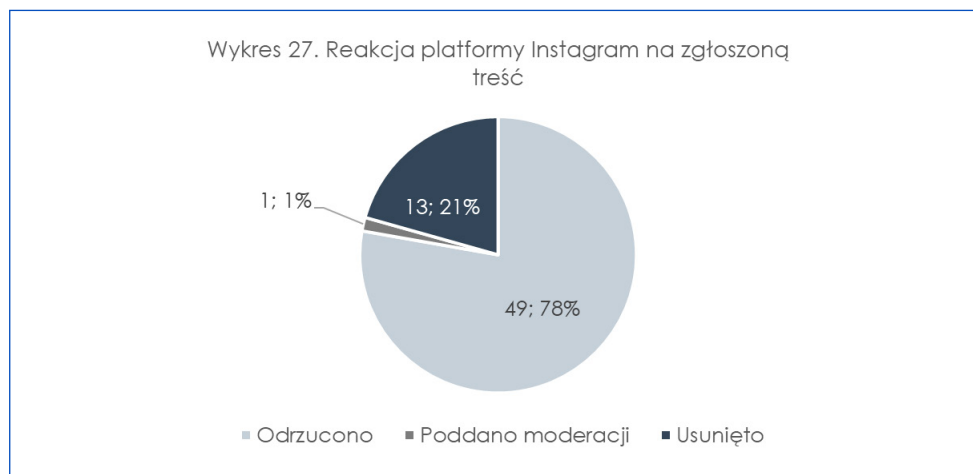
WYKRES 26. Reakcja platformy Facebook na zgłoszoną treść

ŹRÓDŁO: OPRACOWANIE WŁASNE

W analizowanym okresie platforma Facebook usunęła 36 proc. zgłoszonych treści, czyli 454 materiałów, a niecałe 4 proc. zgłoszonych treści, czyli

45 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 61 proc. zgłoszeń, czyli 769 przypadków.

Instagram



WYKRES 27. Reakcja platformy Instagram na zgłoszoną treść

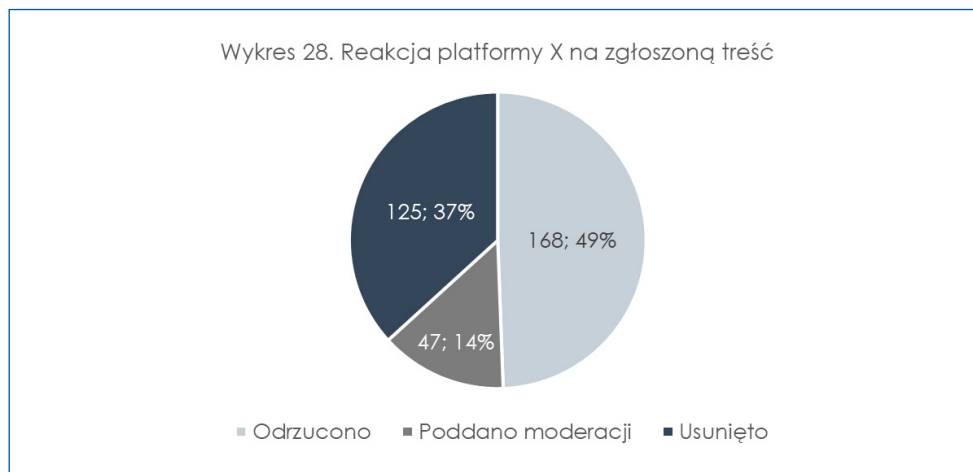
ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram usunęła 21 proc. zgłoszonych treści, czyli 13 materiałów, a mniej niż 2 proc. zgłoszonych treści, czyli 1 materiał został poddany moderacji. Jednocześnie moderatorzy odrzucili 78 proc. zgłoszeń, czyli 49 przypadków.

Threads

Do administracji platformy Threads przekazano 2 materiały, z których żaden nie został usunięty ani poddany moderacji.

X.com

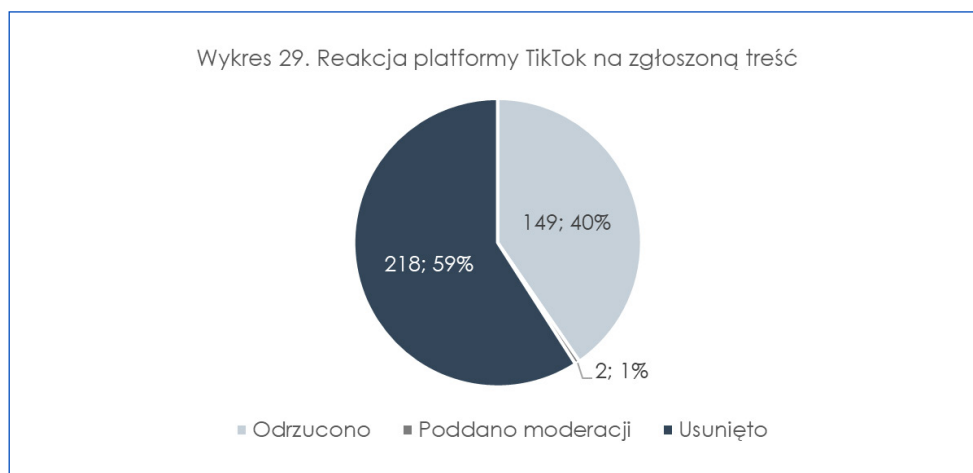


WYKRES 28. Reakcja platformy X na zgłoszoną treść

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 37 proc. zgłoszonych treści, czyli 125 materiałów, a 14 proc. zgłoszonych treści, czyli 47 materiałów zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 49 proc. zgłoszeń, czyli 168 przypadków.

TikTok

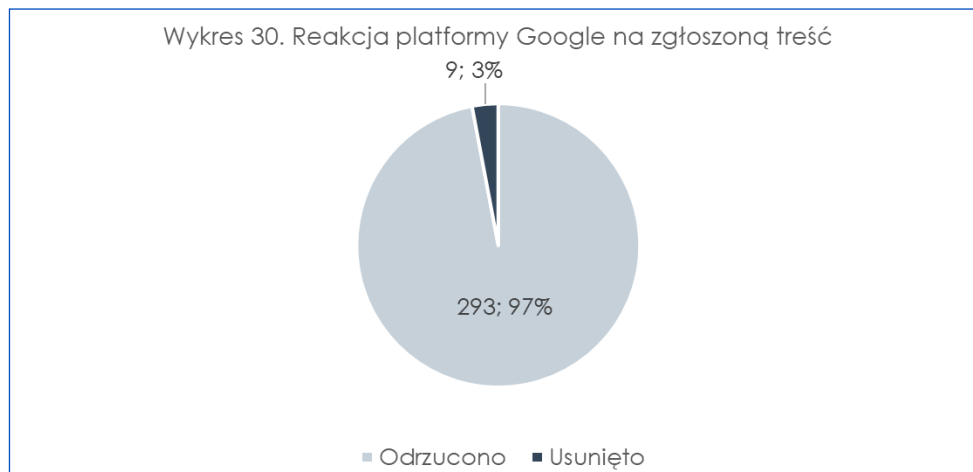


WYKRES 29. Reakcja platformy TikTok na zgłoszoną treść

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok usunęła 59 proc. zgłoszonych treści, czyli 218 materiałów, a 1 proc. zgłoszonych treści, czyli 2 materiały zostały poddane moderacji. Jednocześnie moderatorzy odrzucili 40 proc. zgłoszeń, czyli 149 przypadków.

Google (YouTube i Google Ads)



WYKRES 30. Reakcja platformy Google na zgłoszoną treść

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do firmy Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 3 proc. zgłoszonych treści, czyli 9 materiałów. Nie usunięto natomiast 97 proc. zgłoszonych treści, czyli 293 materiałów.

LinkedIn

Administracja platformy LinkedIn odrzuciła wszystkich **7 zgłoszeń**.

Discrod

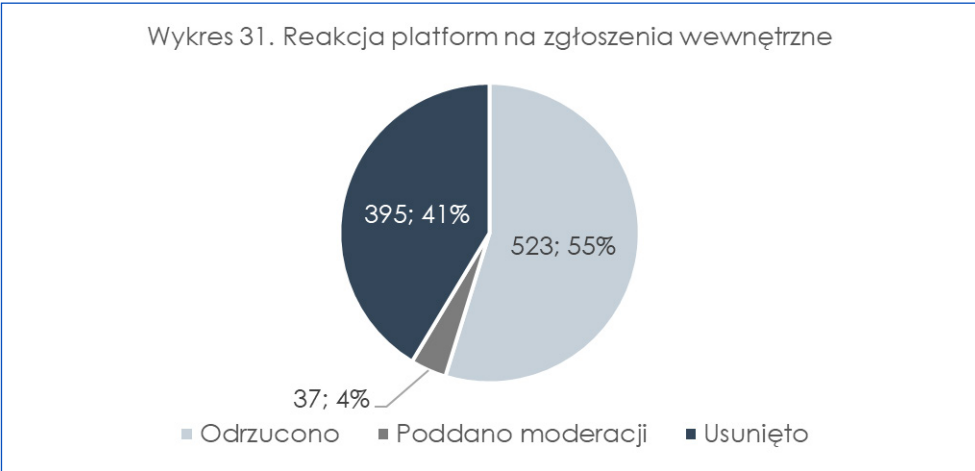
Administracja platformy Discrod odrzuciła wszystkich **14 zgłoszeń**.

Telegram

Administracja platformy Telegram odrzuciła **1 przesłane do niej zgłoszenie**.

Zgłoszenia wewnętrzne

W przypadku zgłoszeń wewnętrznych (pochodzących z bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej) w I kwartale 2025 roku z platform społecznościowych usunięto 41 proc. zgłoszonych treści, czyli 395 wpisów. Dalszych 4 proc. zgłoszonych treści, czyli 37 wpisów, zostało poddanych moderacji, natomiast 55 proc. zgłoszeń, czyli 523 przypadki, nie spotkało się z żadną widoczną reakcją ze strony moderacji.



WYKRES 31. Reakcja platform na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń usuniętych	57	88	250
Liczba zgłoszeń odrzuconych	163	179	181
Liczba zgłoszeń poddanych moderacji	13	24	0
Całkowita liczba zgłoszeń	233	291	431

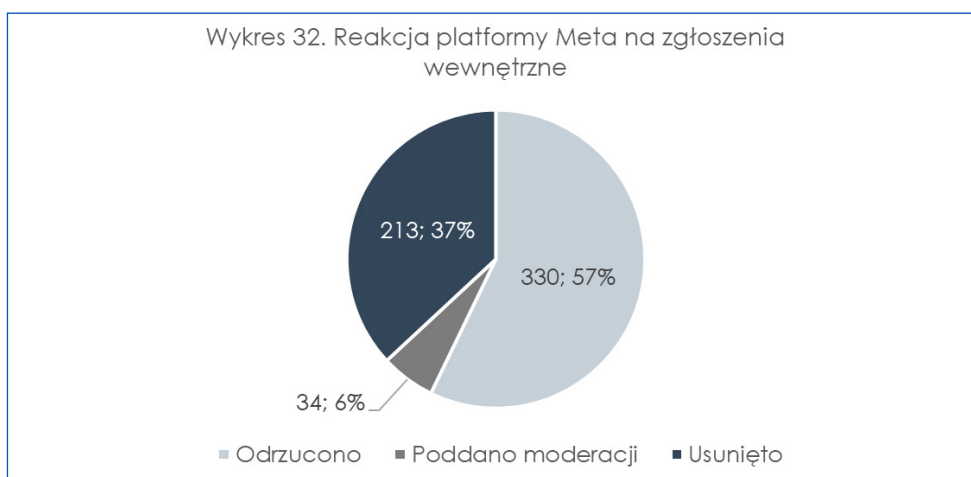
TABELA 2. Rodzaj przypisanego naruszenia w zgłoszeniach wewnętrznych

ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń⁹. Wśród zgłoszeń wewnętrznych zidentyfikowano:

- **24,4 proc. (233 przypadki)** wpisów zawierających fałszywe informacje;
- **30,5 proc. (291 przypadków)** wpisów zawierających manipulację;
- **45,1 proc. (431 przypadków)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

Meta



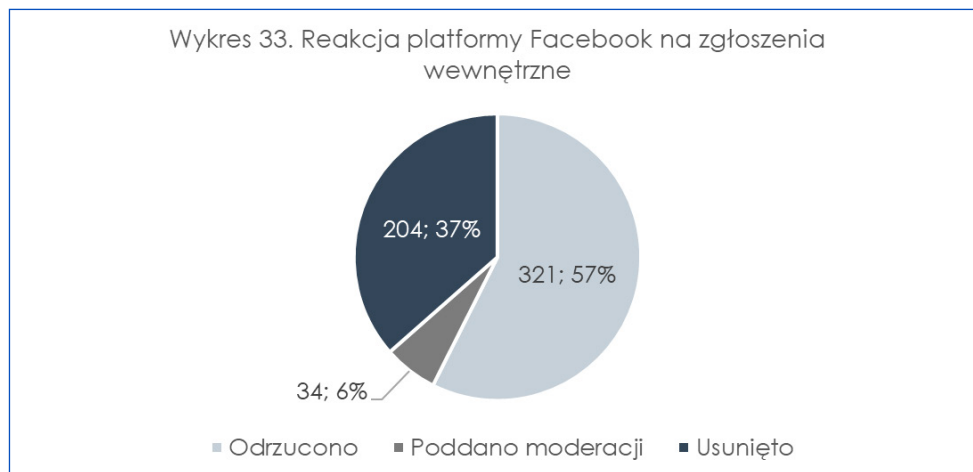
WYKRES 32. Reakcja platformy Meta na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszeń wewnętrznych przekazanych do administracji platform firmy Meta w związku z treściami pochodzącymi z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 37 proc. zgłoszonych treści, czyli 213 materiałów, 6 proc., czyli 34 materiały, poddano moderacji, a 57 proc. zgłoszeń, czyli 330 przypadków, zostało odrzuconych.

⁹ Podział na wskazane kategorie został dokonany wewnętrznie przez Zespół Szybkiego Reagowania i Wykrywania Dezinformacji w celu uproszczenia procedury realizacji zgłoszeń.

Facebook

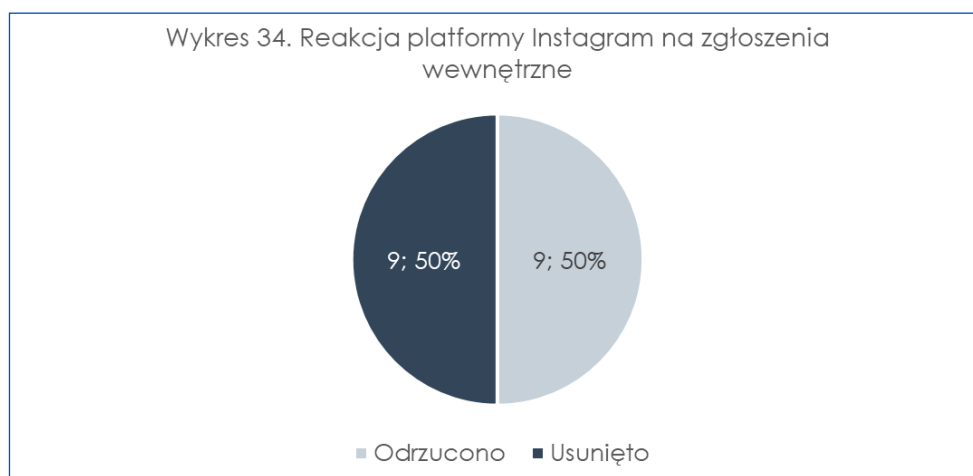


WYKRES 33. Reakcja platformy Facebook na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Facebook usunęła 37 proc. zgłoszonych treści, czyli 204 materiały, a 6 proc. zgłoszonych treści, czyli 34 materiały, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 57 proc. zgłoszeń, czyli 321 przypadków.

Instagram

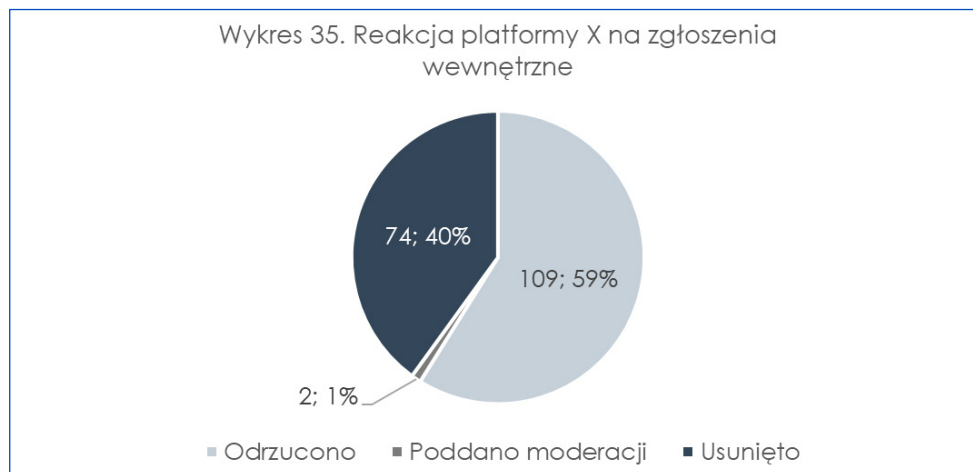


WYKRES 34. Reakcja platformy Instagram na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram usunęła 9 materiałów, czyli 50 proc. zgłoszonych treści. Pozostałych 9 zgłoszeń (50 proc) zostało odrzuconych.

X.com

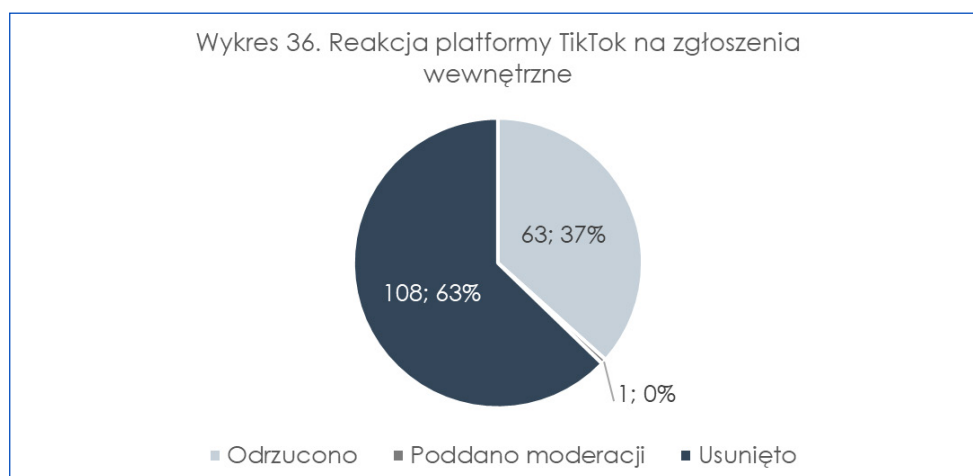


WYKRES 35. Reakcja platformy X na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 40 proc. zgłoszonych treści, czyli 74 materiały, a 1 proc. zgłoszonych treści, czyli 2 materiały zostały poddane moderacji. Jednocześnie moderatorzy odrzucili 59 proc. zgłoszeń, czyli 109 przypadków.

TikTok



WYKRES 36. Reakcja platformy TikTok na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok usunęła 63 proc. zgłoszonych treści, czyli 108 materiałów, a mniej niż 1 proc. zgłoszonych treści, czyli dokładnie 1 materiał został poddany moderacji. Jednocześnie moderatorzy odrzucili 37 proc. zgłoszeń, czyli 63 przypadki.

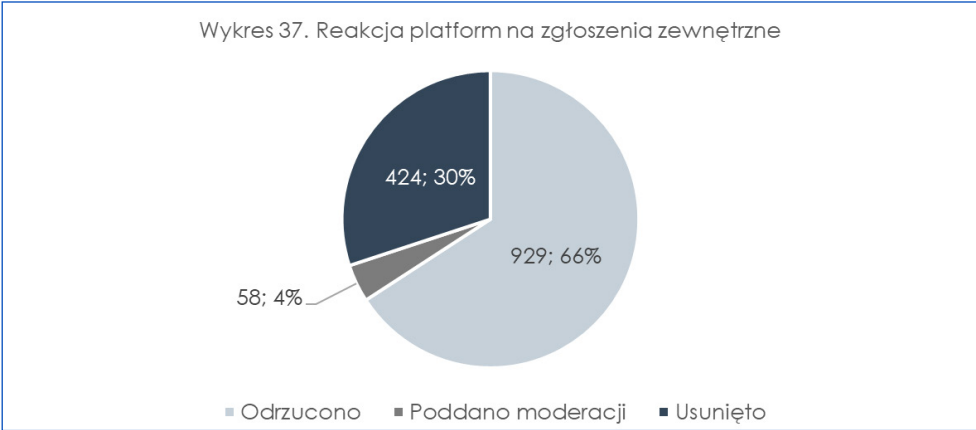
Google (YouTube i Google Ads)

Platformy Google odrzuciły wszystkich 8 zgłoszeń.

Discrod

Platforma Discrod odrzuciła wszystkich 13 zgłoszeń.

Zgłoszenia zewnętrzne



WYKRES 37. Reakcja platform na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku zgłoszeń zewnętrznych (przekazywanych do Zespołu Szybkiego Reagowania i Wykrywania Dezinformacji przez obywateli i obywatelki, podmioty publiczne oraz inne instytucje) w I kwartale 2025 roku z platform społecznościowych usunięto 30 proc. zgłoszonych treści, czyli 424 wpisy. Dalszych 4 proc. zgłoszonych treści, czyli 58 wpisów, zostało poddanych moderacji, natomiast 66 proc. zgłoszeń, czyli 929 przypadków, nie spotkało się z żadną widoczną reakcją ze strony moderacji. 604 materiały zgłoszone do OAD NASK zostały ocenione jako niezasadne, czyli takie, które nie stanowiły naruszenia regulaminów bądź innych regulacji prawnych.

Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń usuniętych	45	21	358
Liczba zgłoszeń odrzuconych	218	497	214
Liczba zgłoszeń poddanych moderacji	8	5	45
Całkowita liczba zgłoszeń	271	523	617

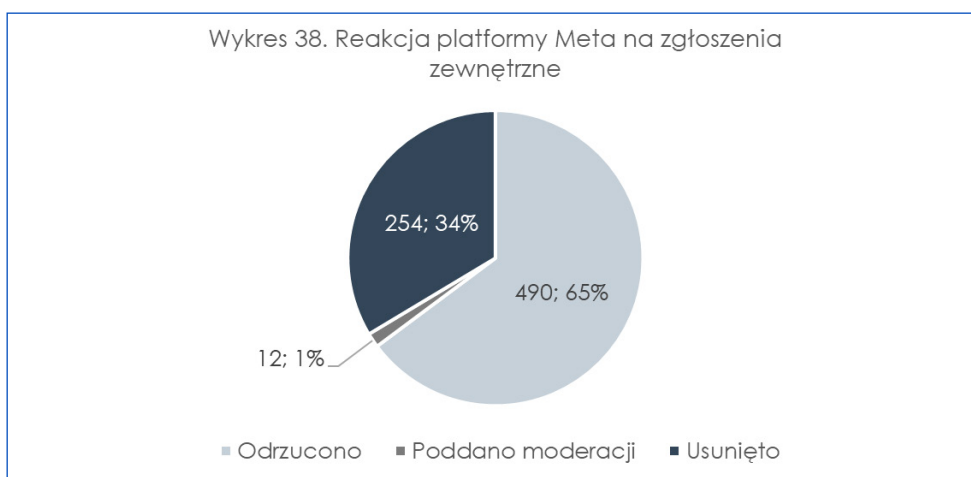
TABELA 3. Rodzaj naruszenia w zgłoszeniach zewnętrznych

ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń¹⁰. Wśród zgłoszeń zewnętrznych zidentyfikowano:

- **19 proc. (271 przypadki)** wpisów zawierających fałszywe informacje;
- **37 proc. (523 przypadki)** wpisów zawierających manipulacje;
- **44 proc. (617 przypadki)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

Meta



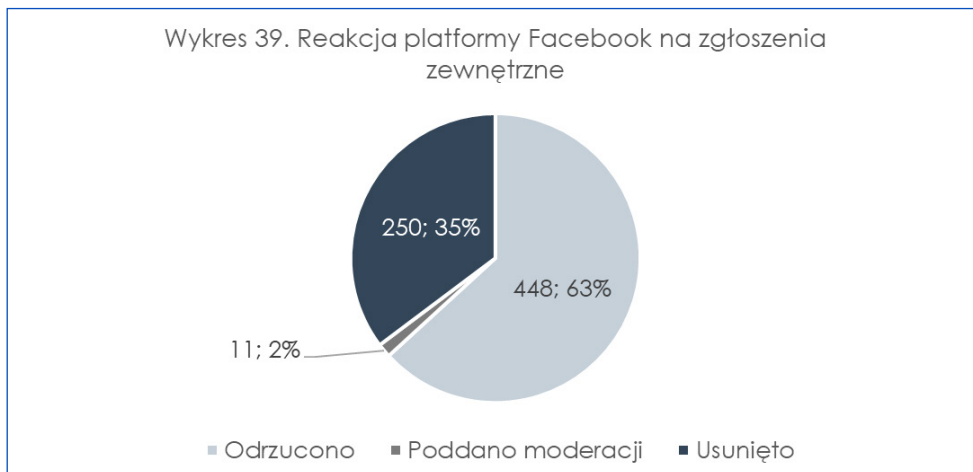
WYKRES 38. Reakcja platformy Meta na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszeń zewnętrznych przekazanych do administracji platform firmy Meta w związku z treściami pochodzącymi z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 34 proc. zgłoszonych treści, czyli 254 materiały, 2 proc., czyli 12 materiałów, poddano moderacji, a 65 proc. zgłoszeń, czyli 490 przypadków, zostało odrzuconych.

¹⁰ Podział na wskazane kategorie został dokonany wewnętrznie przez Zespół Szybkiego Reagowania i Wykrywania Dezinformacji w celu uproszczenia procedury realizacji zgłoszeń.

Facebook

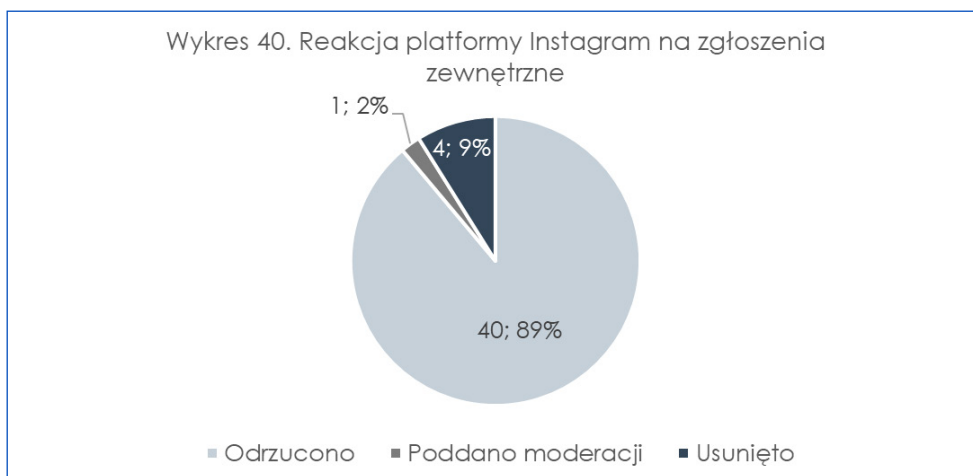


WYKRES 39. Reakcja platformy Facebook na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Facebook usunęła 35 proc. zgłoszonych treści, czyli 250 materiałów, a 2 proc. zgłoszonych treści, czyli 11 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 63 proc. zgłoszeń, czyli 448 przypadki.

Instagram



WYKRES 40. Reakcja platformy Instagram na zgłoszenia zewnętrzne

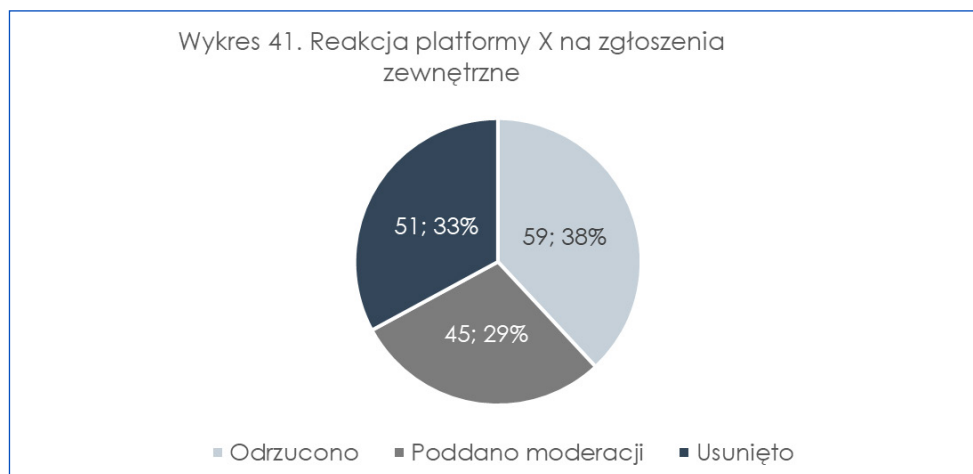
ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram usunęła 9 proc. zgłoszonych treści, czyli 4 materiały, a 2 proc. zgłoszonych treści, czyli 1 materiał został poddany moderacji. Jednocześnie moderatorzy odrzucili 89 proc. zgłoszeń, czyli 40 przypadków.

Threads

Platforma Threads odrzuciła wszystkie 2 zgłoszenia.

X.com



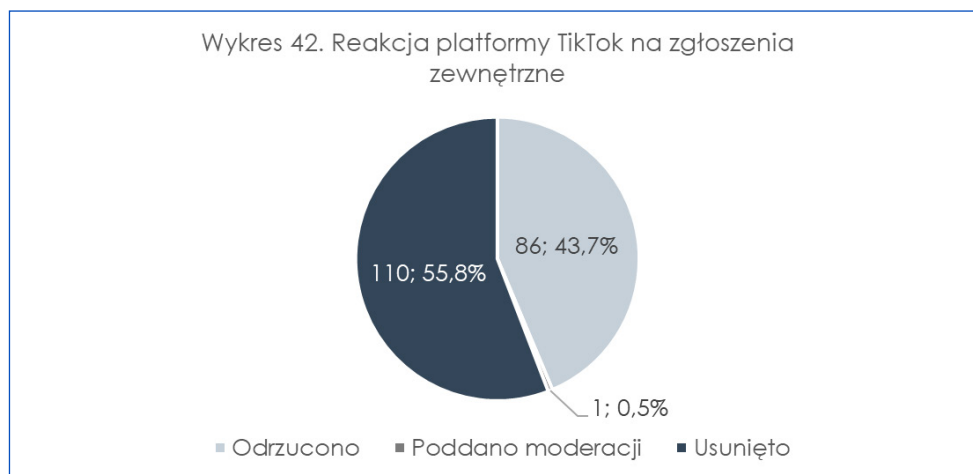
WYKRES 41.

Reakcja platformy X na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 33 proc. zgłoszonych treści, czyli 51 materiałów, a 29 proc. zgłoszonych treści, czyli 45 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 38 proc. zgłoszeń, czyli 59 przypadków.

TikTok



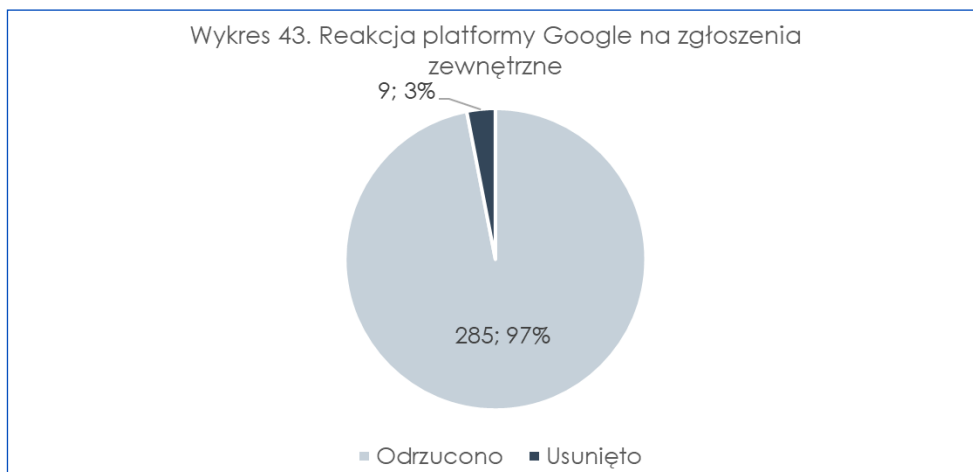
WYKRES 42.

Reakcja platformy TikTok na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok usunęła 56 proc. zgłoszonych treści, czyli 110 materiałów, a mniej niż 1 proc. zgłoszonych treści, czyli 1 materiał został poddany moderacji. Jednocześnie moderatorzy odrzucili 44 proc. zgłoszeń, czyli 86 przypadków.

Google (YouTube i Google Ads)



WYKRES 43.

Reakcja platformy Google na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 3 proc. zgłoszonych treści, czyli 9 materiałów. Nie usunięto natomiast 97 proc. zgłoszonych treści, czyli 285 materiałów.

LinkedIn

Administracja platformy LinkedIn odrzuciła wszystkich 7 zgłoszeń.

Discrod

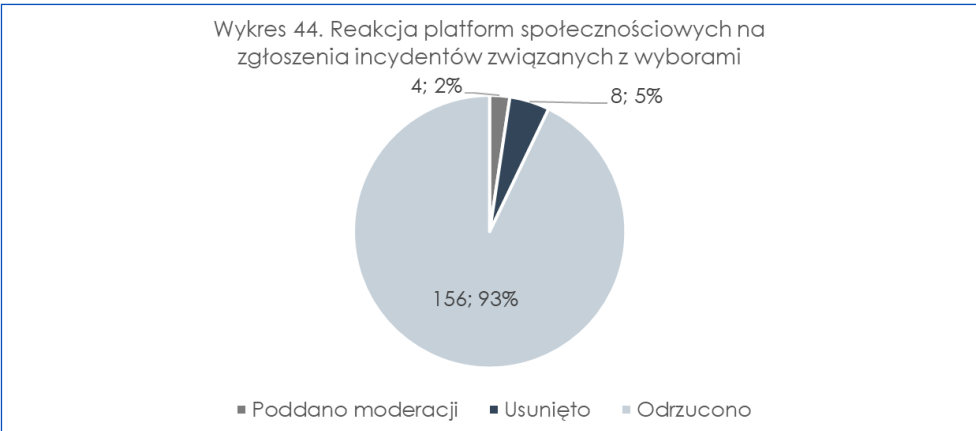
Administracja platformy Discrod odrzuciła 1 przesłane do niej zgłoszenie.

Telegram

Administracja platformy Telegram odrzuciła 1 przesłane do niej zgłoszenie.

Case studies

Zgłoszenia dotyczące kampanii wyborczej 2025



WYKRES 44. Reakcja platform społecznościowych na zgłoszenia incydentów związanych z wyborami
ŹRÓDŁO: OPRACOWANIE WŁASNE

Od początku 2025 roku OAD NASK prowadził dedykowany monitoring przestrzeni informacyjnej w związku wyborami prezydenckimi. W I kwartale do platform społecznościowych przekazano zgłoszenia dotyczące 168 materiałów. 8 z nich (5 proc.) zostało usuniętych, z kolei 4 (2 proc.) zostały poddane moderacji. 156 zgłoszeń (93 proc.) zostało odrzuconych¹¹.

Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/institucję
Liczba zgłoszeń usuniętych	2	3	3
Liczba zgłoszeń odrzuconych	104	51	1
Liczba zgłoszeń poddanych moderacji	2	1	1
Całkowita liczba zgłoszeń	108	55	5

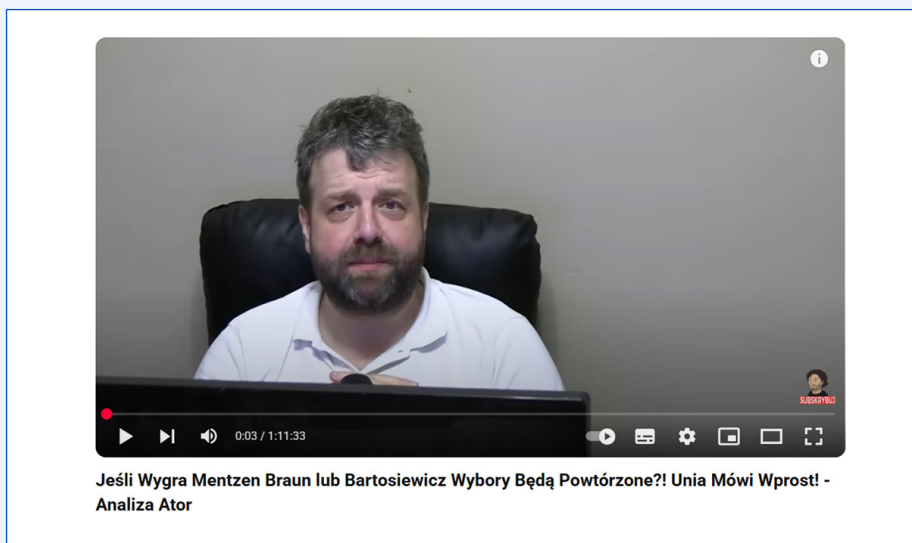
TABELA 4. Rodzaj naruszenia w zgłoszeniach zewnętrznych
ŹRÓDŁO: OPRACOWANIE WŁASNE

¹¹ Więcej informacji w raporcie: <https://www.gov.pl/web/cyfryzacja/raport-z-programu-parasol-wyborczy-kompleksowa-ochrona-wyborow-prezydenckich-2025> (dostęp: 26.01.2026 godz. 10:50)

Przykłady zgłoszonej dezinformacji wyborczej

Wśród zgłaszanych do platform społecznościowych materiałów znalazły się treści zawierające twierdzenia o rzekomej ingerencji Unii Europejskiej i państw sojusznicznych w proces wyborczy. Zgłaszano również wpisy, które sugerowały, że w przypadku wygranej lub przegranej określonego kandydata (występowały różne konfiguracje) wybory zostaną unieważnione lub powtórzone ze względu na oskarżenia o fałszowanie głosów. W badanym okresie pojawiały się także przykłady wykorzystywania wizerunku kandydatów w materiałach typu deepfake.

Opisane powyżej przypadki, mimo wielokrotnego ich zgłaszania, nie spotkały się z żadną reakcją zespołów odpowiedzialnych za moderację treści. Warto zaznaczyć, że brak reakcji platform społecznościowych na zgłoszone materiały sprawił, że dotarły one do szerokiego grona użytkowników.



OBRAZ 3.

Nagranie zawierające fałszywe informacje i manipulacje (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM



OBRAZ 4. Wpis zawierający fałszywe informacje (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



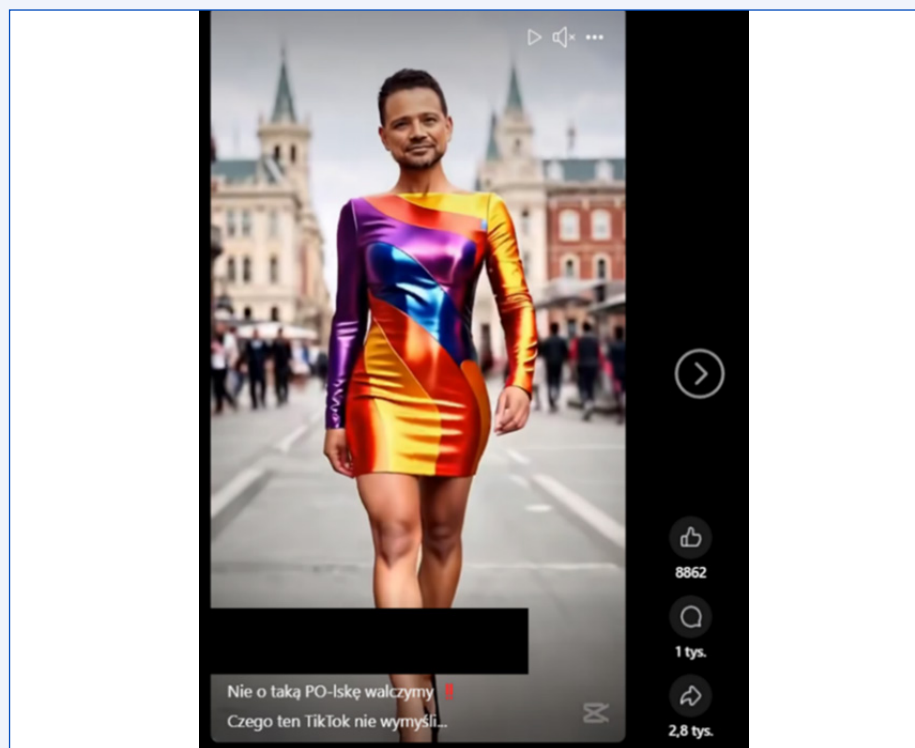
OBRAZ 5. Nagranie zawierające fałszywe informacje (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM



OBRAZ 6. Wpis zawierający fałszywe informacje (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM



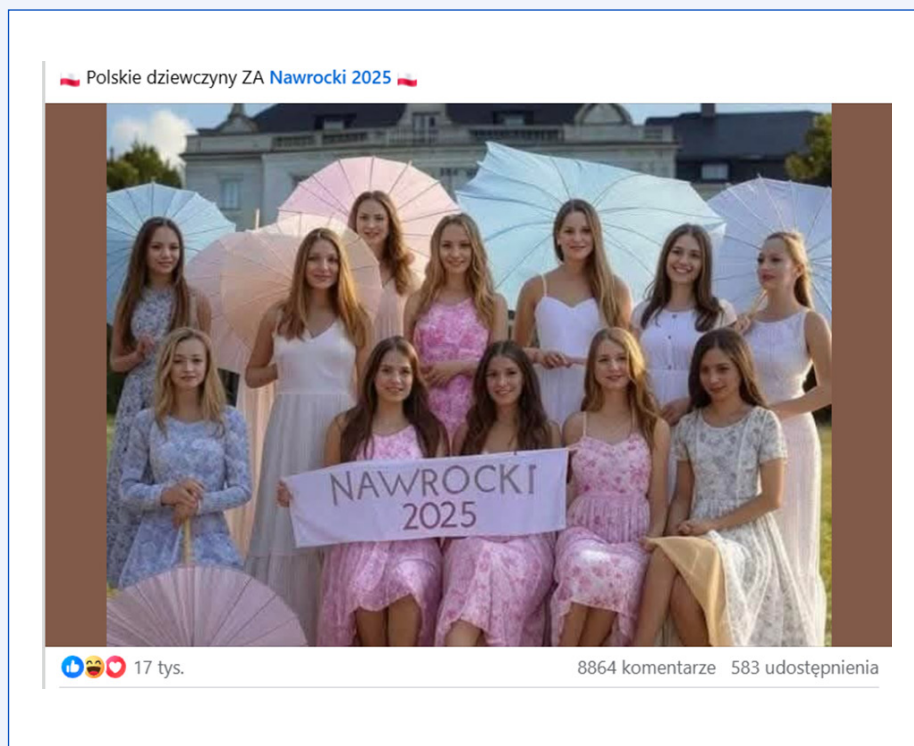
OBRAZ 7. Deepfake z udziałem kandydata (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



OBRAZ 8. Wpis zawierający fałszywe informacje (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM



OBRAZ 9. Nieoznaczony materiał wygenerowany przez AI (zgłoszenie odrzucone)

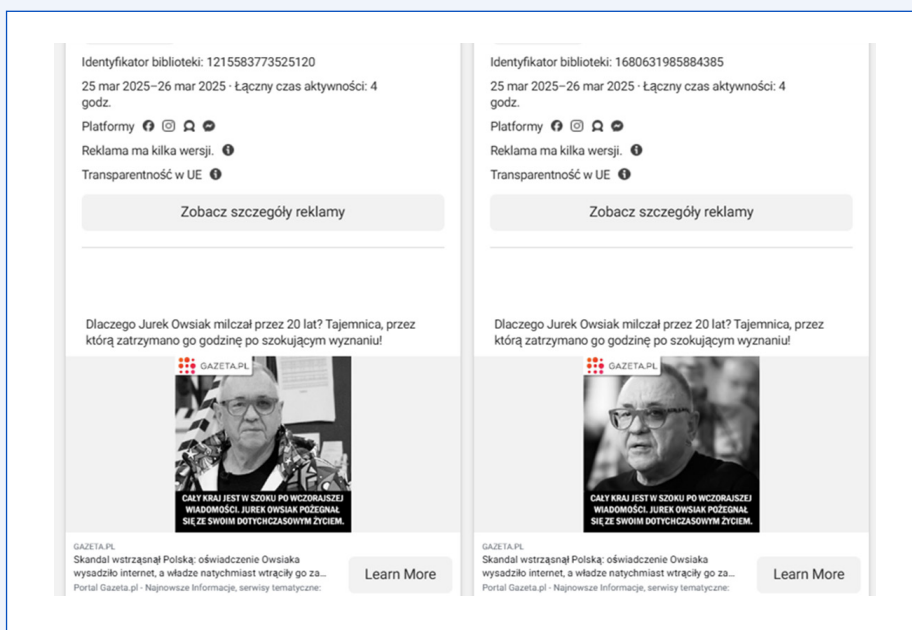
ŹRÓDŁO: FACEBOOK.COM

Reklamy z wizerunkiem osób publicznych

Jedną z regularnie zgłaszanych kategorii treści są materiały reklamowe wykorzystujące wizerunek osób publicznych. Ich celem jest nakłonienie użytkowników do odwiedzenia zewnętrznych witryn, gdzie najczęściej dochodzi do prób wyłudzenia danych lub środków finansowych. W niektórych przypadkach, oprócz oczywistego oszustwa, materiały te mogą też manipulować odbiorcami i potencjalnie wpływać na zmianę ich poglądów.

W przypadku platformy Meta pojedyncze reklamy były stosunkowo szybko usuwane przez moderatorów. Kluczowym problemem było jednak to, że działania administracji ograniczały się głównie do blokowania konkretnych treści, a nie całych profili reklamodawców. W rezultacie osoby odpowiedzialne za publikowanie takich reklam mają możliwość zamieszczania kolejnych szkodliwych treści bez ograniczeń i korzystają z niej. Decyzja platformy usunięcia pojedynczej reklamy jest traktowana jako zgłoszenie „poddane moderacji”.

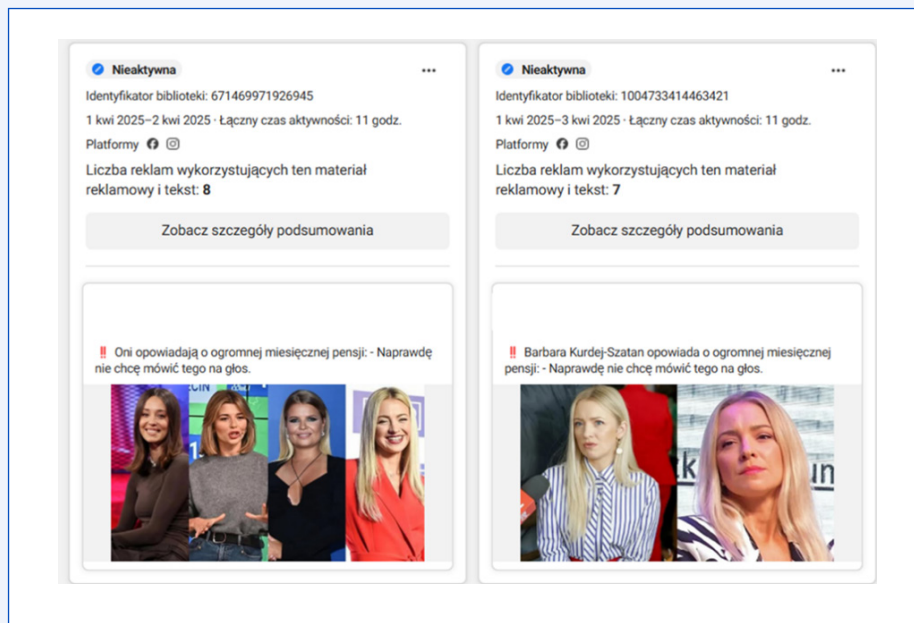
Kilka pojedynczych profili reklamowych zostało zablokowanych. Była to jednak reakcja na eskalację zgłoszeń przekazanych przez OAD NASK bezpośrednio do punktu kontaktowego, a nie efekt działania moderacji w ramach standardowej realizacji zgłoszenia.



OBRAZ 10.

Przykłady reklam zamieszczanych przez zgłoszony profil (zgłoszenie poddano moderacji, to jest: usunięte zostały poszczególne reklamy tego reklamodawcy)

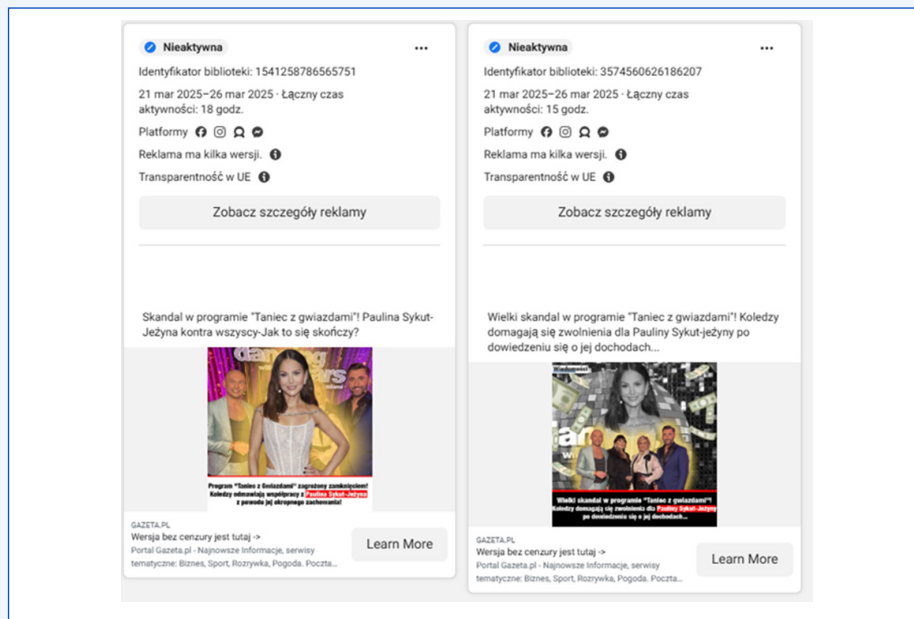
ŹRÓDŁO: BIBLIOTEKA REKLAM META



OBRAZ 11.

Przykłady reklam zamieszczanych przez zgłoszony profil (zgłoszenie poddano moderacji, to jest: usunięte zostały pojedyncze reklamy)

ŹRÓDŁO: BIBLIOTEKA REKLAM META



OBRAZ 12.

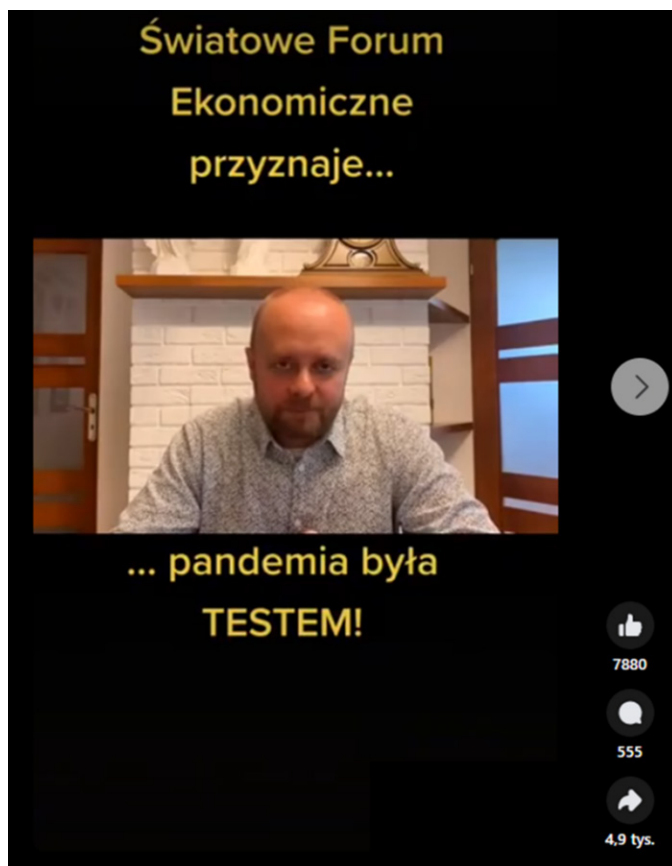
Przykłady reklam zamieszczanych przez zgłoszony profil (zgłoszenie poddano moderacji, to jest: usunięte zostały pojedyncze reklamy)

ŹRÓDŁO: BIBLIOTEKA REKLAM META

Szkodliwe treści medyczne

Wśród zgłaszanych na przestrzeni I kwartału 2025 roku treści znalazły się materiały zawierające szkodliwe i wprowadzające w błąd porady medyczne. W polskojęzycznej infosferze funkcjonują profile, które skupiają się na promowaniu wyłącznie tego typu treści. Nierzadko uzyskiwana w ten sposób popularność jest monetyzowana – twórcy rekomendują na przykład zażywanie określonych specyfików, na sprzedaży których zarabiają.

W mediach społecznościowych nadal utrzymuje się także zainteresowanie tematem pandemii COVID-19. Aktorzy dezinformacji powielają teorię spiskową, według której pandemia była testem zaplanowanym przez światowe elity, w tym Światowe Forum Ekonomiczne, a szczepienia, które mają ochronić przed chorobą, tak naprawdę prowadzą do śmierci.

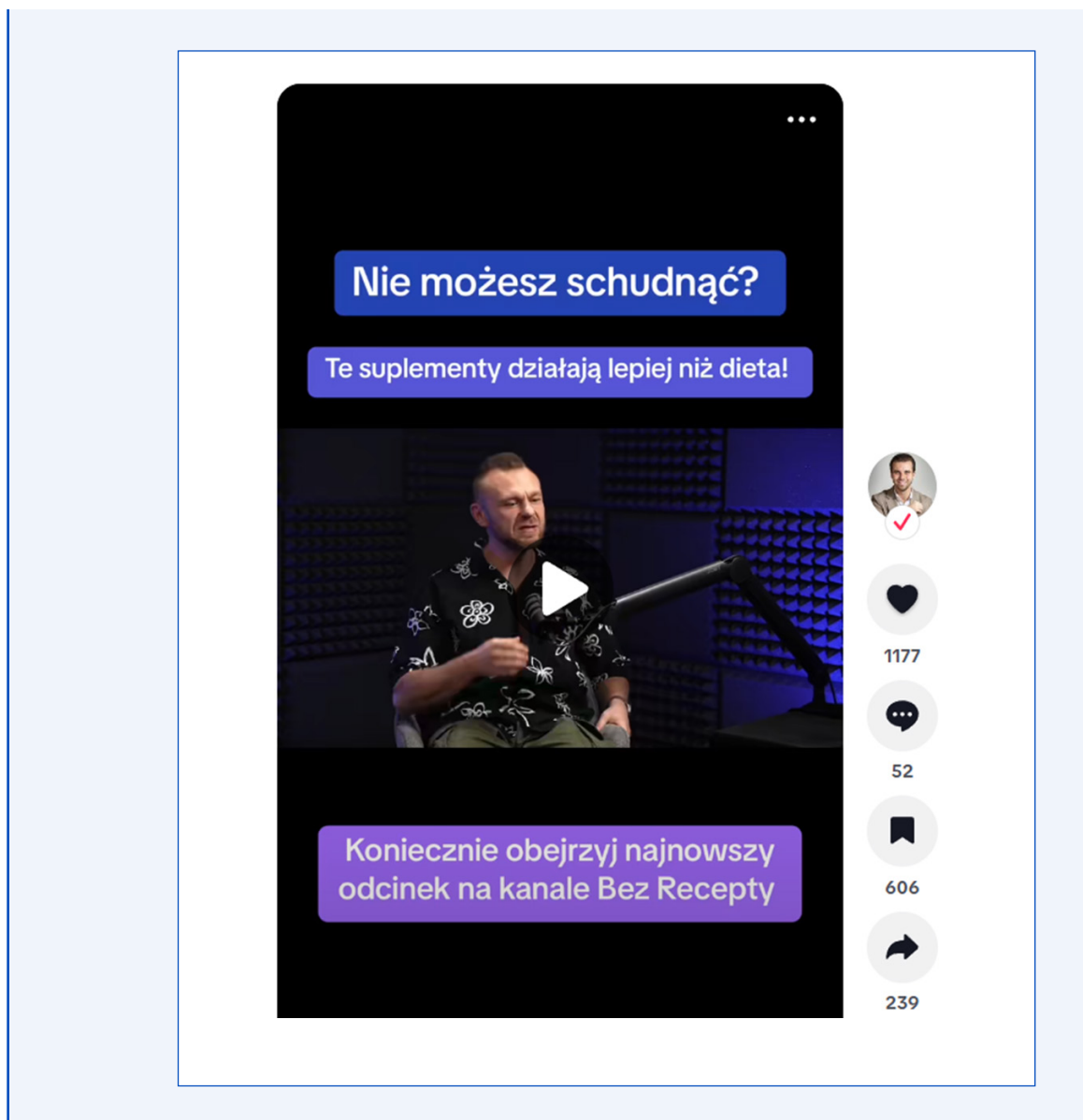


OBRAZ 13.

Materiał dotyczący pandemii COVID-19 (zgłoszenie odrzucone)

ŹRÓDŁO: TIKTOK.COM

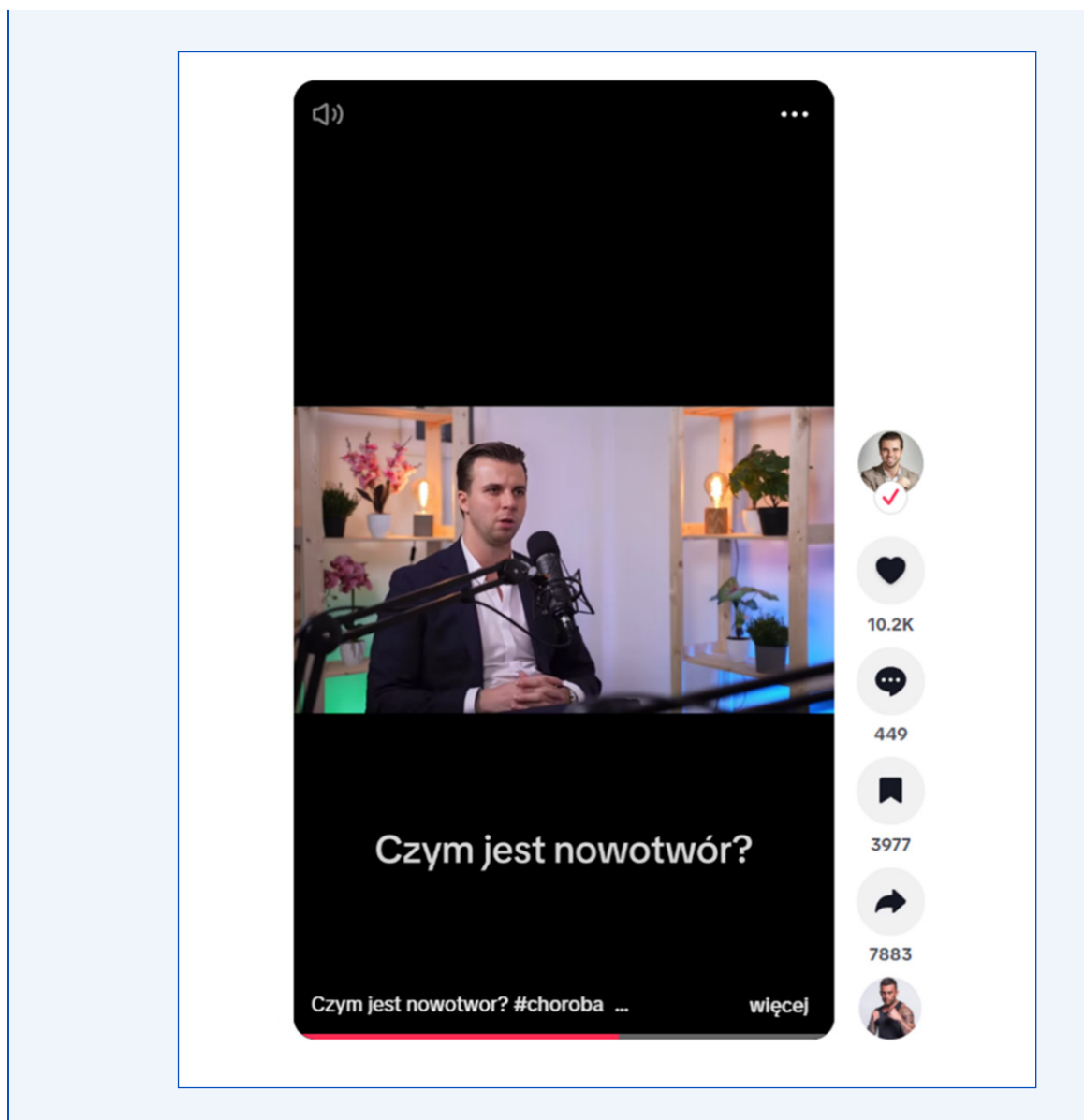
W innych zgłoszonych materiałach samozwańczy eksperci rekomendowali stosowanie takich metod radzenia sobie z chorobami jak spożywanie ziemi czy przyjmowanie ogromnych dawek witaminy D3 oraz soli. Platformy społecznościowe uznały jednak, że tego typu treści nie naruszają ich wewnętrznych regulaminów, dlatego wiele wpisów zawierających szkodliwe narracje wciąż pozostaje dostępnych.



OBRAZ 14.

Materiał zawierający rekomendację dotyczącą spożywania ziemi (zgłoszenie odrzucone)

ŹRÓDŁO: TIKTOK.COM

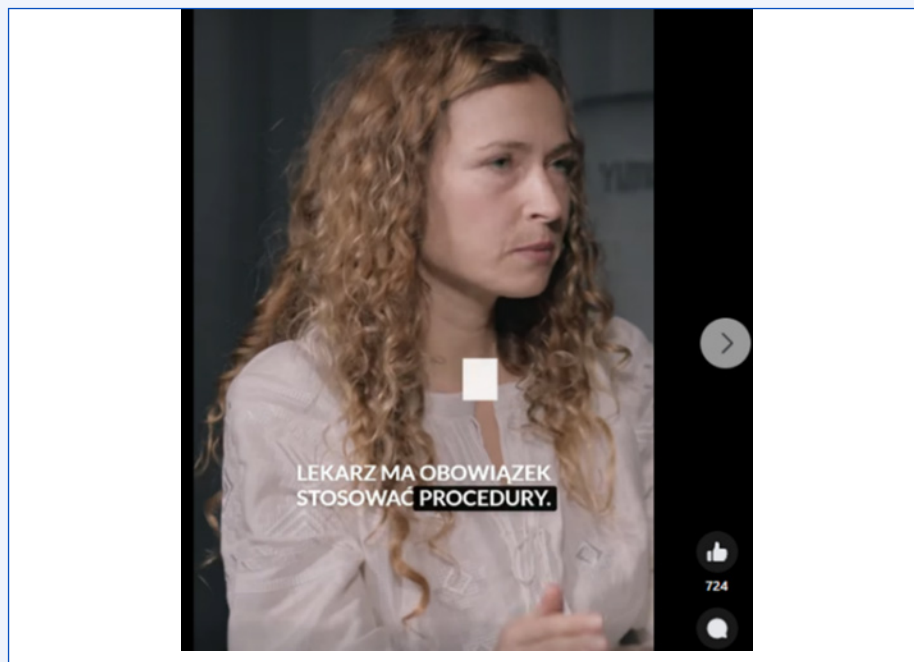


OBRAZ 15.

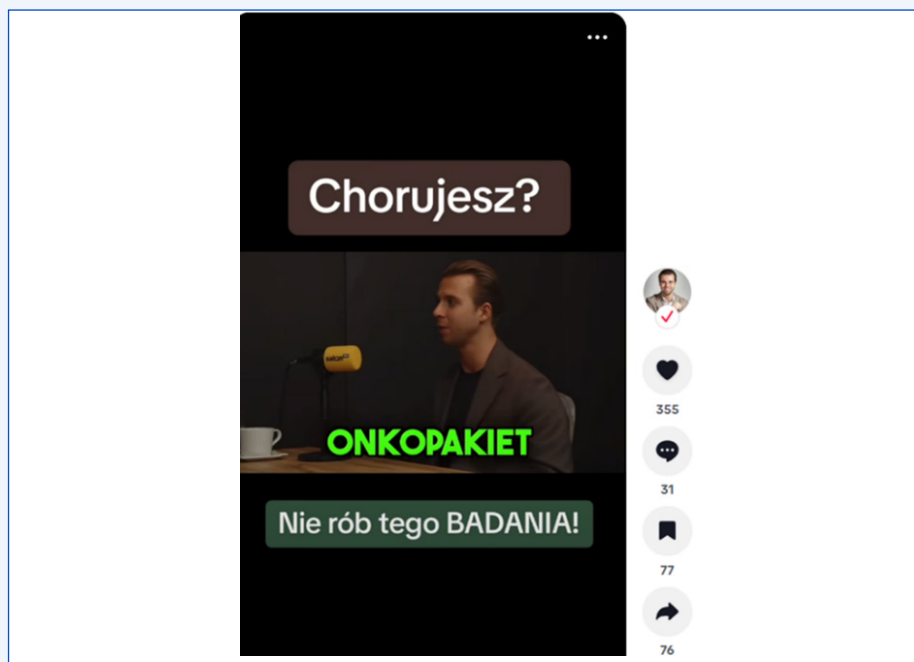
Materiał zawierający rekomendację przedawkowywania witaminy D3 (zgłoszenie odrzucone)

ŹRÓDŁO: TIKTOK.COM

Wśród zgłoszonych treści wystąpiły także próby przekonania obserwujących, że nie warto korzystać z diagnostyki i procedur medycznych oferowanych przez polski system zdrowia. Przekazane do administracji platform społecznościowych wpisy zawierały tezę, że badania wykonane w ramach onkopakiety nie pozwalają na wykrycie nowotworu. Ponadto zgłoszono nagranie, w którym pada teza, że nie warto korzystać z indukcji porodu, ponieważ skurcze wywołane przez podanie oksytocyny szkodzą dziecku.



OBRAZ 16. Materiał zawierający nieprawdziwe informacje na temat indukcji porodu (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM



OBRAZ 17. Materiał zawierający nieprawdziwe informacje na temat onkopakiету (zgłoszenie odrzucone)
ŹRÓDŁO: TIKTOK.COM

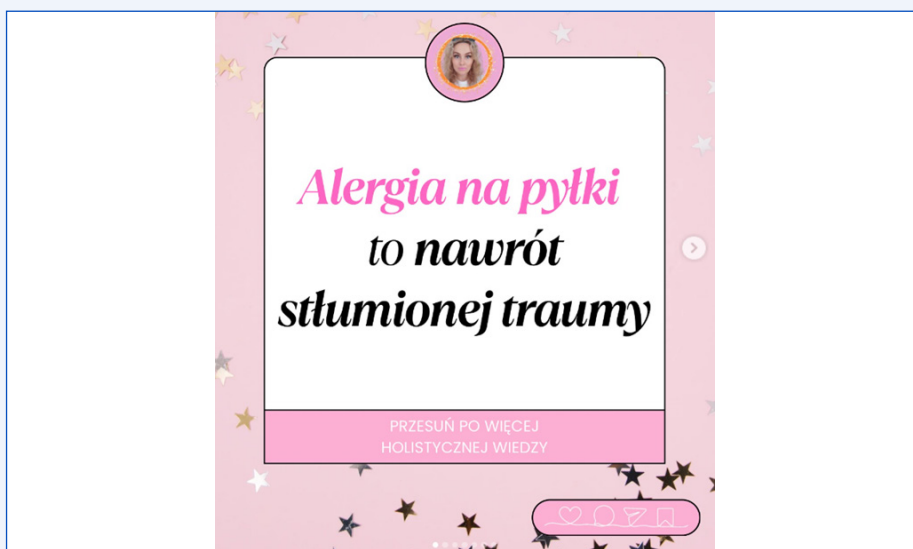
W polskojęzycznej infosferze pojawiały się także zmanipulowane materiały wskazujące traumę i problemy emocjonalne jako przyczyny wszelkich chorób. Zgodnie z nimi rozdzielenie z ojcem ma prowadzić do uczulenia na słońce, a alergia na pyłki to rzekomo nawrót traumy. Materiały zawierające przedstawione narracje zostały odrzucone przez platformy społecznościowe.



OBRAZ 18.

Wpis zawierający szkodliwą treść medyczną (zgłoszenie odrzucone)

ŹRÓDŁO: [INSTAGRAM.COM](https://www.instagram.com)



OBRAZ 19.

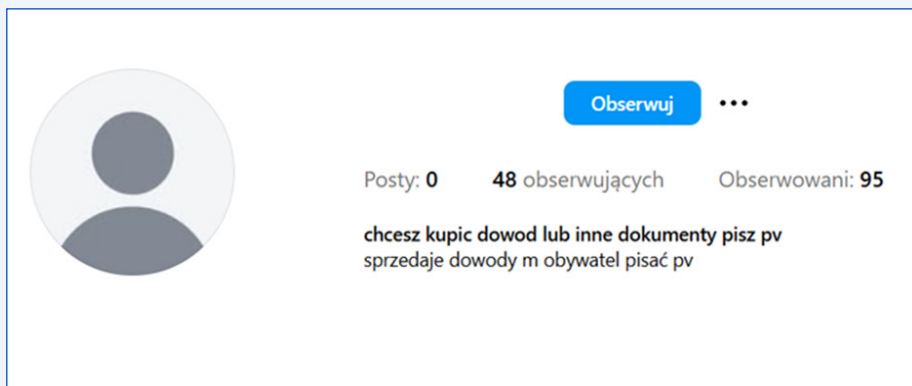
Wpis zawierający szkodliwą treść medyczną (zgłoszenie odrzucone)

ŹRÓDŁO: [INSTAGRAM.COM](https://www.instagram.com)

Oferta sprzedaży fałszywej aplikacji mObywatel

Na początku stycznia 2025 roku w mediach społecznościowych pojawiły się oferty sprzedaży fałszywej wersji aplikacji mObywatel, która do złudzenia przypominała oficjalne oprogramowanie.

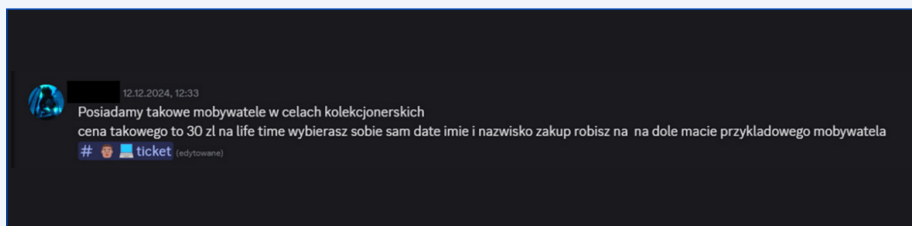
Najwięcej tego typu treści było dostępnych na platformie TikTok. 93 ze 105 zgłoszonych do serwisu materiałów zostało usuniętych (89 proc.). Cyberprzestępcy do promowania fałszywych ofert wykorzystali także platformę Discrod. W przypadku tego serwisu wiadomości zachęcające do zakupu „kolekcyjnej” wersji aplikacji pojawiały się na zamkniętych serwerach. Administracja platformy Discrod zdecydowała o odrzuceniu zgłoszeń wszystkich 13 kanałów. Kampania obecna była również na platformie Instagram, gdzie nadal znaleźć można konta użytkowników, którzy proponują zakup fałszywej wersji mObywatela. Zablokowanych zostało 9 z 13 zgłoszonych profili (69 proc.), co oznacza, że przynajmniej 4 konta pozostają aktywne.



OBRAZ 20.

Konto oferujące sprzedaż aplikacji mObywatel (zgłoszenie odrzucone)

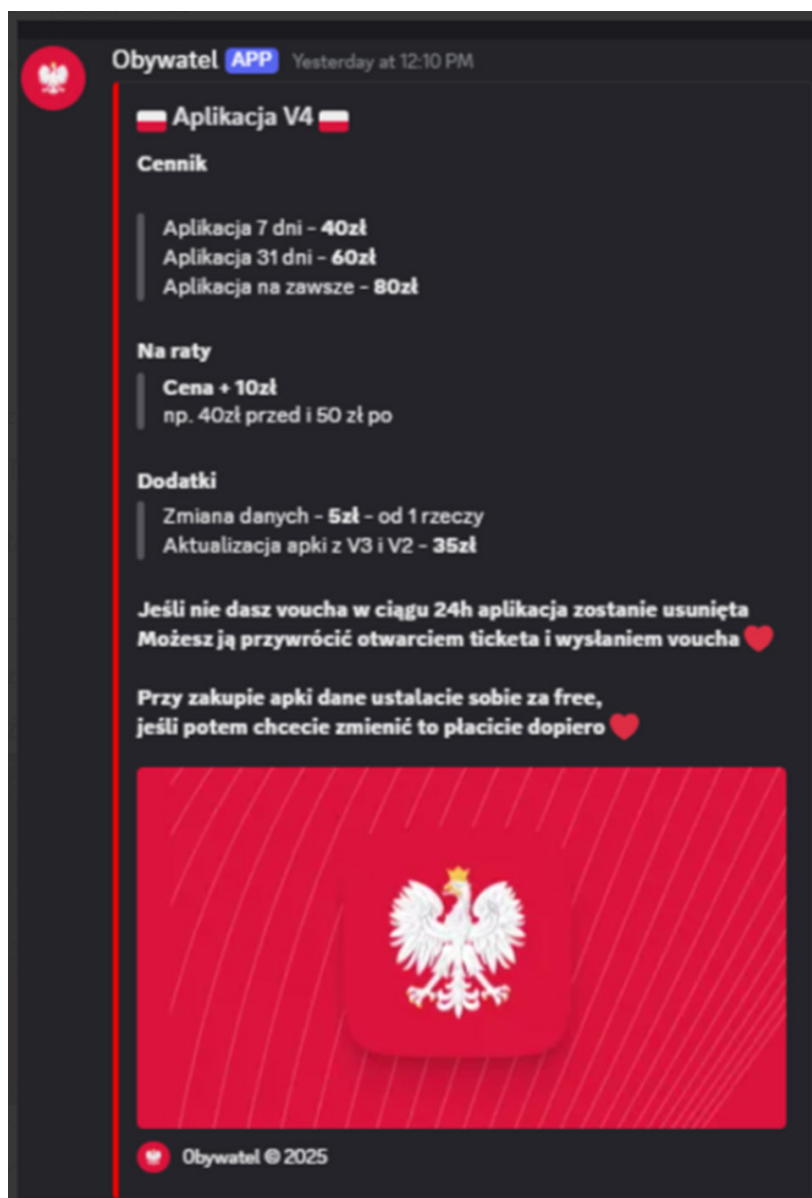
ŹRÓDŁO: [INSTAGRAM.COM](https://www.instagram.com)



OBRAZ 21.

Konto oferujące sprzedaż aplikacji mObywatel (zgłoszenie odrzucone)

ŹRÓDŁO: [DISCORD.COM](https://www.discord.com)



OBRAZ 22.

Konto oferujące sprzedaż aplikacji mObywatel (zgłoszenie odrzucone)

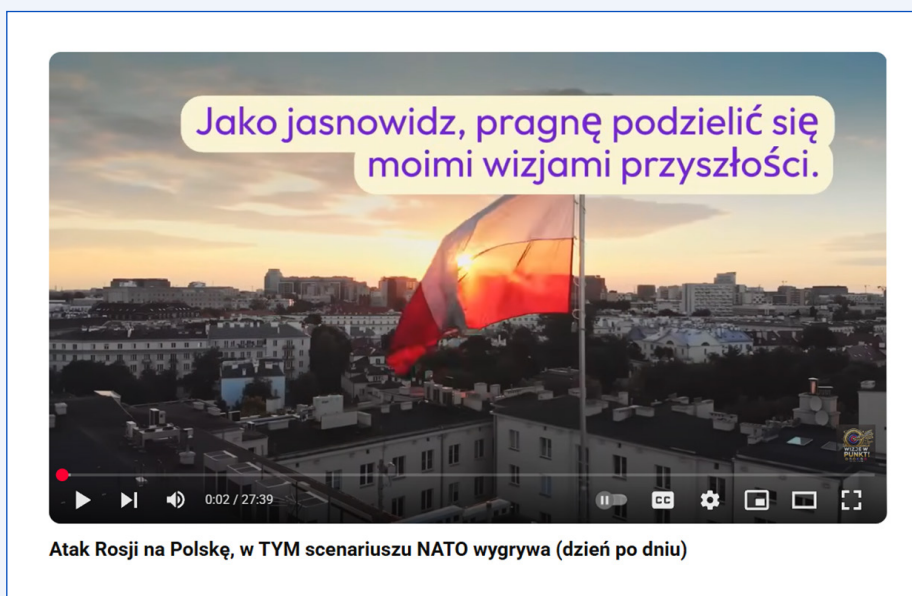
ŹRÓDŁO: DISCROD.COM

Kanały aktywnie publikujące treści dezinformacyjne

Na platformie YouTube aktywni są aktorzy dezinformacji specjalizujący się w promowaniu szokujących treści, których celem jest wzbudzenie strachu w społeczeństwie, podważanie zaufania do instytucji publicznych czy osłabienie sojuszy międzynarodowych. Twórcy posługują się takimi technikami manipulacji jak wykorzystywanie clickbaitowych, krzykliwych nagłówków czy formułowanie przekazu w sposób celowo nacechowany emocjonalnie.

Wśród materiałów zgłoszonych do administracji platformy znalazły się nagrania samozwańczych jasnowidzów, którzy dzielili się z widzami swoimi, zazwyczaj niepokojącymi, wizjami. Twórcy poruszali w nich przede wszystkim takie tematy jak rzekomo zbliżająca się wielkimi krokami wojna oraz możliwość opuszczenia Polski przez NATO w razie wybuchu konfliktu z Rosją.

Autorzy tego typu materiałów próbują obejść regulamin platformy, oznaczając treść jako satyrę lub część kreacji artystycznej. Dodanie takiego kontekstu do publikacji sprawia, że administracja serwisu może zrobić wyjątek i potraktować treść jako nienaruszającą zasad społeczności¹².



OBRAZ 23.

Materiał zawierający niepokojące wizje kreowane przez jasnowidza (zgłoszenie odrzucone)

ŹRÓDŁO: [YOUTUBE.COM](https://www.youtube.com)

¹² <https://support.google.com/youtube/answer/6345162?sjid=8111128354612446914-EU#zippy=%2Cexamples-of-basic-facts%2Cexamples-of-condemnation-opposing-views-or-satire%2Cexamples-of-discouraging-dangerous-behavior> (dostęp: 17.02.2026, godz. 14:15)

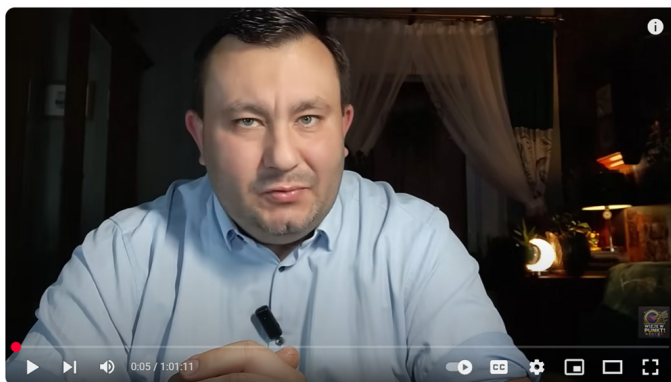


Czy wchodząc do Polski Putin powoła się na Ustawę o bratniej pomocy nr 1066? Ustawa 1066. Kazachstan

OBRAZ 24.

Materiał zawierający niepokojące wizje kreowane przez jasnowidza (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM



🌊 Wielka Fala 🇷🇺 Ruszy W Kierunku Europy. Co Z Polską? 🔥 Tajemnicze Wizje Ojca Klimuszki O Polsce

OBRAZ 25.

Materiał zawierający niepokojące wizje kreowane przez jasnowidza (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM

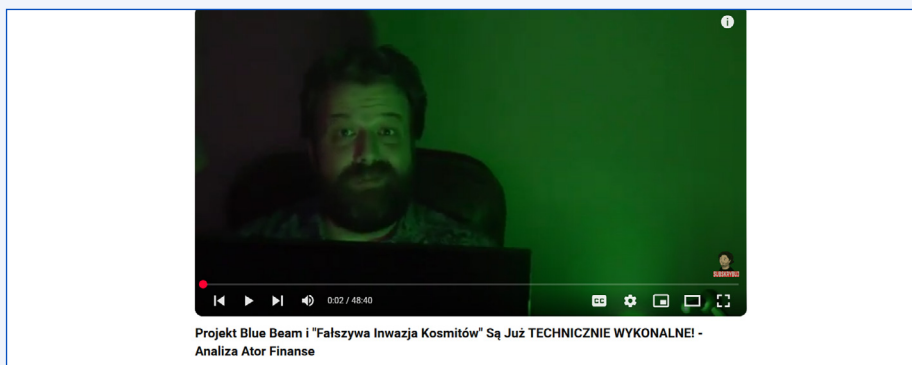
ABY NIE ZAMKNĘLI MI KANAŁU: TO DOWCIP, TO KAWAŁ TEN FILM! SAME TEORIE SPISKOWE I HIPOTEZY, NIE NALEŻY NA POWAŻNIE TEGO FILMU BRAĆ.
TEN FILM JEST W PEŁNI W KONWENCJI TEORII SPISKOWYCH JAK KAŻDY NA MOIM KANAŁE.
INFORMUJĘ, ŻE WSZYSTKO, CO USŁYSZycIE NA TYM FILMIE MA CHARAKTER EKSPRESJI ARTYSTYCZNEJ I OPARTE ZOSTAŁO NA PARAPSYCHOLOGICZNYM PODEJŚCIU I ŹRÓDLE. NIE ZASTĘPUJE ON RACJONALNEGO MYŚLENIA. TO ZABAWA.

OBRAZ 26.

Informacja zamieszczona przez jednego z aktorów dezinformacji

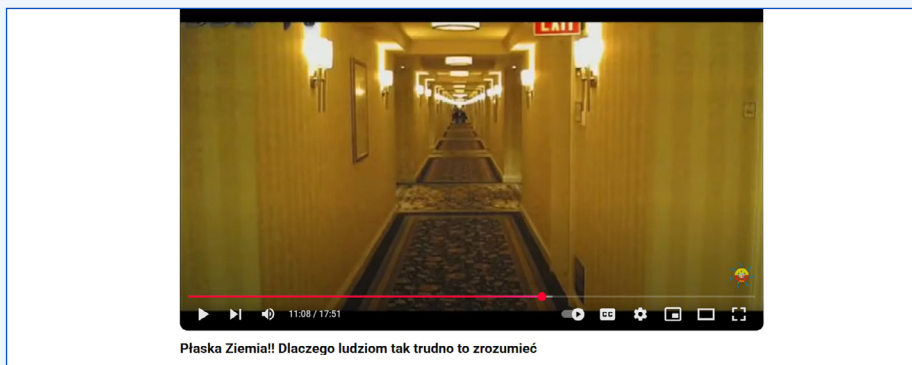
ŹRÓDŁO: YOUTUBE.COM

Na platformie YouTube popularnością cieszą się również materiały rozpowszechniające teorie spiskowe. Do administracji serwisu zgłoszono treści promujące między innymi narrację o istnieniu płaskiej Ziemi czy projekcie Blue Beam, mającym rzekomo służyć przejęciu kontroli nad umysłami ludzkości. Podobnie jak w przypadku nagrań przedstawiających wizje przyszłości, w opisach takich materiałów często zamieszcza się zastrzeżenie, że mają one charakter rozrywkowy.



OBRAZ 27. Materiał rozpowszechniający teorię spiskową (zgłoszenie odrzucone)

ŹRÓDŁO: [YOUTUBE.COM](https://www.youtube.com)



OBRAZ 28. Materiał rozpowszechniający teorię spiskową (zgłoszenie odrzucone)

ŹRÓDŁO: [YOUTUBE.COM](https://www.youtube.com)

Zapraszam na film z cyklu "w zielonej poświacie" gdzie analizujemy W CELACH ROZRYWKOWYCH pod względem logicznym przeróżne teorie z internetu pt.: Projekt Blue Beam i "Falszywa Inwazja Kosmitów" Są Już TECHNICZNIE WYKONALNE! - Analiza Ator Finance

OBRAZ 29. Informacja zamieszczona przez jednego z aktorów dezinformacyjnych

ŹRÓDŁO: [YOUTUBE.COM](https://www.youtube.com)

Podszywanie pod osobę publiczną/instytucję

Wśród naruszeń zgłaszanych do platform społecznościowych na szczególną uwagę zasługują incydenty związane z podszywaniem się pod publicznie znaną osobę lub pod instytucję. Niejednokrotnie oprócz promowania przekazu dezinformacyjnego noszą one także znamiona cyberprzestępstwa (zachęcanie do zakupu usług poprzez reklamowanie ich z wykorzystaniem wizerunku znanych osób lub autorytetu instytucji zaufania publicznego, przekierowywanie do zewnętrznych witryn w celu wyłudzenia danych lub środków).

W I kwartale 2025 roku zgłoszenia dotyczyły między innymi bezprawnego wykorzystania wizerunku premiera Donalda Tuska, prezesa NBP Adama Glapińskiego czy generała Rajmunda Andrzejczaka. Zgłaszano także incydenty związane z wykorzystaniem wizerunku instytucji publicznych, w tym Interpolu, jednostek Wojska Polskiego, Zakładu Ubezpieczeń Społecznych czy marek należących do Totalizatora Sportowego. Dzięki zabiegowi podszywania pod osobę/instytucję opisywane konta mogły nakłonić użytkowników do wpłacenia środków na rzecz nieistniejącego funduszu czy przejścia na niebezpieczną witrynę.

Poniżej przedstawiono przykłady zgłoszonych kont oraz reakcję platform społecznościowych. Wśród nich znalazły się profile podszywające się pod premiera Donalda Tuska oraz markę Mini Lotto, należącą do Totalizatora Sportowego. Do serwisów społecznościowych zgłoszono wiele podobnych kont, a jedynie część z nich została zablokowana.

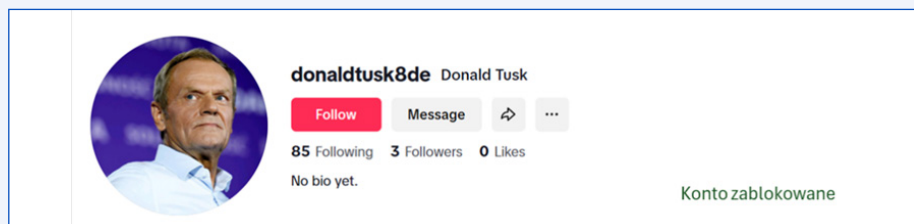
Z uwagi na brak przejrzystości ze strony platform społecznościowych pracownicy OAD NASK nie są w stanie jednoznacznie określić, dlaczego zespoły moderujące podejmują tak różne decyzje. W przypadku odrzuconych zgłoszeń nieskuteczne pozostają również odwołania od pierwszej decyzji moderatorów.



OBRAZ 30.

Konto podszywające się pod Donalda Tuska (zgłoszenie odrzucone)

ŹRÓDŁO: [TIKTOK.COM](https://www.tiktok.com)



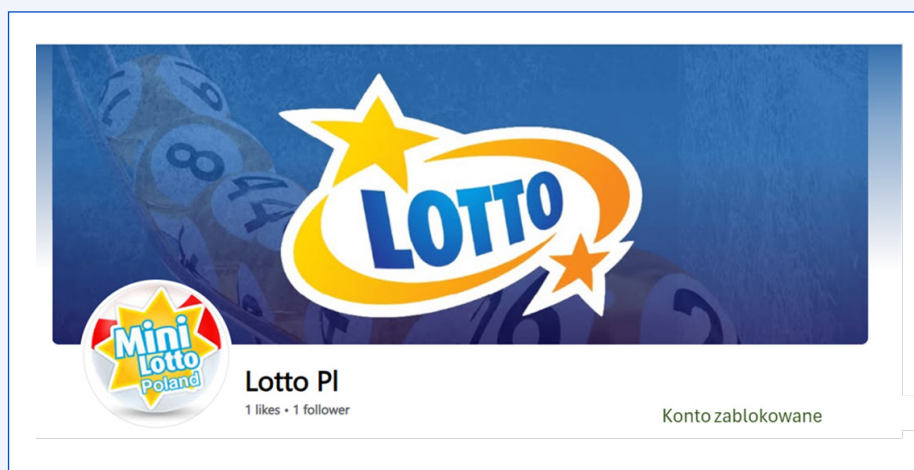
OBRAZ 31. Konto podszywające się pod Donalda Tuska (konto zablokowane)

ŹRÓDŁO: TIKTOK.COM



OBRAZ 32. Konto podszywające się pod Mini Lotto (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



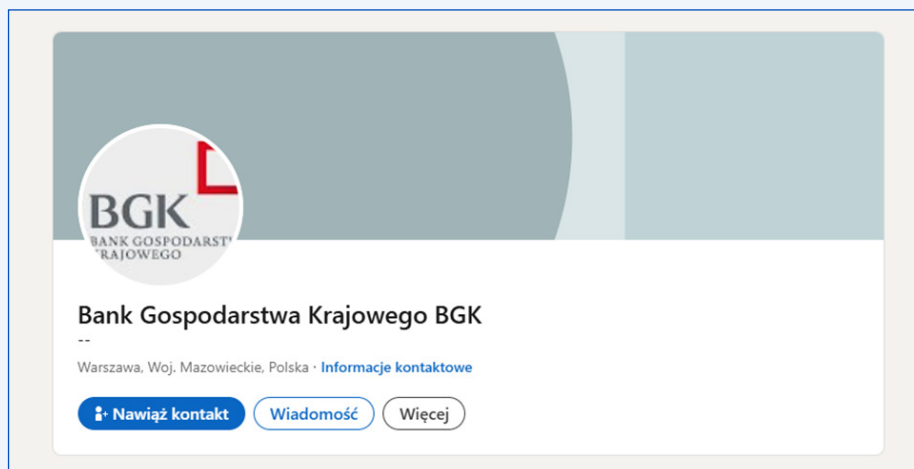
OBRAZ 33. Konto podszywające się pod Mini Lotto (konto zablokowane)

ŹRÓDŁO: FACEBOOK.COM



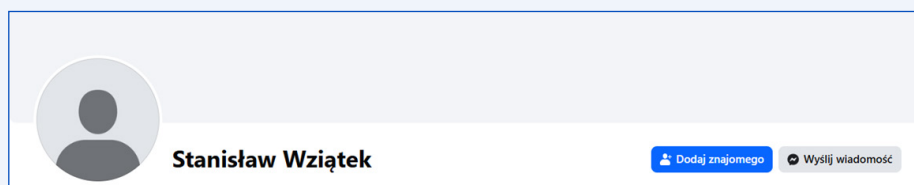
OBRAZ 34. Konto podszywające się pod ZUS (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM



OBRAZ 35. Konto podszywające się pod BGK (zgłoszenie odrzucone)

ŹRÓDŁO: LINKEDIN.COM



OBRAZ 36. Konto podszywające się pod Stanisława Wziątka, podsekretarza stanu w Ministerstwie Obrony Narodowej (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

II kwartał 2025

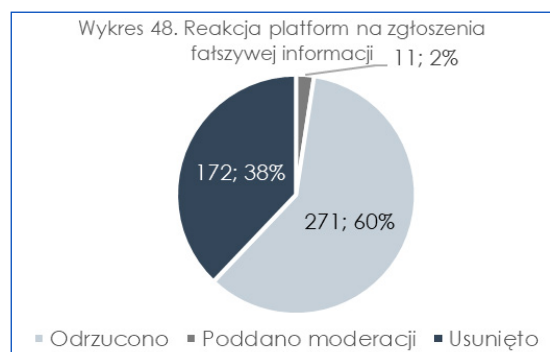
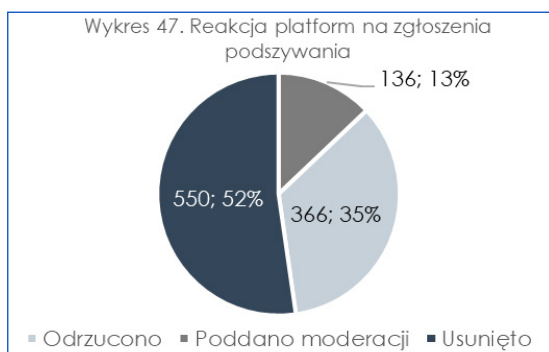
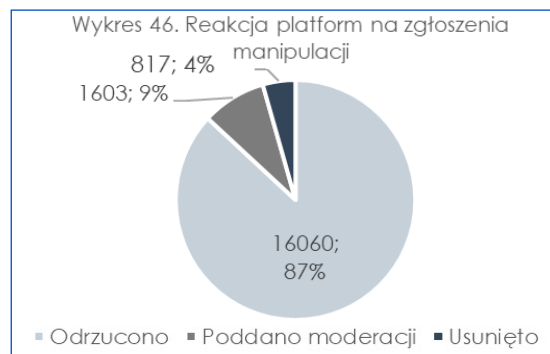
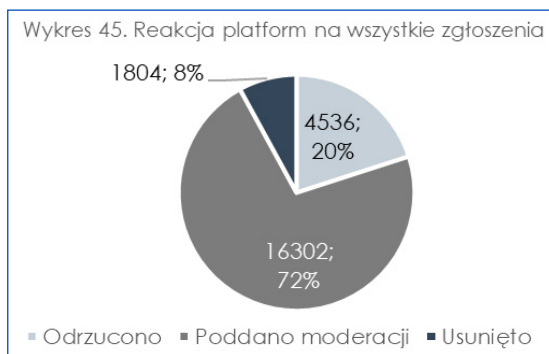
Q2.

II kwartał 2025

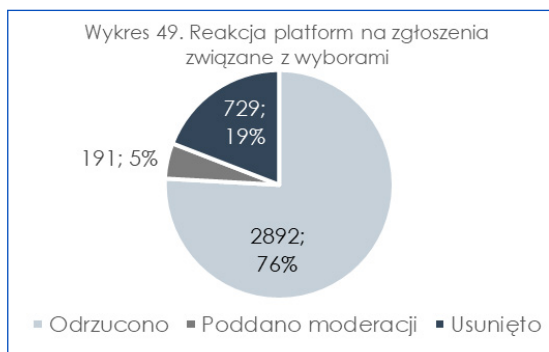
W II kwartale 2025 roku OAD NASK zgłosił do administracji platform społecznościowych 22 642 przypadków treści naruszających regulacje. Usuniętych zostało 8 proc. zgłoszonych materiałów, czyli 1804 publikacje. Dalsze 72 proc., czyli 16 302 wpisy, zostało poddanych moderacji, a 20 proc. zgłoszeń, czyli 4536 przypadków, nie spotkało się z żadną zauważalną reakcją ze strony moderacji. Spośród wszystkich zgłoszonych treści 454 materiały stanowiły fałszywą informację, 18 480 zawierało manipulacje, a 1056 związane było z podszywaniem się pod osobę lub instytucję publiczną. Pozostałych 2656 zgłoszeń dotyczyło incydentów naruszenia ciszy wyborczej.

- Platformy najczęściej podejmowały działania w reakcji na incydenty związane z manipulacją – w analizowanym okresie usunięto 4 proc. zgłoszonych treści, czyli 817 materiałów, zaś moderacji poddano 87 proc. takich wpisów, czyli 16 078 incydentów. Oznacza to jednak, że 9 proc. zmanipulowanych treści, czyli 1603 przypadki, nadal pozostają dostępne.
- 550 z 1056 zgłoszonych kont podszywających się pod osoby publiczne lub instytucje (w sumie 52 proc. zgłoszeń) zostało usuniętych. Moderacji poddano 136 takich profili (13 proc. incydentów), co oznacza, że ich widoczność została ograniczona lub konto zostało zawieszone. 366 zgłoszeń (35 proc. incydentów) zostało odrzuconych przez serwisy społecznościowe.
- Platformy podejmowały mniej działań w przypadku treści fałszywych. Usunięto 172 wpisy z 454 zgłoszonych (38 proc.), a moderacji poddano 11 materiałów (2 proc.). Odrzucono natomiast zgłoszenia dotyczące 271 wpisów (60 proc.).
- Platformom przekazano 3812 zgłoszeń dotyczących naruszeń związanych z procesem wyborczym, w wyniku których 729 treści (19 proc.) zostało usuniętych, a 191 (5 proc.) poddanych moderacji. 2892 zgłoszonych materiałów (76 proc.) jest nadal dostępnych i mogło mieć negatywny wpływ na dyskurs wyborczy w Polsce.
- Zgłoszenia treści ocenionych przez OAD NASK jako naruszające regulaminy platform lub mogące łamać inne regulacje prawne w wielu przypadkach zostały odrzucone bez podania powodu lub z adnotacją, że nie naruszają regulaminu danej platformy społecznościowej.

- Na platformach społecznościowych pojawiały się dezinformujące reklamy, mające na celu nie tylko rozpowszechnianie fałszywych treści, lecz także próbę wyłudzenia danych lub środków od odbiorców. Analitycy OAD NASK zaobserwowali, że działania podejmowane przez zespoły moderujące treści na platformach są w takich przypadkach niewystarczające. Zgłoszenia tego rodzaju często kończyły się jedynie blokadą pojedynczego wpisu, a nie całego konta reklamodawcy, co umożliwia kontynuowanie przestępczej działalności.
- Dodatkowym utrudnieniem związanym z kwestią zgłaszania szkodliwych reklam pojawiających się na platformie Facebook niezmiennie pozostaje temat opóźnień, z jakim dane są aktualizowane w ogólnodostępnej bibliotece reklam. Ograniczona jest także możliwość wyszukiwania i filtrowania zamieszczanych przez reklamodawców treści.



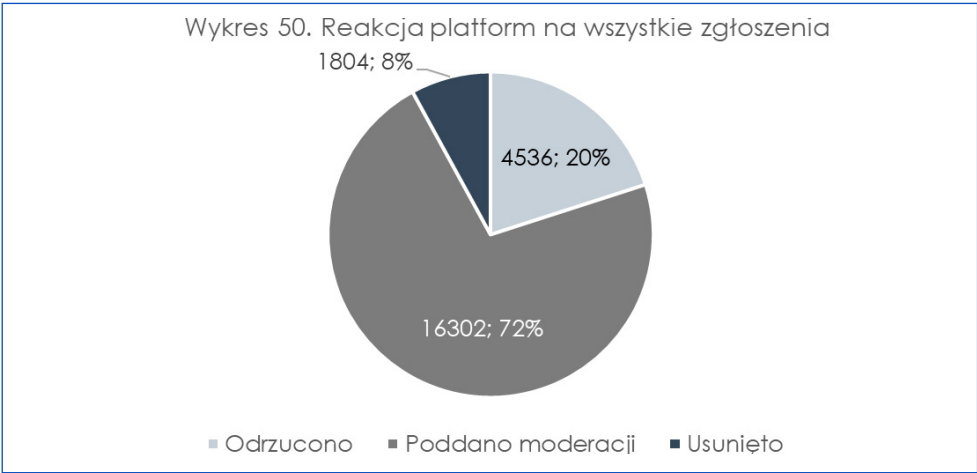
WYKRESY 45-48. Reakcja platform na zgłoszenia: wszystkie (45); manipulacja (46); podszywanie (47); fałszywa informacja (48)



WYKRES 49. Reakcja platform na incydenty związane z wyborami

ŹRÓDŁO: OPRACOWANIE WŁASNE

Reakcja platform na wszystkie zgłoszenia OAD NASK



WYKRES 50. Reakcja platform na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

W I kwartale 2025 roku z platform społecznościowych usunięto 35 proc. zgłoszonych przez OAD NASK treści, czyli 819 wpisów. Dalszych 4 proc. zgłoszonych treści, czyli 95 wpisów, zostało poddanych moderacji, natomiast 61 proc. zgłoszeń, czyli 1452 przypadki, nie spotkało się z żadną widoczną reakcją ze strony moderacji.

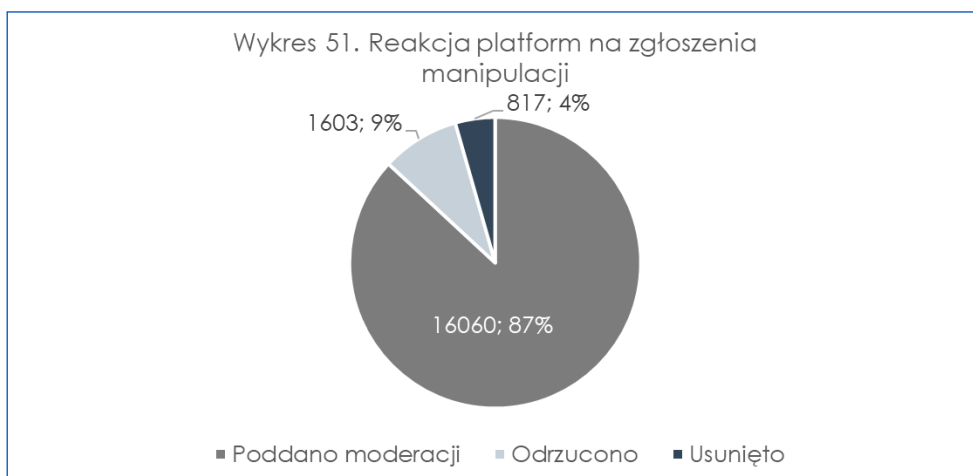
Rodzaj naruszenia	Fałszywa informacja	Łamanie ciszy wyborczej	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń odrzuconych	271	2296	1603	366
Liczba zgłoszeń poddanych = moderacji	11	95	16 060	136
Liczba zgłoszeń usuniętych	172	265	817	550
Całkowita liczba zgłoszeń	454	2 656	18 480	056

TABELA 5. Rodzaj naruszenia w przekazanych zgłoszeniach

ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń. Wśród wszystkich zgłoszeń zidentyfikowano:

- **2,0 proc. (454 przypadki)** wpisów zawierających fałszywe informacje;
- **11,7 proc. (2656 przypadków)** wpisów stanowiących możliwe naruszenie ciszy wyborczej;
- **81,6 proc. (18 480 przypadków)** wpisów zawierających manipulację;
- **4,7 proc. (1056 przypadków)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

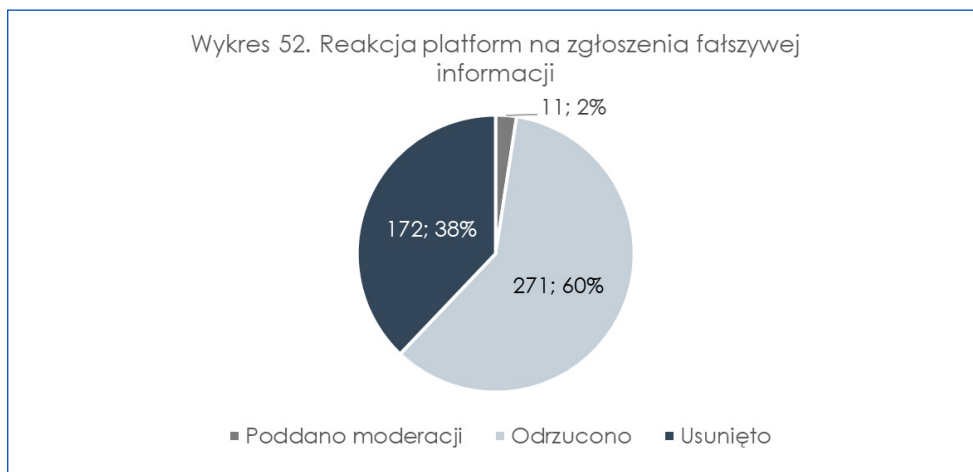


WYKRES 51.

Reakcja platform na zgłoszenia manipulacji

ŹRÓDŁO: OPRACOWANIE WŁASNE

Analiza zgłoszonych treści wykazała, że platformy społecznościowe najczęściej reagowały na incydenty polegające na publikowaniu zmanipulowanych informacji. Usunięto 817 takich materiałów, co odpowiada 4 proc. zgłoszonych treści. Moderacji poddano z kolei 16 060 wpisów, czyli 87 proc. zgłoszonych treści. Pozostałe 1603 zgłoszenia (9 proc.) zostały odrzucone. Jako incydent wystąpienia manipulacji zakwalifikowane zostały między innymi treści publikowane przez konta-boty, wykazujące oznaki automatycznego działania.

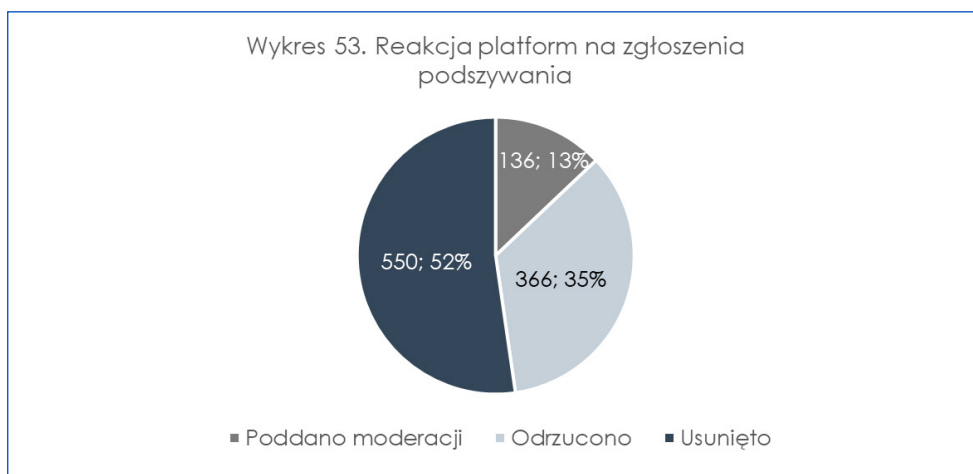


WYKRES 52.

Reakcja platform na zgłoszenia fałszywej informacji

ŹRÓDŁO: OPRACOWANIE WŁASNE

W wyniku zgłoszeń fałszywych informacji usunięto 172 materiały (38 proc.), a moderacji poddano 11 wpisów (2 proc.). Platformy odrzuciły 271 zgłoszeń (60 proc. przypadków).



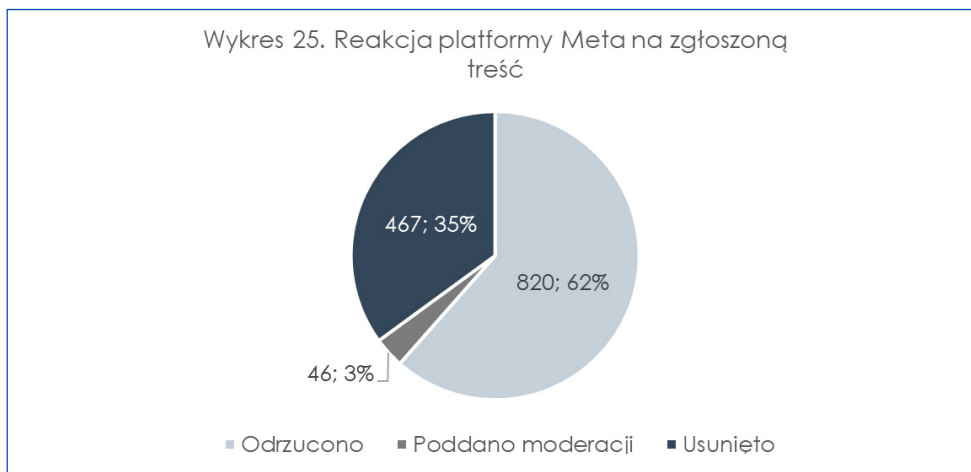
WYKRES 53.

Reakcja platform na zgłoszenia podszywania

ŹRÓDŁO: OPRACOWANIE WŁASNE

Jeżeli chodzi o działania podejmowane przez platformy społecznościowe w odpowiedzi na aktywność kont podszywających się pod osoby publiczne / instytucje, zdecydowano o usunięciu 550 kont (52 proc.), a moderacji poddano 136 profili (13 proc.); moderacja oznacza w tym przypadku, że profil został zawieszony lub nie jest widoczny dla odbiorców przebywających na terenie Polski. Dalszych 366 zgłoszeń (35 proc.) zostało odrzuconych przez osoby odpowiadające za moderację treści.

Meta

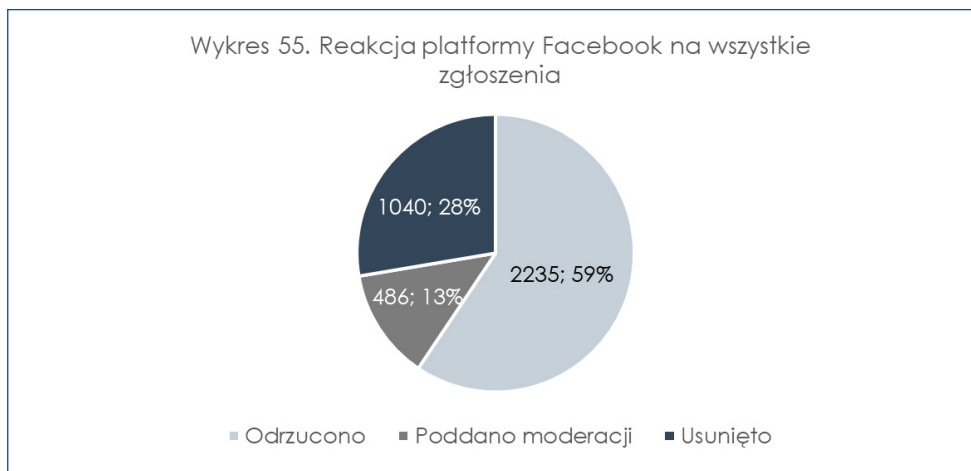


WYKRES 54. Reakcja platformy Meta na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszonych do administracji platform firmy Meta treści pochodzących z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 26 proc. zgłoszonych treści, czyli 1063 materiały, 12 proc., czyli 490 materiałów, poddano moderacji, a 62 proc. zgłoszeń, czyli 2597 przypadków, zostało odrzuconych.

Facebook



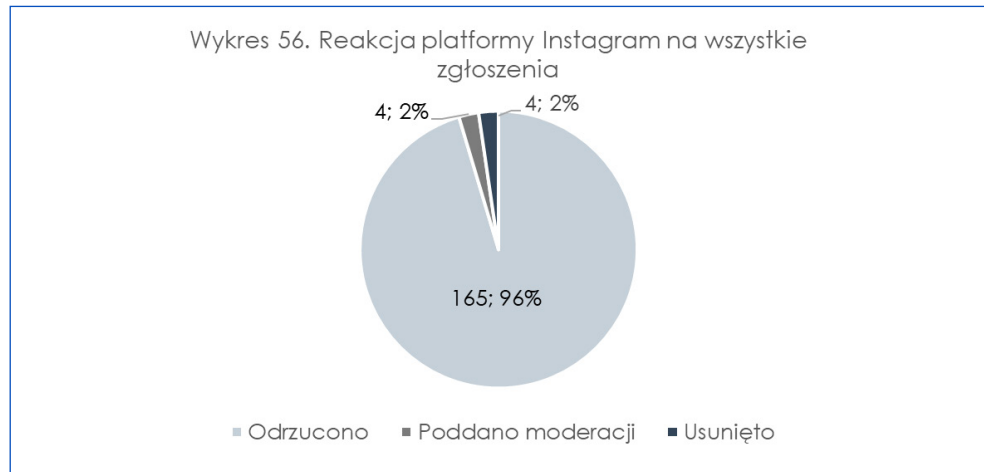
WYKRES 55. Reakcja platformy Facebook na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

W analizowanym okresie platforma Facebook usunęła 28 proc. zgłoszonych treści, czyli 1040 materiały, a 13 proc. zgłoszonych treści, czyli 486 materiałów,

zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 59 proc. zgłoszeń, czyli 2235 przypadków.

Instagram

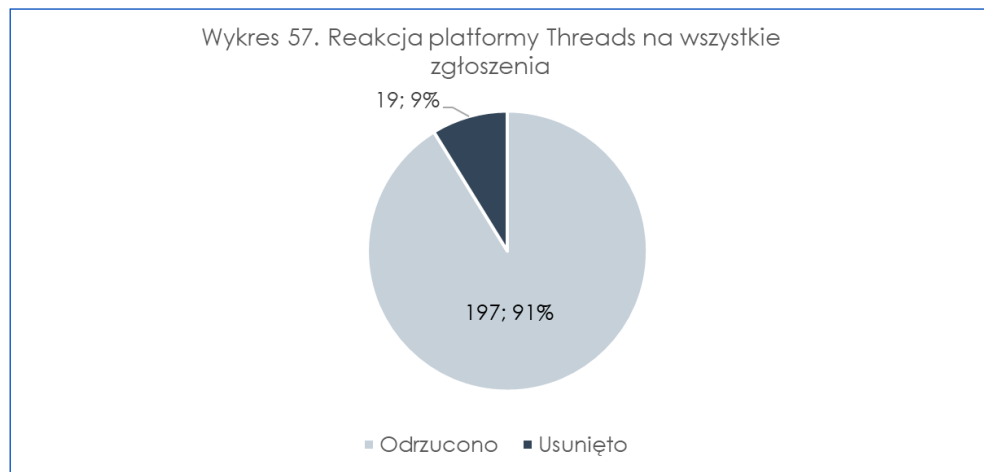


WYKRES 56. Reakcja platformy Instagram na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram usunęła 2 proc. zgłoszonych treści, czyli 4 materiały, a 2 proc. zgłoszonych treści, czyli 4 materiały, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 96 proc. zgłoszeń, czyli 165 przypadków.

Threads

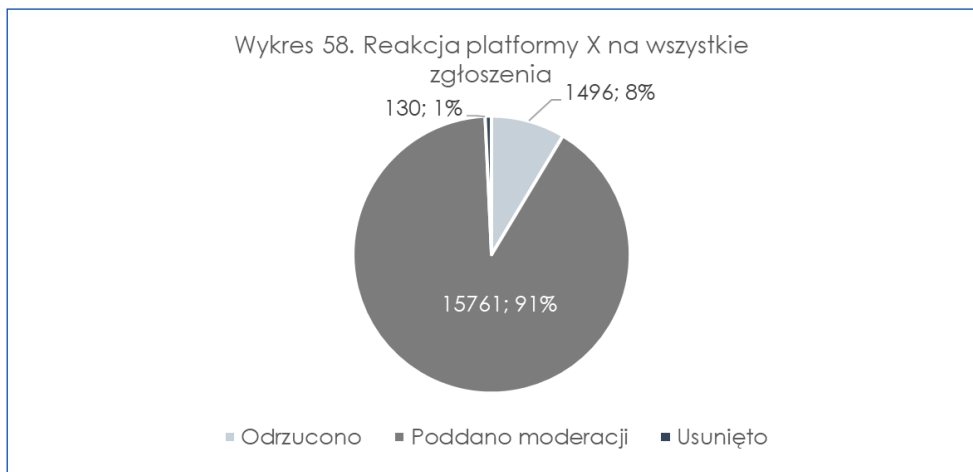


WYKRES 57. Reakcja platformy Threads na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram zdecydowała o usunięciu 9 proc. zgłoszonych treści, czyli 19 materiałów. Nie usunięto natomiast 91 proc. zgłoszonych treści, czyli 197 materiałów.

X.com

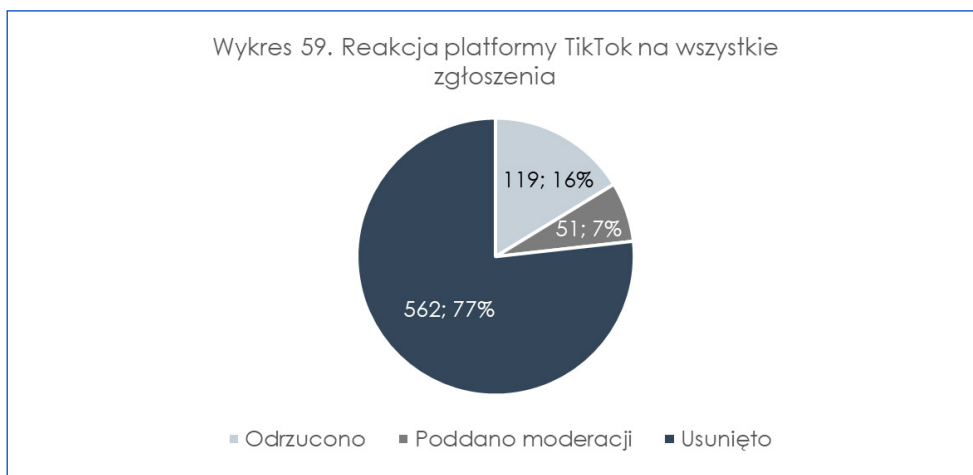


WYKRES 58. Reakcja platformy X na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 1 proc. zgłoszonych treści, czyli 130 materiałów, a 91 proc. zgłoszonych treści, czyli 15 761 materiałów zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 8 proc. zgłoszeń, czyli 1496 przypadków.

TikTok

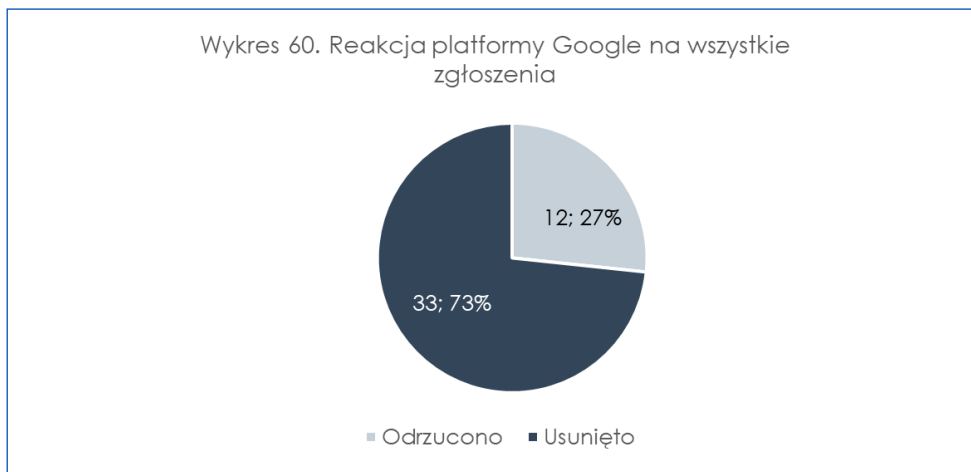


WYKRES 59. Reakcja platformy TikTok na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok usunęła 77 proc. zgłoszonych treści, czyli 562 materiały, a 7 proc. zgłoszonych treści, czyli 51 materiałów zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 16 proc. zgłoszeń, czyli 119 przypadków.

Google (YouTube i Google Ads)



WYKRES 60.

Reakcja platformy Google na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do firmy Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 73 proc. zgłoszonych treści, czyli 33 materiałów. Nie usunięto natomiast 27 proc. zgłoszonych treści, czyli 12 materiałów.

LinkedIn

Administracja platformy LinkedIn odrzuciła wszystkich **11 zgłoszeń**.

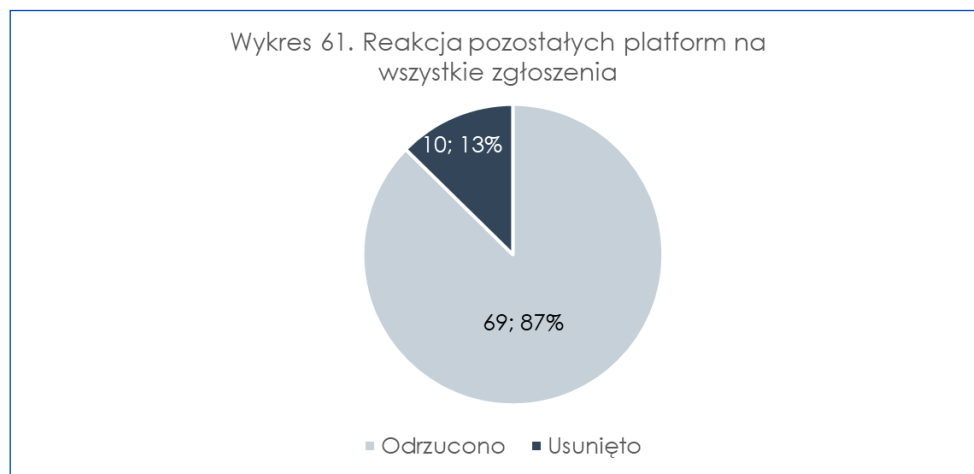
Telegram

Administracja platformy Telegram odrzuciła wszystkich **8 zgłoszeń**.

Bluesky

Administracja platformy Bluesky odrzuciła wszystkie **3 zgłoszenia**.

Pozostałe platformy

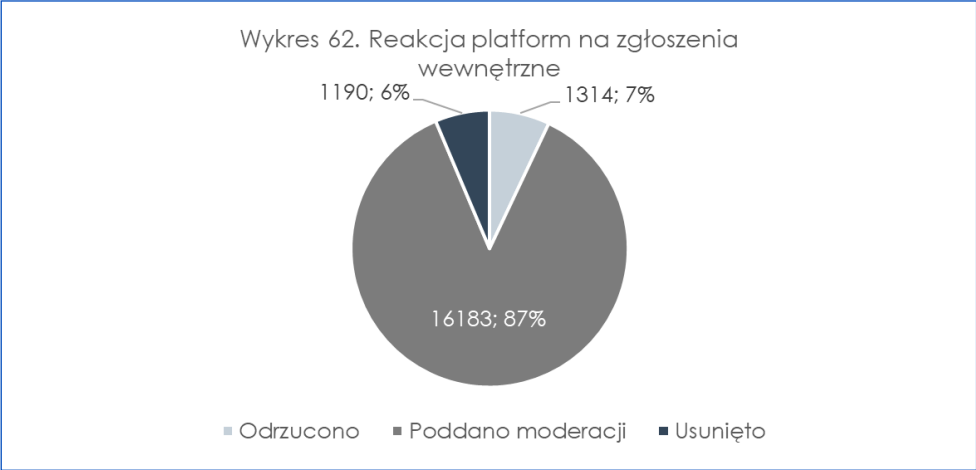


WYKRES 61. Reakcja pozostałych platform na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Pozostałe platformy usunęły 13 proc. zgłoszonych treści, czyli 10 materiałów. Odrzucono natomiast 87 proc. zgłoszeń, czyli 69 przypadków.

Zgłoszenia wewnętrzne



WYKRES 62. Reakcja platform na zgłoszenia wewnętrzne
ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku zgłoszeń wewnętrznych (pochodzących z bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej) w II kwartale 2025 roku z platform społecznościowych usunięto 6 proc. zgłoszonych treści, czyli 1190 wpisów. Dalszych 87 proc. zgłoszonych treści, czyli 16 183 wpisy, zostało poddanych moderacji, natomiast 7 proc. zgłoszeń, czyli 1314 przypadków, nie spotkało się z żadną widoczną reakcją ze strony moderacji.

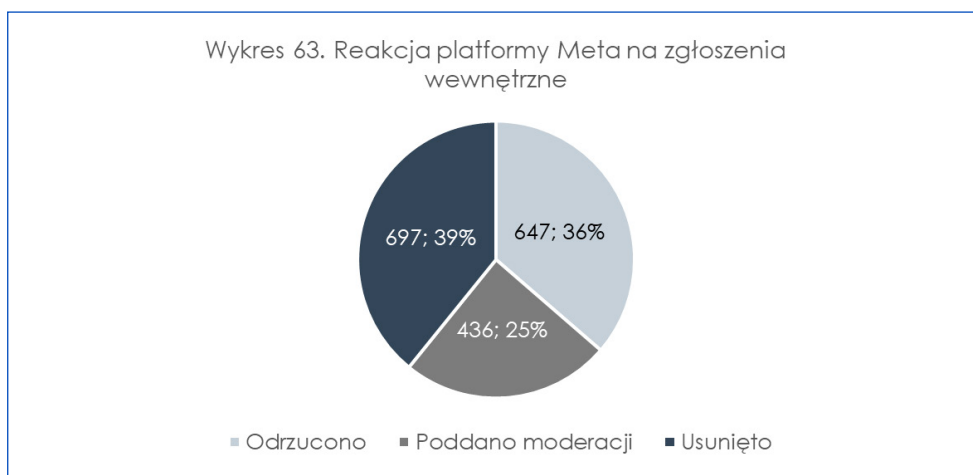
Rodzaj naruszenia	Fałszywa informacja	Łamanie ciszy wyborczej	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń usuniętych	140	6	682	362
Liczba zgłoszeń odrzuconych	110	207	797	200
Liczba zgłoszeń poddanych moderacji	9	4	16 038	132
Całkowita liczba zgłoszeń	259	217	17 517	694

TABELA 6. Rodzaj przypisanego naruszenia w zgłoszeniach wewnętrznych
ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń¹³. Wśród zgłoszeń wewnętrznych zidentyfikowano:

- **1,4 proc. (259 przypadków)** wpisów zawierających fałszywe informacje;
- **1,2 proc. (217 przypadków)** wpisów stanowiących możliwe naruszenie ciszy wyborczej;
- **93,7 proc. (17 517 przypadków)** wpisów zawierających manipulacje;
- **3,7 proc. (694 przypadków)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

Meta



WYKRES 63.

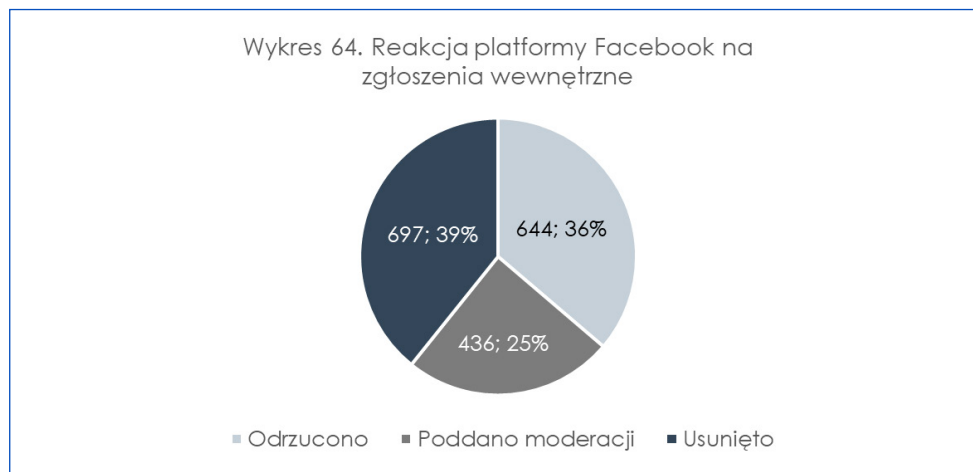
Reakcja platformy Meta na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszeń wewnętrznych przekazanych do administracji platform firmy Meta w związku z treściami pochodzącymi z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 39 proc. zgłoszonych treści, czyli 697 materiałów, 25 proc., czyli 436 materiały, poddano moderacji, a 36 proc. zgłoszeń, czyli 647 przypadków, zostało odrzuconych.

¹³ Podział na wskazane kategorie został dokonany wewnętrznie przez Zespół Szybkiego Reagowania i Wykrywania Dezinformacji w celu uproszczenia procedury realizacji zgłoszeń.

Facebook



WYKRES 64. Reakcja platformy Facebook na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Facebook usunęła 39 proc. zgłoszonych treści, czyli 697 materiałów, a 25 proc. zgłoszonych treści, czyli 436 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 36 proc. zgłoszeń, czyli 644 przypadki.

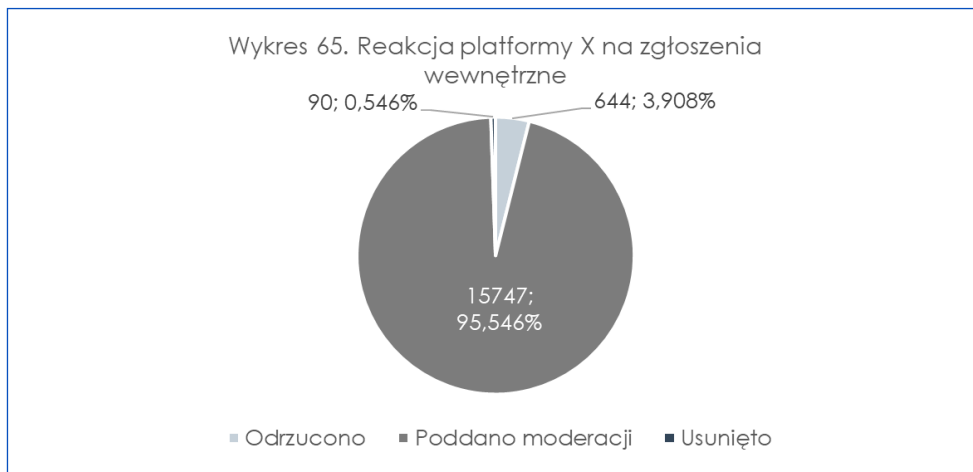
Instagram

Administracja platformy Instagram odrzuciła wszystkie 2 zgłoszenia.

Threads

Administracja platformy Threads odrzuciła 1 przesłane do niej zgłoszenie.

X.com

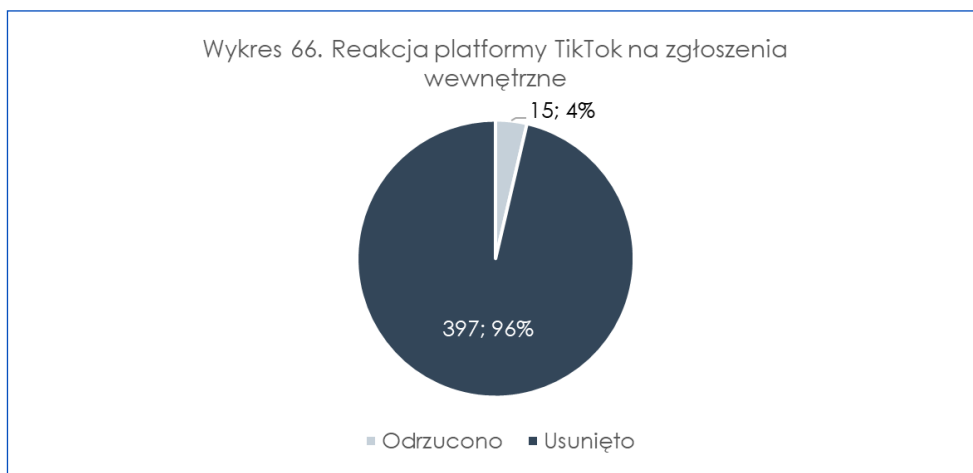


WYKRES 65. Reakcja platformy X na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła mniej niż 1 proc. zgłoszonych treści, czyli 90 materiałów, a 95 proc. zgłoszonych treści, czyli 15 747 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 4 proc. zgłoszeń, czyli 644 przypadki.

TikTok



WYKRES 66. Reakcja platformy TikTok na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok zdecydowała o usunięciu 96 proc. zgłoszonych treści, czyli 397 materiałów. Nie usunięto natomiast 4 proc. zgłoszonych treści, czyli 15 materiałów.

Google (YouTube i Google Ads)

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do firmy Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu wszystkich 5 zgłoszonych materiałów.

Telegram

Administracja platformy Telegram odrzuciła 1 przesłane do niej zgłoszenie.

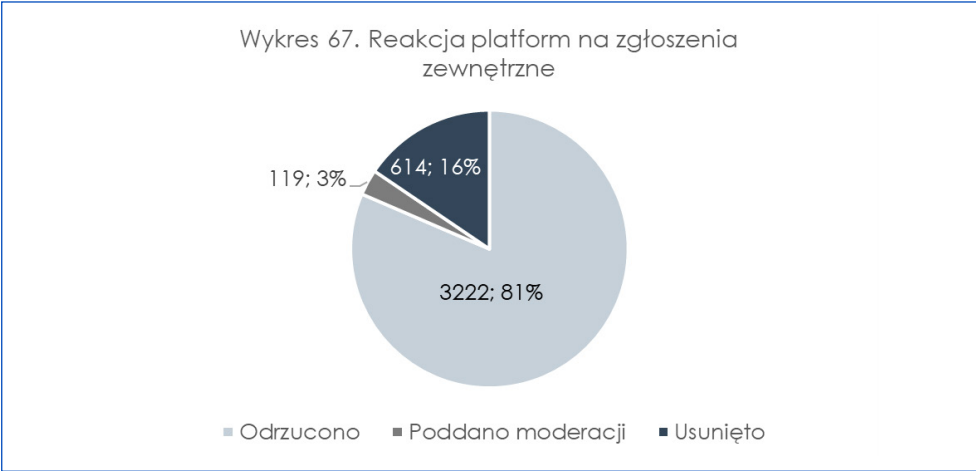
Bluesky

Administracja platformy Bluesky odrzuciła 1 przesłane do niej zgłoszenie.

Pozostałe platformy

Administratorzy pozostałych platform usunęli 1 zgłoszony materiał i odrzucili 1 zgłoszenie.

Zgłoszenia zewnętrzne



WYKRES 67. Reakcja platform na zgłoszenia zewnętrzne
ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku zgłoszeń zewnętrznych (przekazywanych do Zespołu Szybkiego Reagowania i Wykrywania Dezinformacji przez obywateli i obywatelki, podmioty publiczne oraz inne instytucje) w II kwartale 2025 roku z platform społecznościowych usunięto 16 proc. zgłoszonych treści, czyli 614 wpisów. Dalszych 3 proc. zgłoszonych treści, czyli 119 wpisów, zostało poddanych moderacji, natomiast 81 proc. zgłoszeń, czyli 3222 przypadków, nie spotkało się z żadną widoczną reakcją ze strony moderacji. 3267 materiały zgłoszone do OAD NASK zostały ocenione jako niezasadne, czyli takie, które nie stanowiły naruszenia regulaminów bądź innych regulacji prawnych.

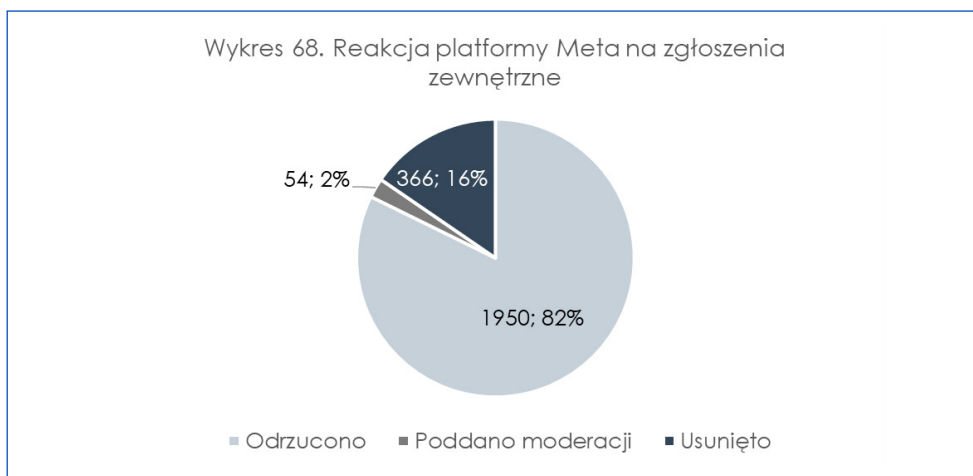
Rodzaj naruszenia	Fałszywa informacja	Łamanie ciszy wyborczej	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń usuniętych	32	259	135	188
Liczba zgłoszeń odrzuconych	161	2089	806	166
Liczba zgłoszeń poddanych moderacji	2	91	22	4
Całkowita liczba zgłoszeń	195	2439	963	358

TABELA 7. Rodzaj naruszenia w zgłoszeniach zewnętrznych
ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń¹⁴. Wśród zgłoszeń zewnętrznych zidentyfikowano:

- **5 proc. (195 przypadków)** wpisów zawierających fałszywe informacje;
- **62 proc. (2439 przypadków)** wpisów stanowiących możliwe naruszenie ciszy wyborczej;
- **24 proc. (963 przypadków)** wpisów zawierających manipulacje;
- **9 proc. (358 przypadków)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

Meta



WYKRES 68.

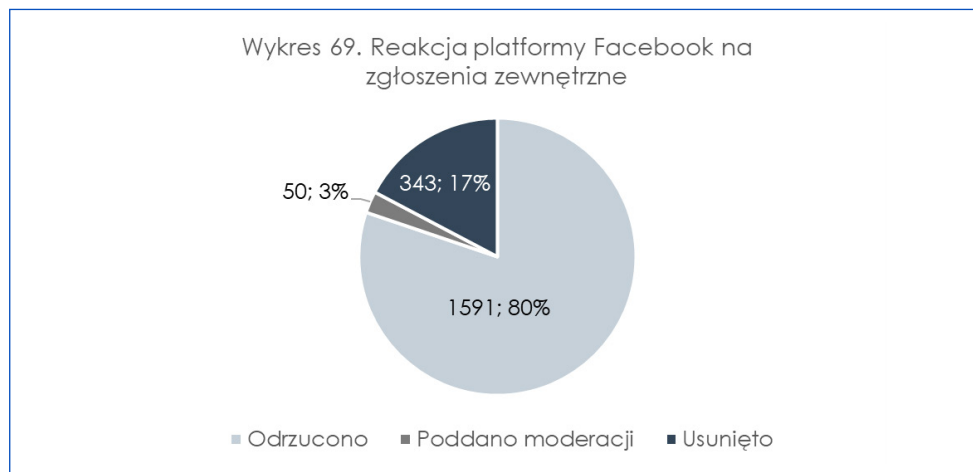
Reakcja platformy Meta na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszeń zewnętrznych przekazanych do administracji platform firmy Meta w związku z treściami pochodzącymi z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 16 proc. zgłoszonych treści, czyli 366 materiały, 2 proc. zgłoszonych treści, czyli 54 materiałów, poddano moderacji, a 82 proc. zgłoszeń, czyli 1950 przypadków, zostało odrzuconych.

¹⁴ Podział na wskazane kategorie został dokonany wewnętrznie przez Zespół Szybkiego Reagowania i Wykrywania Dezinformacji w celu uproszczenia procedury realizacji zgłoszeń.

Facebook

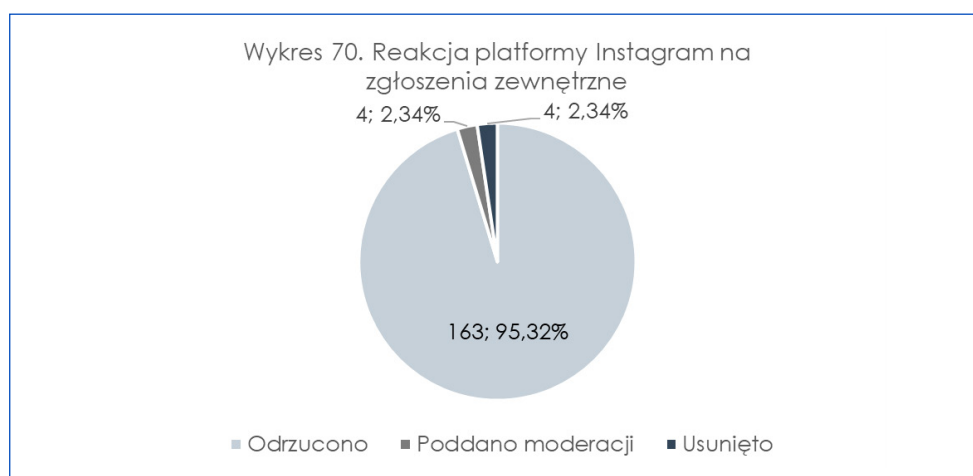


WYKRES 69. Reakcja platformy Facebook na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Facebook usunęła 17 proc. zgłoszonych treści, czyli 343 materiały, a 3 proc. zgłoszonych treści, czyli 50 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 80 proc. zgłoszeń, czyli 1591 przypadków.

Instagram

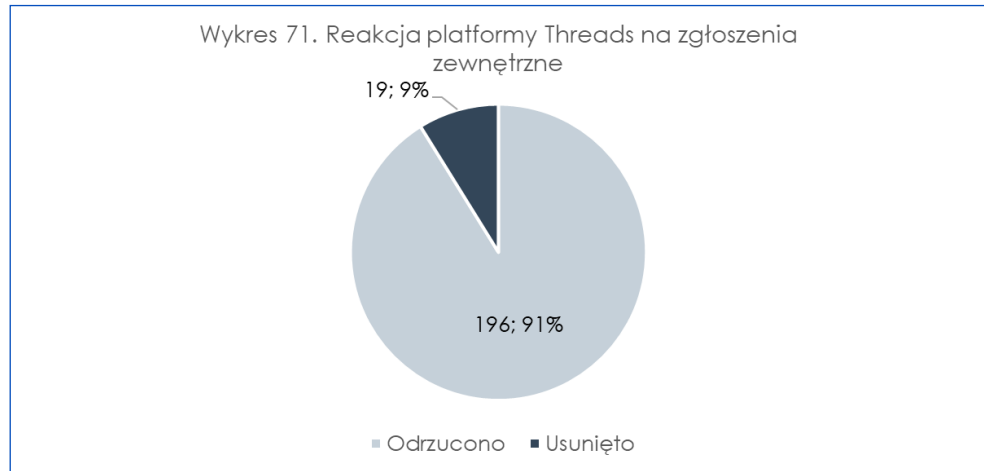


WYKRES 70. Reakcja platformy Instagram na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram usunęła 2 proc. zgłoszonych treści, czyli 4 materiałów, a kolejne 2 proc. zgłoszonych treści, czyli 4 materiały, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili niecałe 96 proc. zgłoszeń, czyli 163 przypadki.

Threads

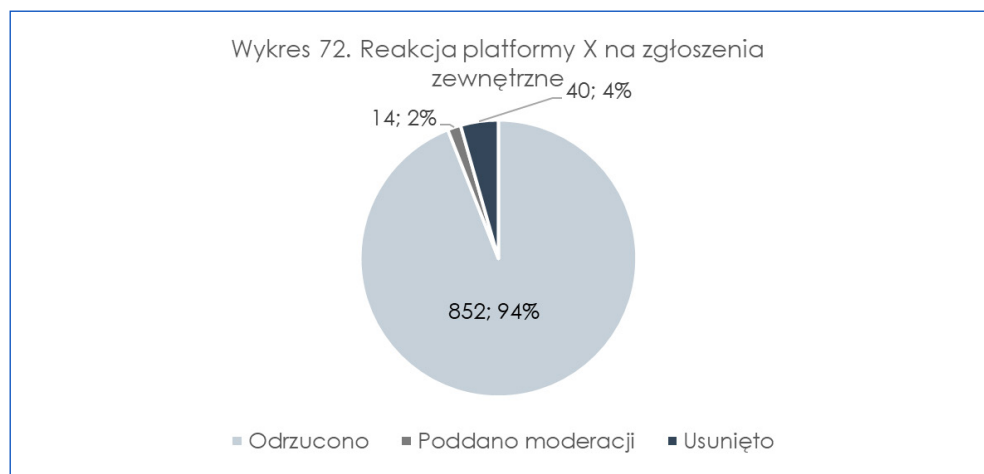


WYKRES 71. Reakcja platformy Threads na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Threads zdecydowała o usunięciu 9 proc. zgłoszeń, czyli 19 materiałów. Nie usunięto natomiast 91 proc. zgłoszonych treści, czyli 196 materiałów.

X.com

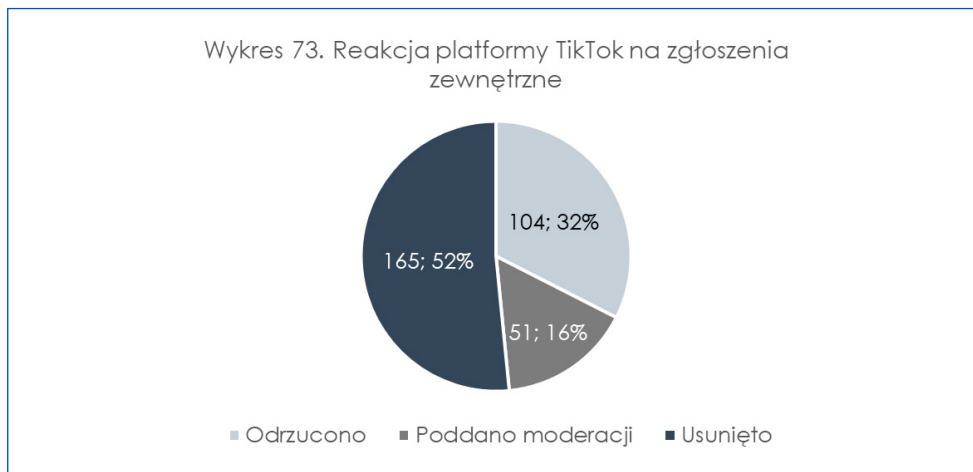


WYKRES 72. Reakcja platformy X na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 4 proc. zgłoszonych treści, czyli 40 materiałów, a 2 proc. zgłoszonych treści, czyli 14 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 94 proc. zgłoszeń, czyli 852 przypadki.

TikTok

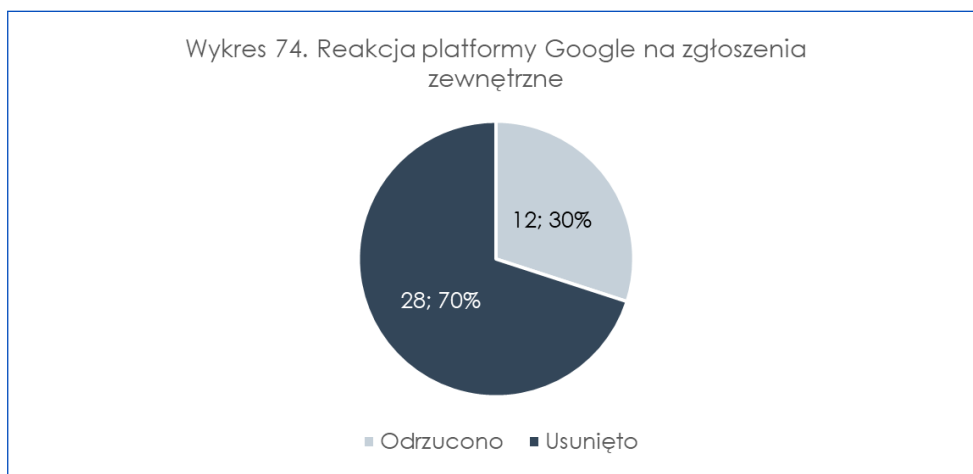


WYKRES 73. Reakcja platformy TikTok na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok usunęła 52 proc. zgłoszonych treści, czyli 165 materiałów, a 16 proc. zgłoszonych treści, czyli 51 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 32 proc. zgłoszeń, czyli 104 przypadki.

Google (YouTube i Google Ads)



WYKRES 74. Reakcja platformy Google na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 30 proc. zgłoszonych treści, czyli 12 materiałów. Nie usunięto natomiast 70 proc. zgłoszonych treści, czyli 28 materiałów.

LinkedIn

Platforma LinkedIn odrzuciła wszystkich 11 zgłoszeń.

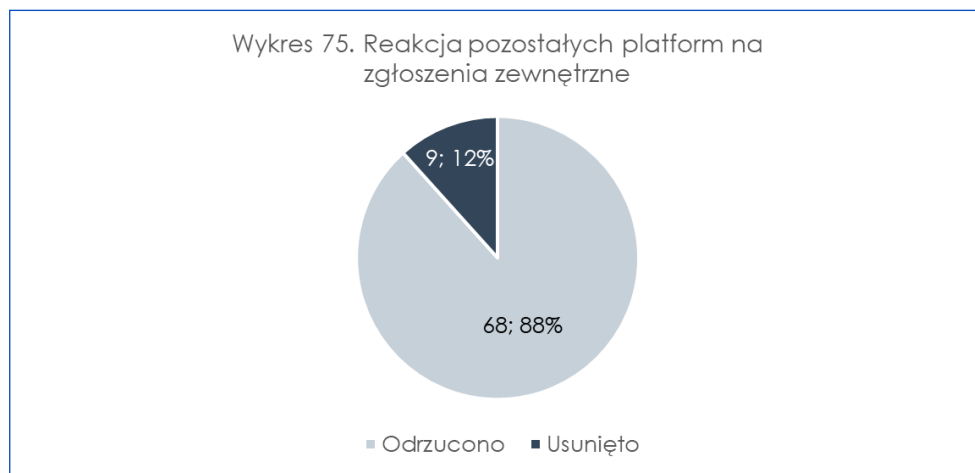
Telegram

Platforma Telegram odrzuciła wszystkich 7 zgłoszeń.

Bluesky

Platforma Bluesky odrzuciła wszystkie 2 zgłoszenia.

Pozostałe platformy



WYKRES 75.

Reakcja pozostałych platform na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Administratorzy pozostałych platform usunęli 12 proc. zgłoszonych treści, czyli 9 materiałów i odrzucili 88 proc. zgłoszeń, czyli 68 przypadków.

Case studies

Dezinformacja na temat zabójstwa na terenie kampusu UW

W maju 2025 roku opinię publiczną poruszyło zabójstwo 53-letniej portierki Uniwersytetu Warszawskiego. Służby jeszcze tego samego dnia ujawniły dane sprawcy¹⁵. Pomimo tego zdarzenie zostało wykorzystane do rozpowszechniania dezinformacji uderzającej w obywateli Ukrainy. W mediach społecznościowych pojawiły się fałszywe informacje, jakoby sprawcą morderstwa był Dmytro Budchuk, obywatel Ukrainy poszukiwany za jazdę pod wpływem alkoholu. Pojawiły się również twierdzenia, że media rzekomo ukrywają prawdziwą narodowość sprawcy.

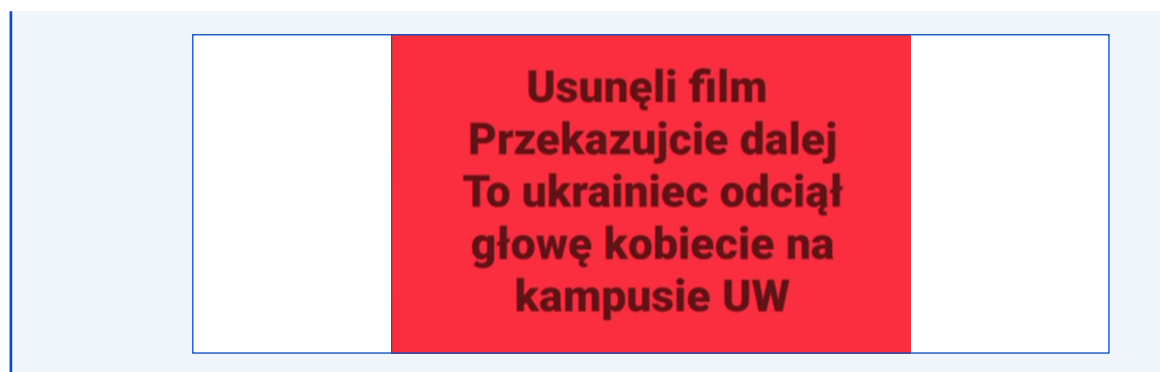
Komunikaty policji i prokuratury, wskazujące na polską narodowość zabójcy, były przekazywane przez analityków do administracji platform społecznościowych w charakterze uzasadnienia zgłoszeń wprowadzających w błąd wpisów jako dezinformacji. Pomimo tego serwisy nie podejmowały działań wobec zgłaszanych treści.

W jednym przypadku serwis Meta potwierdził przekazanie materiału do zespołu fact-checkingowego. Analitycy nie otrzymali jednak informacji o wyniku weryfikacji, a manualne sprawdzenie potwierdziło, że treść pozostaje dostępna.

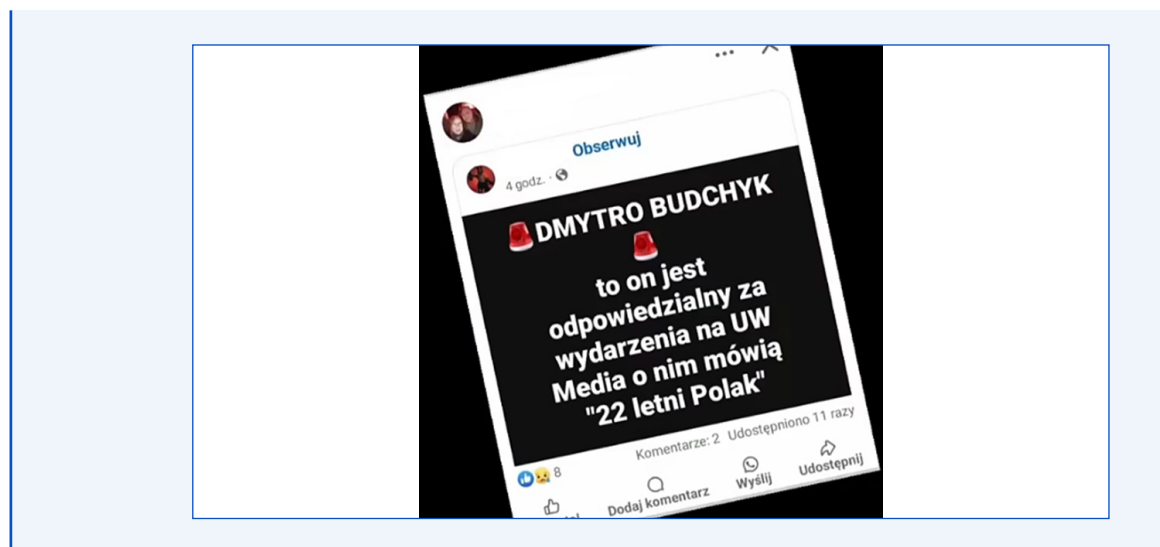
¹⁵ <https://ksp.policja.gov.pl/pl/dzialania/aktualnosci/66279,Zbrodnia-na-kampusie-uniwersyteckim.html> (dostęp 28.08.2025, godz. 12:00)



OBRAZ 37. Wpis zawierający fałszywą informację na temat tożsamości zabójcy na UW (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM



OBRAZ 38. Wpis zawierający fałszywą informację na temat tożsamości zabójcy na UW (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM



OBRAZ 39. Wpis zawierający fałszywą informację na temat tożsamości zabójcy na UW (zgłoszenie odrzucone)
ŹRÓDŁO: YOUTUBE.COM

Hi,

Thank you for your report.

While we don't remove all untrue claims, our approach to misinformation involves reducing the distribution of known misinformation by partnering with independent third-party fact-checkers to review and rate the accuracy of content on Facebook. These partners have been certified through the non-partisan International Fact-Checking Network (IFCN). We have reviewed the following content according to our policies and it is eligible for external fact-checker review. We ask fact-checkers to prioritize provably false claims that are timely, trending and consequential, though they ultimately have discretion on what to review and rate.

[links to content that has been enqueued to 3PFC].

If you have more information or would like us to investigate further, please do not hesitate to let us know.

Kind regards,
Meta Team

OBRAZ 40.

Wiadomość o przekazaniu treści do zespołu fact-checkingowego

ŹRÓDŁO: WIADOMOŚCI E-MAIL PRZESŁANE PRZEZ ZESPÓŁ META

Hi,

As a result of your appeal, we have reviewed the content at issue in your original report again.

However, it still isn't clear to us that the content you reported violates Facebook's Community Standards and/or Instagram's Community Guidelines.

In addition, we have passed the content to third party fact checkers (3PFC) for review for potential misinformation.

If you would be willing to report this content as false news through the standard reporting option, in-product reporting helps us improve our models:

[FACEBOOK] <https://www.facebook.com/help/572838089565953>

[INSTAGRAM] <https://help.instagram.com/1735798276553028>

You may have the right to challenge our decision in court. You may also be able to refer our decision to a certified dispute settlement body.

Thanks,
Meta Team

OBRAZ 41.

Wiadomość stanowiąca odpowiedź na odwołanie od pierwotnej decyzji

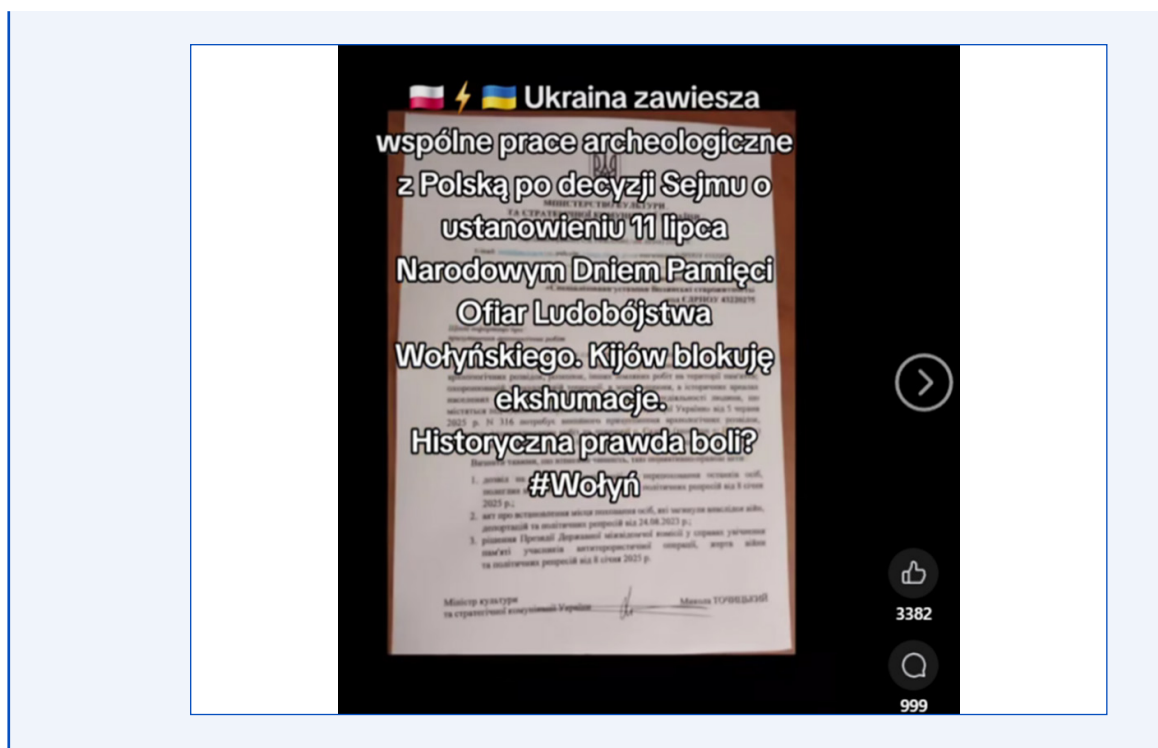
ŹRÓDŁO: WIADOMOŚCI E-MAIL PRZESŁANE PRZEZ ZESPÓŁ META

Dezinformacja na temat ekshumacji ofiar rzezi wołyńskiej

W badanym okresie analitycy OAD NASK zgłaszali do platform społecznościowych dezinformacyjne przekazy związane z prowadzonymi w Ukrainie ekshumacjami ofiar rzezi wołyńskiej.

W mediach społecznościowych pojawiały się fałszywe informacje na temat decyzji o zawieszeniu prac ekshumacyjnych na terenie ukraińskiej miejscowości Puźniki. Według doniesień, które pojawiły się w sieci, była to reakcja władz Ukrainy na ustanowienie w Polsce nowego święta, Narodowego Dnia Pamięci o Polakach – Ofiarach Ludobójstwa dokonanego przez OUN i UPA. Strony ukraińska i polska zaprzeczyły pojawiającym się fałszywym doniesieniom¹⁶.

Po zgłoszeniu dezinformacyjnych wpisów na ten temat platforma Meta zdecydowała o dodaniu notki informującej o fałszywej treści do jednego z przesłanych materiałów. Pozostałe wpisy nie zostały poddane moderacji.



OBRAZ 42.

Materiał zawierający fałszywą informację o zawieszeniu ekshumacji w Puźnikach (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

¹⁶ <https://www.gov.pl/web/kultura/oswiadczenie-dot-nieprawdziwych-informacji-dotyczacych-wstrzymania-pozwolen-na-ekshumacje> (dostęp 28.08.2025, godz. 12:00)

Podobno Ukraińcy za naszą Ustawę o Dniu Pamięci 11 lipca, zablokowali ekshumację w Pużnikach czyli jednej wsi z 5000 wiosek wyróżnionych przez ukraińskich sąsiadów. Ktoś się czymś martwi? Przynajmniej mamy z głowy zamartwianie się o możliwość profanowania polskich kości ukraińskimi rękami oraz możliwość zafalszowania historii przez stronę ukraińską. Poczekamy na lepsze czasy.

Ostatnia zmiana: 8:04 PM · 7 cze 2025 · 9 660 wyświetleń

23

210

768

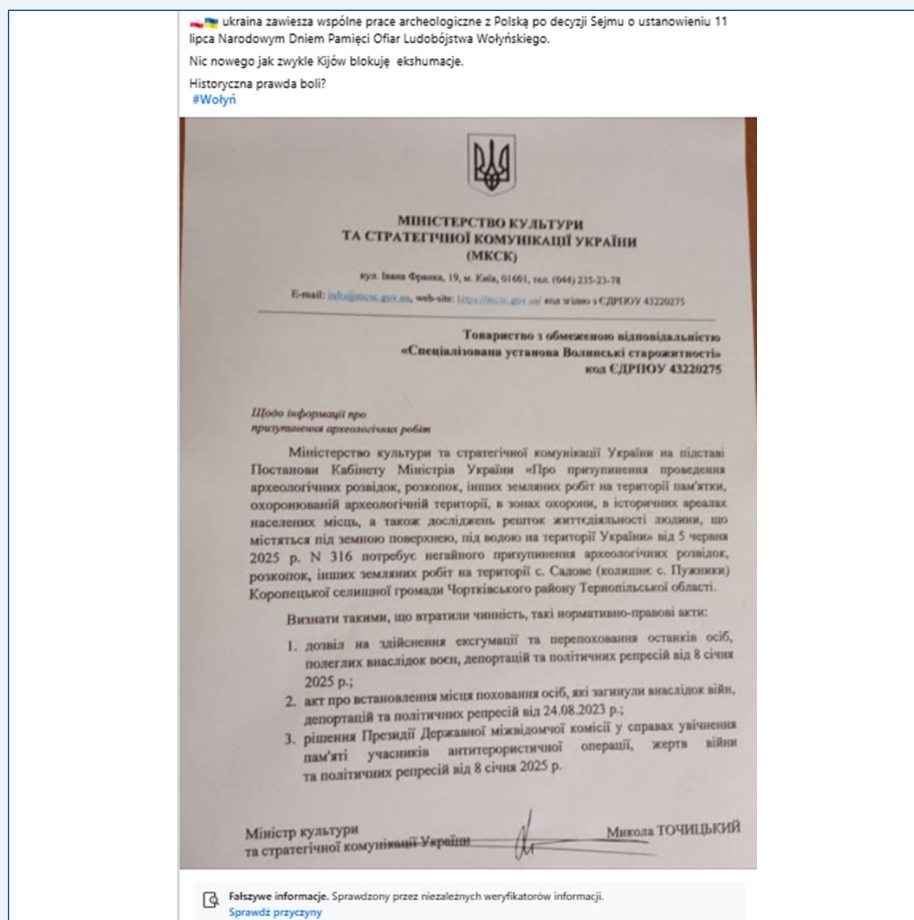
13



OBRAZ 43.

Wpis zawierający fałszywą informację o zawieszeniu ekshumacji w Pużnikach

ŹRÓDŁO: X.COM



OBRAZ 44.

Wpis o ekshumacjach oznaczony przez platformę jako fałszywa informacja

ŹRÓDŁO: FACEBOOK.COM

Kolejnym ze zgłaszanych do platform społecznościowych wątków dezinformacyjnych związanych z ekshumacjami były twierdzenia o rzekomym zacieraniu śladów ludobójstwa przez stronę ukraińską. Według pojawiających się doniesień, Ukraińcy nie chcieli wpuścić dziennikarzy na miejsce poszukiwań, aby prawda o zbrodni nie wyszła na jaw. Jak informowała strona polska¹⁷, w pracach brali udział członkowie Instytutu Pamięi Narodowej, naukowcy z Pomorskiego Uniwersytetu Medycznego oraz przedstawiciele Fundacji Wolność i Demokracja.



OBRAZ 45.

Wpis zawierający fałszywą informację na temat odsunięcia mediów od ekshumacji w Pużnikach (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

¹⁷ <https://www.radiopik.pl/3,128996,rzecznik-resortu-kultury-o-ekshumacjach-w-puznikach-prace-ruszyly-ida-zgodnie-z-planem> (dostęp 28.08.2025, godz. 12:00)

Dezinformacja klimatyczna

W II kwartale 2025 roku zgłaszano wpisy, które zaprzeczały istnieniu globalnego ocieplenia. Aktorzy dezinformacji rozpowszechniali również fałszywe twierdzenia o tym, że wysokie stężenie CO₂ w atmosferze nie ma wpływu na temperaturę Ziemi. W mediach społecznościowych pojawia się także dezinformacja sugerująca, że zmiany klimatu to rzekomo efekt cyklicznych zmian, a nie działalności człowieka.

Mimo że istnieją dowody¹⁸ na fałszywość przedstawionych tez, platformy społecznościowe nie podejmują działań, które miałyby na celu ograniczenie rozprzestrzeniania zgłaszanych przez analityków OAD NASK treści. Obecność dezinformacji klimatycznej w sieci prowadzi między innymi do opóźniania działań na rzecz ochrony klimatu oraz do polaryzacji społeczeństwa. Dlatego w walce z postępującym zagrożeniem ważne jest podjęcie efektywnej moderacji fałszywych twierdzeń dotyczących tego tematu.

Dokładnie, klimat określa się w perspektywie wieloletniej. W ostatnim tysiącleciu mieliśmy zarówno znaczne ocieplenie (winnice na Wielkiej Brytanii) oraz ochłodzenie (mała epoka lodowcowa w XVIIw.). Jest dokładne 0 dowodów na antropocentryczność zjawisk klimatycznych.

1:22 PM · 6 kwi 2025 · 114 wyświetleń

OBRAZ 46.

Wpis zawierający dezinformację klimatyczną (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

Przede wszystkim nie ma dowodów na to, że stężenie CO₂ w atmosferze ma wpływ na temperaturę na Ziemi. Są tylko hipotezy wysnute na podstawie przypuszczeń. I oczywiście konsensus czyli zmowa.

1:49 PM · 6 kwi 2025 · 230 wyświetleń

OBRAZ 47.

Wpis zawierający dezinformację klimatyczną (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

¹⁸ <https://www.gov.pl/web/edukacja-ekologiczna/najwieksze-mity-o-zmianach-klimatu> (dostęp 28.08.2025, godz. 12:00)

Tak, żebyś wiedział ... Globalne ocieplenie to mistyfikacja. Ludzie forsujący globalne ocieplenie to oszuści. Ludzie, którzy wierzą w globalne ocieplenie, to naiwni frajerzy o niskim IQ. Ani jedna przepowiednia kultu klimatycznego zagłady nigdy się nie sprawdziła.

Zapowiadany przez Gretę Thunberg koniec świata odwołany



Greta Thunberg

Fot. Shutterstock

Ten tekst jest częścią
STRONY PREMIUM WNPPL

Greta Thunberg usunęła swój tweet sprzed pięciu lat, w którym ostrzegała, że ludzkość ulegnie zagładzie, jeśli nie odejdzie od paliw kopalnych do 2023 roku.

Autor: JD

11:19 AM · 3 kwi 2025 · 4 022 wyświetlenia

OBRAZ 48.

Wpis zawierający dezinformację klimatyczną (zgłoszenie odrzucone)

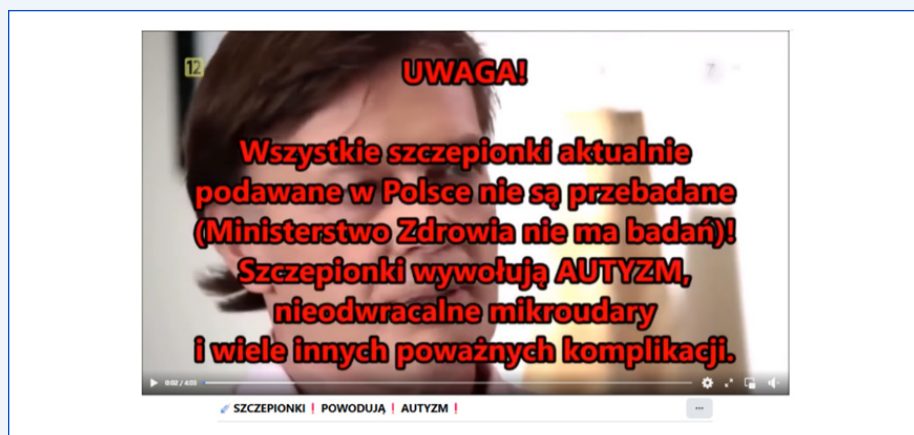
ŹRÓDŁO: X.COM

Dezinformacja medyczna

W badanym okresie analitycy Ośrodka Analizy Dezinformacji NASK zgłaszali do platform społecznościowych treści należące do kategorii dezinformacji medycznej.

Znaczną część zgłoszonych incydentów stanowiły materiały o charakterze antyszczepionkowym. Skupiały się one przede wszystkim na podważaniu bezpieczeństwa oraz skuteczności szczepionek, co mogło wpływać na osłabienie zaufania społecznego wobec instytucji zdrowia publicznego oraz wzbu-
dzać niepokój w odniesieniu do programów szczepień.

Jednym z najczęściej powtarzających się wątków były narracje sugerujące, że szczepionki zawierają substancje szkodliwe, takie jak między innymi tiomersal (organiczny związek rtęci stosowany jako konserwant). Wskazywano, że jego obecność ma wywoływać poważne skutki uboczne, w tym zaburzenia rozwojowe, problemy neurorozwojowe czy, najczęściej przywoływany, autyzm. Kwestionowano przy tym stanowisko instytucji naukowych, między innymi Amerykańskiego Centrum Kontroli i Prewencji Chorób (CDC), które wielokrotnie podkreślało brak związku tiomersalu z autyzmem.



OBRAZ 49.

Wpis zawierający dezinformację na temat szczepień (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



OBRAZ 50.

Wpis zawierający dezinformację na temat szczepień (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



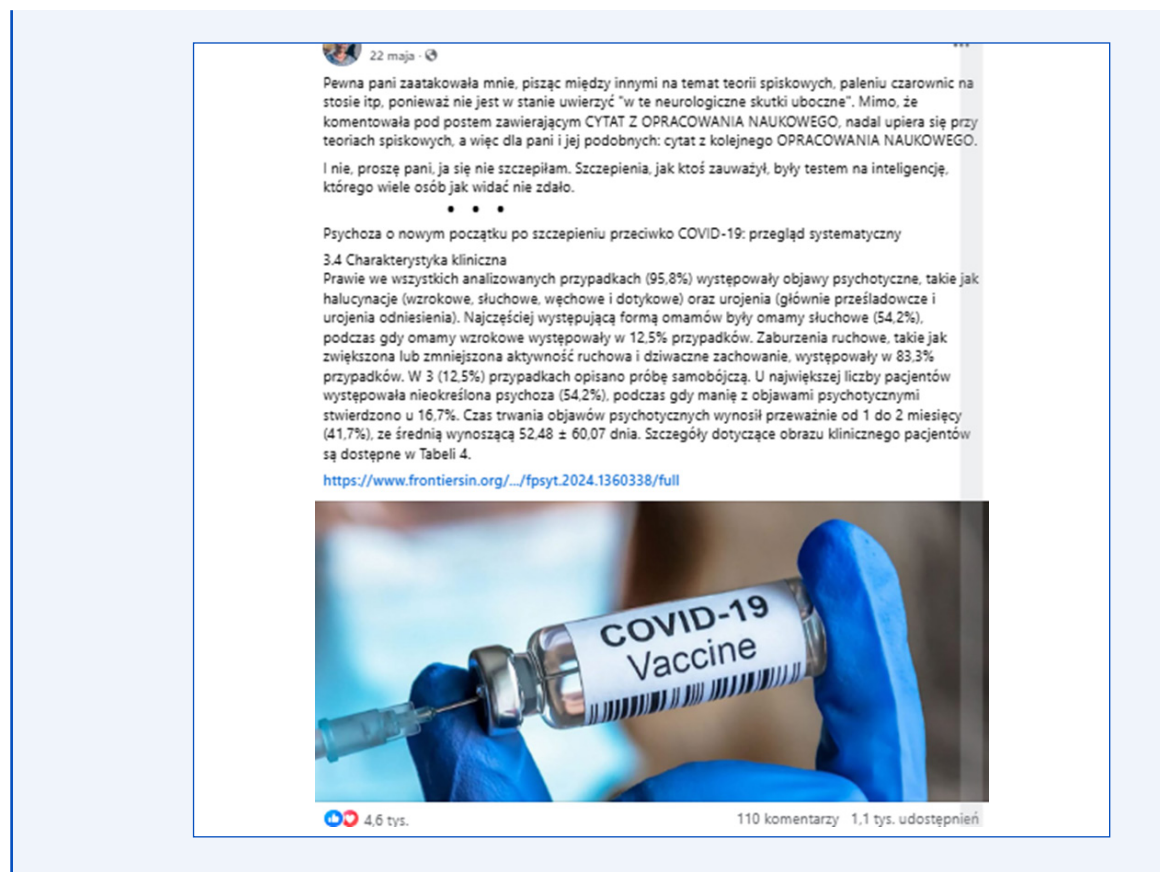
OBRAZ 51.

Wpis zawierający dezinformację na temat szczepień (zgłoszenie odrzucone)

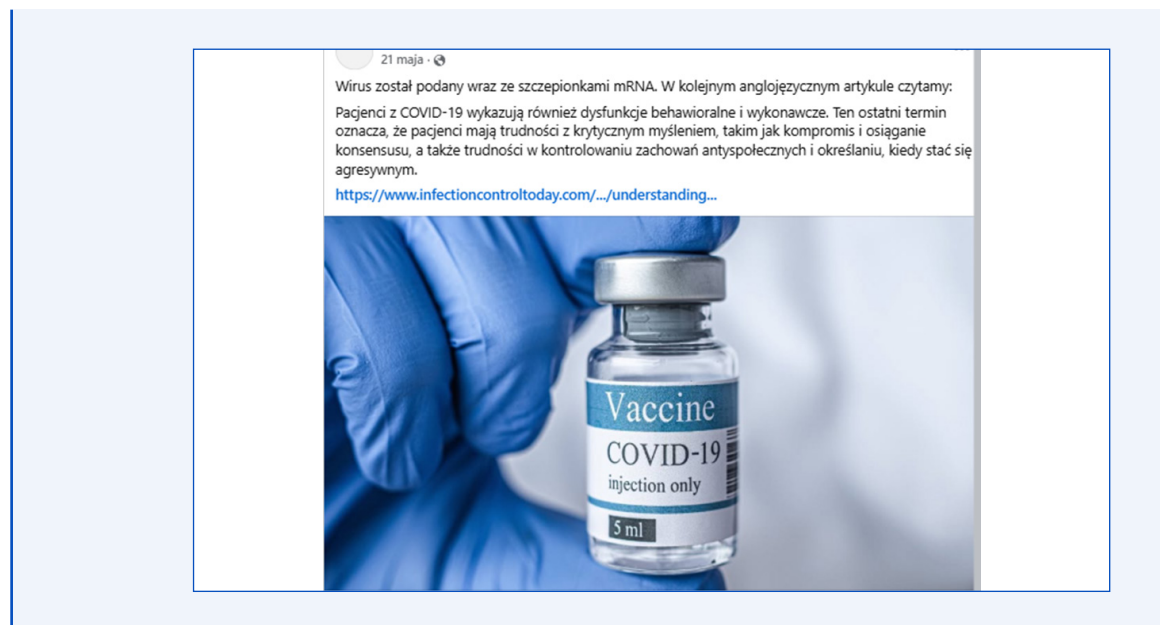
ŹRÓDŁO: FACEBOOK.COM

Wśród treści o charakterze antyszczepionkowym istotne miejsce zajmowały materiały łączące rzekome zagrożenia zdrowotne z teoriami spiskowymi dotyczącymi nowych technologii. Narracje te, pozbawione naukowych dowodów, zyskiwały popularność w mediach społecznościowych, a platformy społecznościowe nie podejmowały wystarczających działań, by ograniczyć ich rozpowszechnianie.

W części zgłaszanych materiałów sugerowano także, że szczepienia miały być testem na inteligencję społeczeństwa, a sama pandemia została wyolbrzymiona lub w pełni sfabrykowana.



OBRAZ 52. Wpis zawierający dezinformację na temat szczepień (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM



OBRAZ 53. Wpis zawierający dezinformację na temat szczepień (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM



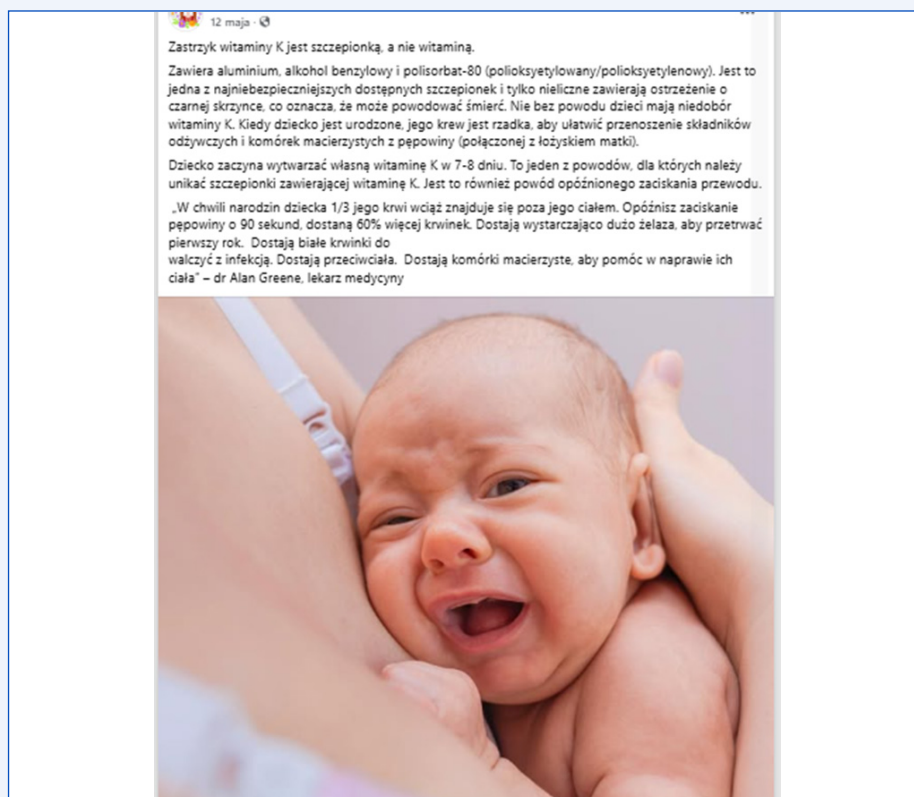
OBRAZ 54.

Wpis zawierający dezinformację na temat szczepień (zgłoszenie odrzucone)

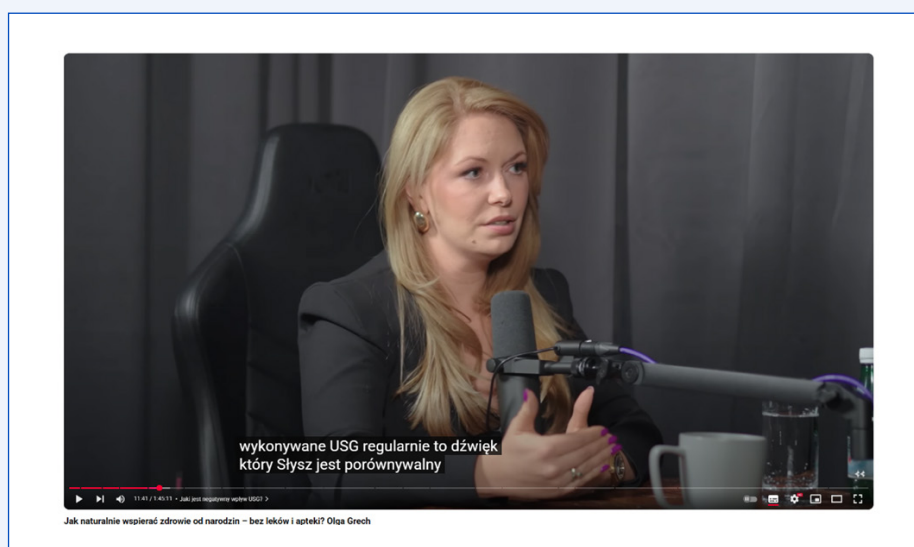
ŹRÓDŁO: FACEBOOK.COM

W obrębie dezinformacji medycznej zgłaszanej przez pracowników OAD NASK wyodrębnić można również kategorię treści związanych z tak zwaną medycyną naturalną i niekonwencjonalnymi metodami leczenia.

Narracje tego typu podkreślały rzekomą wyższość naturalnych metod nad procedurami medycyny konwencjonalnej i zachęcały do unikania leków oraz standardowych procedur medycznych. Wśród przedstawianych tez pojawiało się między innymi negowanie potrzeby podawania noworodkom witaminy K, krytyka badań USG, określanych jako niebezpieczne, sprzeciw wobec stosowania antybiotyków i farmakoterapii u kobiet w ciąży, a także ogólne podważanie skuteczności szczepień i leków konwencjonalnych.



OBRAZ 55. Wpis zawierający dezinformację na temat leków i procedur medycznych (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM



OBRAZ 56. Wpis zawierający dezinformację na temat leków i procedur medycznych (zgłoszenie odrzucone)
ŹRÓDŁO: YOUTUBE.COM

W ramach tego nurtu promowano również szereg alternatywnych praktyk i preparatów, które miałyby zastąpić oficjalne procedury medyczne. Przykłady obejmowały stosowanie oleju rycynowego jako rzekomo skutecznej metody leczenia nowotworów, leczenie autyzmu lekami przeciwgrzybiczymi lub probiotykami, a także kuracje fenbendazolem – lekiem przeciw pasożytniczym stosowanym u zwierząt, który miałby zmniejszać guzy nowotworowe, całkowicie je eliminować albo przynajmniej zatrzymywać rozwój choroby poprzez blokowanie Metabolizmu glukozy w komórkach rakowych. Jednocześnie narracje te promowały odrobaczanie oraz inne niekonwencjonalne praktyki jako kluczowe elementy wspierania odporności i profilaktyki nowotworowej, promując jednocześnie tezę, że medycyna konwencjonalna celowo ignoruje lub ukrywa takie metody.



OBRAZ 57.

Wpis zawierający dezinformację na temat leków i procedur medycznych (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



OBRAZ 58.

Wpis zawierający dezinformację na temat leków i procedur medycznych (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



OBRAZ 59.

Wpis zawierający dezinformację na temat leków i procedur medycznych (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

W zgłoszonych materiałach podważano także bezpieczeństwo i skuteczność standardowych procedur diagnostycznych i terapeutycznych. Sugero- wano, że biopsja może przyspieszać rozwój nowotworu, a konwencjonalne metody leczenia, w tym chemioterapia, radioterapia czy operacje chirurgiczne, prowadzą do powstawania przerzutów i dają jedynie fałszywą remisję. Twierdzono również, że mammografia jest szkodliwa oraz generuje fałszywie dodatnie wyniki.

4 lutego · 🌐

"Biopsja jest niemal gwarancją przerzutów i progresu choroby nowotworowej."
Dr Thomas Lodi, MD

Problem z biopsją jest taki, że kiedy guz nowotworowy rośnie, organizm zawiera go w osłonce fibrynowej. W chwili, gdy osłonka zostaje przerwana przez igłę biopsji, odizolowana wcześniej zawartość rozprzestrzenia się po całym organizmie.

Organizm robi wszystko, co może, aby odizolować zawartość guza nowotworowego od reszty ciała. Przerwanie tej izolacji może mieć śmiertelne konsekwencje.

Wymuszanie diagnozy histopatologicznej zamiast nieinwazyjnego alternatywnego badania diagnostycznego raka jest działaniem zbrodniczym. Jest częścią techniki sprzedaży opracowanej przez Big Pharma. Ta procedura diagnostyczna uzasadnia te konkretne zatwierdzone przez sponsorowane przez Big Pharmę leki na raka, które są opłacane skorumpowany system ubezpieczeń zdrowotnych. A lekarze twierdzą, że jest to właściwa rzecz i jedyna rzecz do zrobienia.

Czego lekarze ci nie mówią:

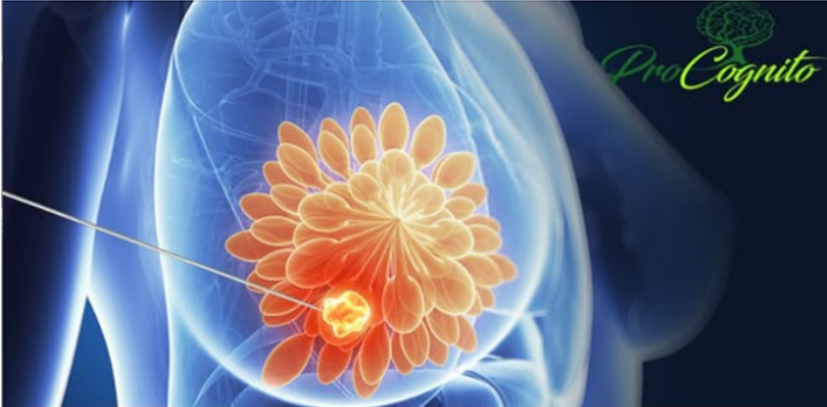
- ✖️ Badania pokazują, że te leki niemal gwarantują przerzuty. Dlatego w ulotkach uwzględniony jest rak jako skutek uboczny terapii przeciwnowotworowej tym lekiem.
- ✖️ Jeśli wykonasz biopsję i/lub operację, najprawdopodobniej będziesz miał przerzuty.
- ✖️ Jeśli poddasz się wysokim dawkom chemioterapii, najprawdopodobniej będziesz miał przerzuty.
- ✖️ Jeśli poddasz się radioterapii, najprawdopodobniej będziesz miał przerzuty.

Dają ci krótkoterminową redukcję pierwotnego nowotworu, osiągając kliniczną fałszywą remisję, a po 9 miesiącach rak powraca.

Holistyczny onkolog wykorzysta nieinwazyjne testy diagnostyczne z wykorzystaniem tomografii komputerowej, rezonansu magnetycznego, badań krwi, biomarkerów nowotworowych i termografii. Istnieje wiele sposobów diagnozy bez ryzyka rozprzestrzeniania się i przyspieszenia. Istnieją sposoby zapobiegania i leczenia raka bez ryzyka i szkód związanych z chemioterapią i radioterapią.

- 📌 Biopsja igłowa przyspiesza raka piersi
<https://www.sciencedirect.com/.../pii/S2666379123005475>
- 📌 Rozsiewanie raka. Tworzenie przerzutów
<https://pmc.ncbi.nlm.nih.gov/articles/PMC3473763/>
- 📌 Protokół leczenia raka: Keto/Post/Iwermektyna
<https://isom.ca/.../targeting-the-mitochondrial-stem.../>

#rak #nowotwór #nowotwory #biopsja #przerzuty



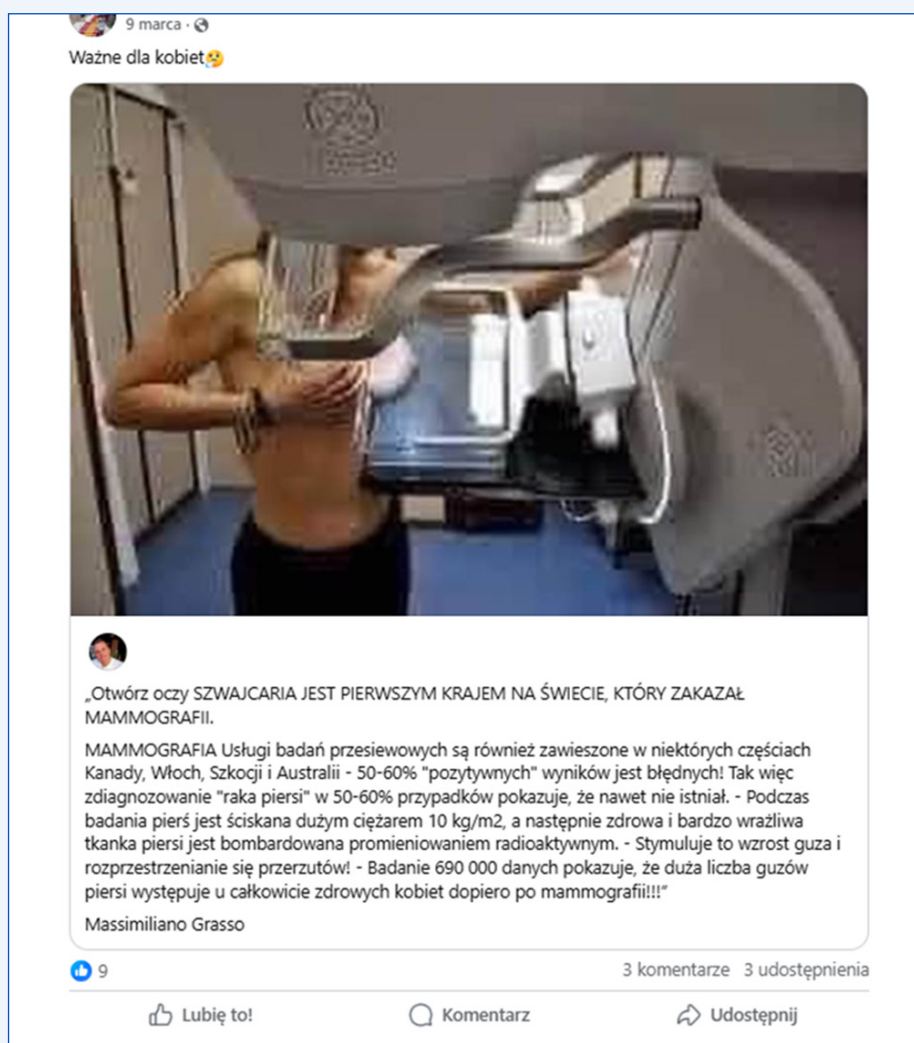
"Biopsja jest niemal gwarancją przerzutów i progresu choroby nowotworowej."
Dr Thomas Lodi, MD

Problem z biopsją jest taki, że kiedy guz nowotworowy rośnie, organizm otacza go osłonką fibrynową. W chwili, gdy osłonka

OBRAZ 60.

Wpis zawierający dezinformację na temat leków i procedur medycznych (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

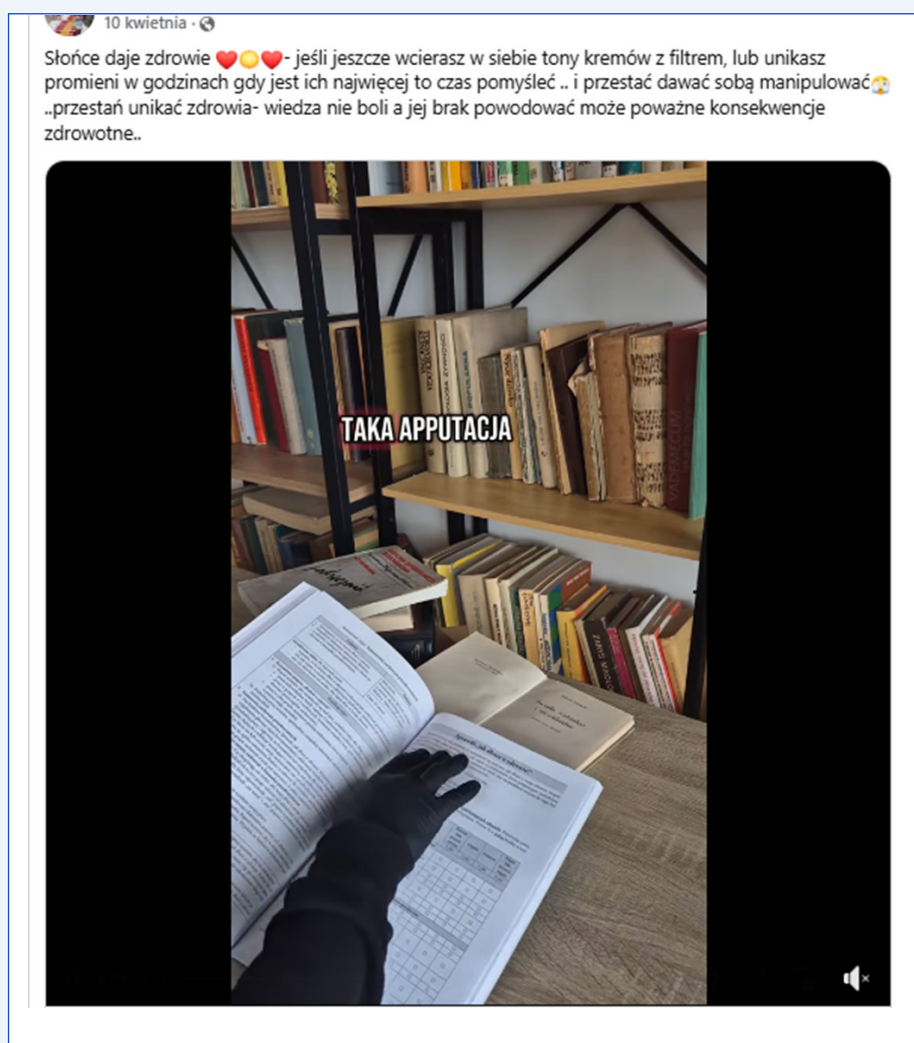


OBRAZ 61.

Wpis zawierający dezinformację na temat leków i procedur medycznych (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

W zgłoszeniach pojawiały się także treści promujące potencjalnie niebezpieczne zachowania prozdrowotne. Odradzały one między innymi stosowanie kremów z filtrem przeciwsłonecznym, zachęcając zamiast tego do przebywania na słońcu bez ochrony skóry, czemu przypisywano rzekome korzyści zdrowotne. Takie zalecenia stoją w kontrze do aktualnej wiedzy medycznej dotyczącej profilaktyki nowotworowej oraz ochrony przed promieniowaniem UV.



OBRAZ 62.

Wpis zawierający dezinformację na temat promieniowania UV (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

Wszystkie te treści kreowały obraz oficjalnych procedur medycznych jako zagrożenia, a alternatywnych metod diagnostycznych i terapeutycznych jako jedynej skutecznej formy terapii. Narracje te mogły wprowadzać użytkowników w błąd i skłaniać ich do rezygnacji z właściwego leczenia, stwarzając realne zagrożenie dla zdrowia i życia. Mimo tych zagrożeń i zgłoszeń ze strony analityków administratorzy platform społecznościowych nie podejmowali działań w celu usunięcia lub moderacji treści.

III kwartał 2025

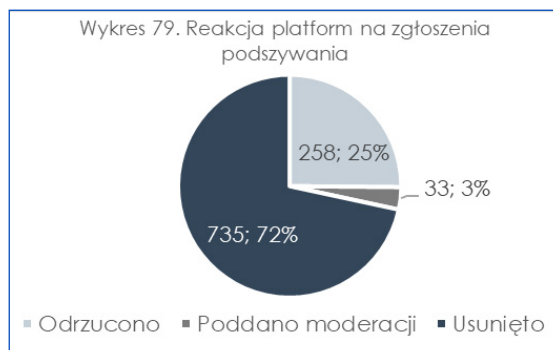
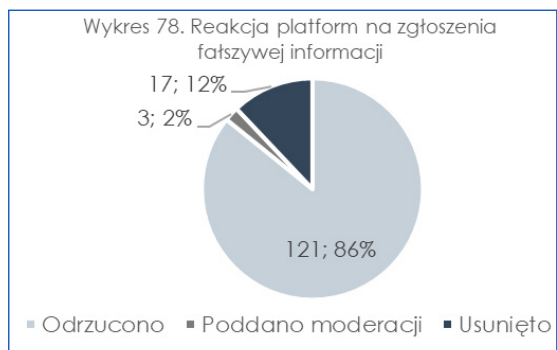
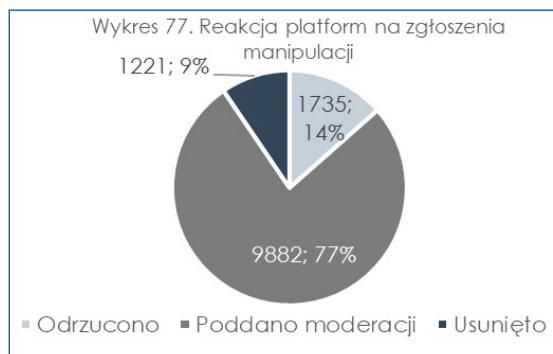
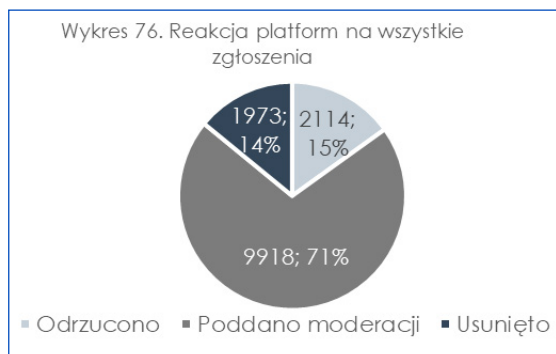
Q3.

III kwartał 2025

W III kwartale 2025 roku zgłoszono 14 005 treści. Platformy usunęły 1973 z nich, 9918 incydentów poddały moderacji, a 2114 materiałów pozostawiły bez żadnej reakcji. Główne wnioski są następujące:

- 15 proc. treści dezinformacyjnych nie spotkało się z żadną reakcją ze strony platform społecznościowych, co oznacza, że w polskiej infosferze nadal obecnych jest 2114 zgłoszonych w tym okresie materiałów, ocenionych przez OAD NASK jako naruszające regulaminy platform lub łamiące inne regulacje.
- W III kwartale 2025 roku spośród 14 005 zgłoszonych treści fałszywą informację zawierało 141 materiałów, 12 838 materiałów zawierało manipulacje, a 1026 incydentów związanych było z podszywaniem się pod osobę lub instytucję.
- Platformy najczęściej podejmowały działania w przypadku incydentów polegających na stosowaniu manipulacji. W analizowanym okresie poddano moderacji 77 proc. spośród nich (9882 materiałów), usunięto natomiast 9 proc. (1221 materiałów). Aktywnych pozostaje nadal 14 proc. zmanipulowanych treści (1735 przypadków).
- W przypadku kont podszywających się pod osoby publiczne lub instytucje usunięto 72 proc. zgłoszonych treści (735 materiałów). Moderacji poddano 3 proc. zgłoszonych treści, czyli 33 materiały. 25 proc. zgłoszeń (258 przypadków) zostało odrzuconych.
- Jeśli chodzi o wystąpienie fałszywych informacji, usunięto 12 proc. zgłoszonych treści (17 materiałów), a moderacji poddano 2 proc. zgłoszonych treści, czyli 3 materiały. Odrzucono 86 proc. zgłoszeń, a więc 121 przypadków.
- Zgłoszenia treści ocenionych przez OAD NASK jako naruszające regulaminy platform pochodzące z analizowanego okresu w wielu przypadkach zostały odrzucone bez podania powodu lub z adnotacją, że nie naruszają regulaminu danej platformy społecznościowej.
- Trudności w obsłudze zgłoszeń sprawiają konta publikujące reklamy z wykorzystaniem technologii deepfake. W takich przypadkach OAD NASK klasyfikuje je zarówno szerzenie dezinformacji, jak i podszywanie się pod osoby czy instytucje w celu wyłudzenia danych lub środków finansowych.
- Zgłoszenia tego typu działań często kończą się blokadą pojedynczego wpisu a nie całego profilu lub strony.

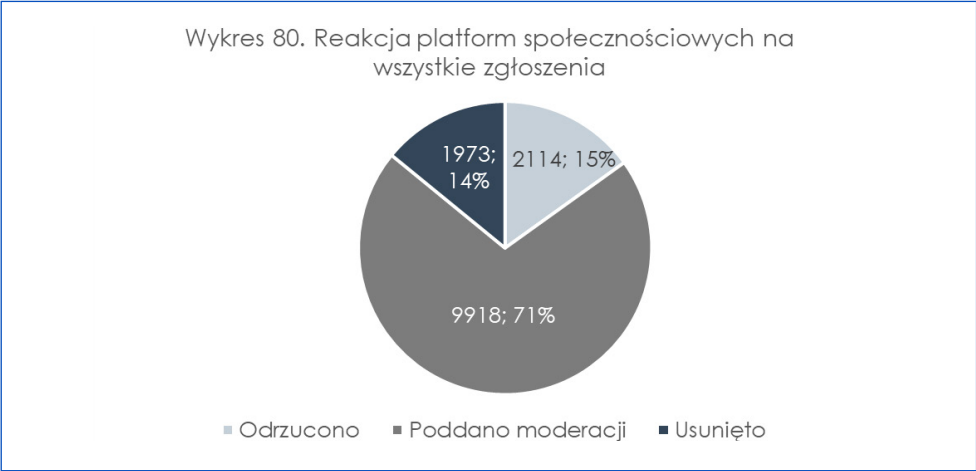
- Dodatkowym utrudnieniem związanym z kwestią zgłaszania szkodliwych reklam pojawiających się na platformach Facebook oraz Google Ads jest to, że ogólnodostępne biblioteki reklam są aktualizowane z opóźnieniem. Ograniczona jest także możliwość wyszukiwania i filtrowania treści zamieszczanych przez poszczególnych reklamodawców.
- Portal X wprowadził dedykowaną platformę zgłoszeniową przeznaczoną do raportowania treści nielegalnych. Działa ona skutecznie w przypadku kont podszywających się pod instytucje. X nie posiada natomiast systemu zgłoszeniowego, który umożliwiłby zgłaszanie treści noszących znamiona dezinformacji. W tym celu administracja zaleca zgłaszanie zapotrzebowania na dodanie community notes z pozycji użytkownika. Opcja ta dostępna jest wyłącznie dla kont zweryfikowanych numerem telefonu. Dodatkowo przykład reakcji na wpisy dezinformujące po naruszeniu polskiej przestrzeni powietrznej przez drony dowodzi, że sposób prezentacji i treść community notes nie są wyczerpujące i nie zapewniają odbiorcom informacji o rzeczywistym kontekście.
- Przypadek dezinformacji dotyczącej naruszenia polskiej przestrzeni powietrznej przez drony pokazał także ograniczoną reakcję platform w sytuacjach kryzysowych dotyczących bezpieczeństwa kraju i jego obywateli.
- Brak reakcji platform na działania takie jak kampania propagandowa Izraela skierowana do polskich internautów wskazuje na ograniczone możliwości reagowania na zagrożenia typu FIMI (z ang. *Foreign Information Manipulation and Interference*), prowadzone przez podmioty posiadające polityczny wpływ na koncerny technologiczne.



WYKRESY 76-79. Reakcja platform na zgłoszenia: wszystkie (76); manipulacja (77); fałszywa informacja (78); podszywanie (79)

ŹRÓDŁO: OPRACOWANIE WŁASNE

Reakcja platform na wszystkie zgłoszenia OAD NASK



WYKRES 80. Reakcja platform społecznościowych na wszystkie zgłoszenia
ŹRÓDŁO: OPRACOWANIE WŁASNE

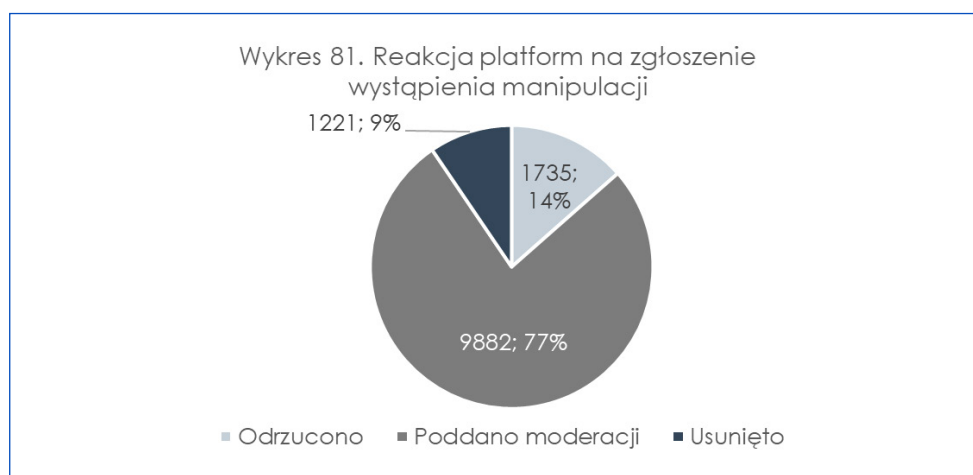
W III kwartale 2025 roku z platform społecznościowych usunięto 14 proc. zgłoszonych przez OAD NASK treści, czyli 1973 wpisy. Dalszych 71 proc. zgłoszonych treści, czyli 9918 wpisów, zostało poddanych moderacji, natomiast 15 proc. zgłoszeń, czyli 2114 przypadki, nie spotkało się z żadną widoczną reakcją ze strony moderacji.

Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/institucję
Liczba zgłoszeń odrzuconych	121	1735	258
Liczba zgłoszeń poddanych moderacji	3	9882	33
Liczba zgłoszeń usuniętych	17	1221	735
Całkowita liczba zgłoszeń	141	12838	1026

TABELA 8. Rodzaj naruszenia w przekazanych zgłoszeniach
ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszenia. Wśród wszystkich zgłoszeń zidentyfikowano:

- **1 proc. (141 przypadków)** wpisów zawierających fałszywe informacje;
- **92 proc. (12 838 przypadków)** wpisów zawierających manipulacje;
- **7 proc. (1026 przypadków)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

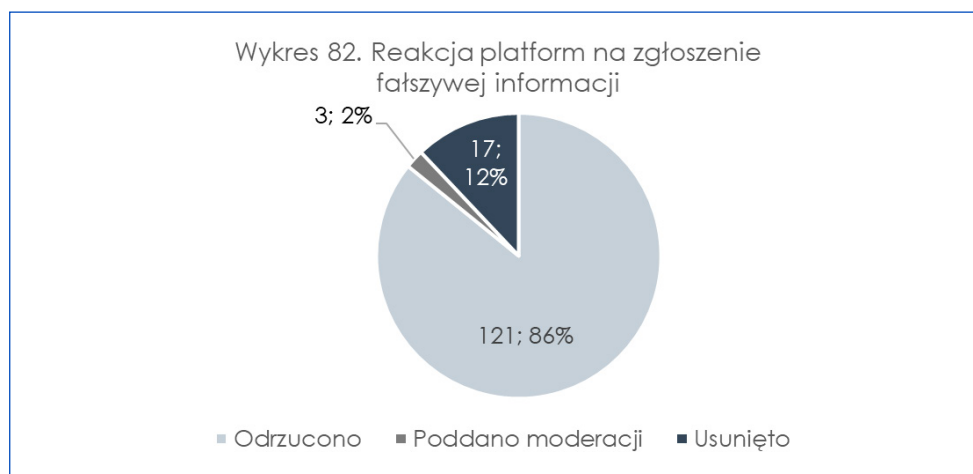


WYKRES 81.

Reakcja platform na zgłoszenie wystąpienia manipulacji

ŹRÓDŁO: OPRACOWANIE WŁASNE

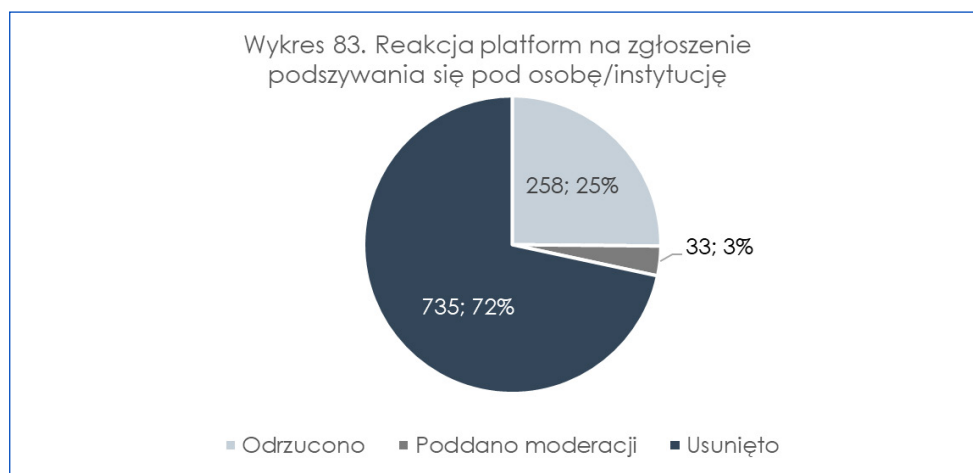
Analiza zgłoszonych treści wykazała, że platformy społecznościowe najczęściej reagowały na incydenty polegające na publikowaniu zmanipulowanych informacji. Usunięto 9 proc. takich treści, a więc 1221 materiałów, zaś 77 proc. zgłoszonych treści, czyli 9882 zgłoszenia, poddano moderacji. Bez reakcji pozostawiono natomiast 14 proc. zgłoszeń, czyli 1735 przypadków. Jako incydent wystąpienia manipulacji zakwalifikowane zostały między innymi treści publikowane przez konta-boty, wykazujące oznaki automatycznego działania.



WYKRES 82. Reakcja platform na zgłoszenie fałszywej informacji

ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku wystąpienia fałszywych informacji usunięto 12 proc. zgłoszonych treści, a więc 17 materiałów. Moderacji z kolei poddano 2 proc. zgłoszonych treści (3 materiały). Odrzucono 86 proc. zgłoszeń, czyli 121 przypadków.

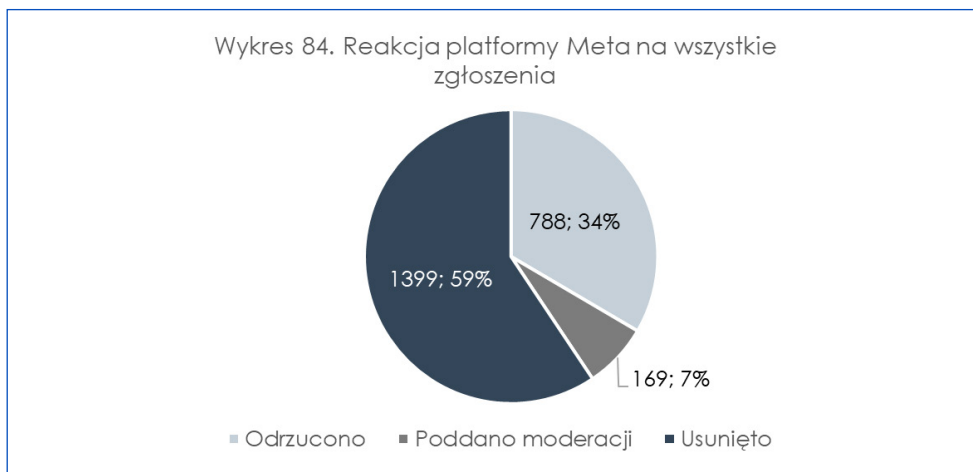


WYKRES 83. Reakcja platform na zgłoszenie podszywania się pod osobę/institucję

ŹRÓDŁO: OPRACOWANIE WŁASNE

Jeżeli chodzi o moderację treści publikowanych przez konta podszywające się pod osoby publiczne lub instytucje, platformy społecznościowe usunęły 735 zgłoszonych materiałów (72 proc.), natomiast moderacji poddały 33 materiały (3 proc. zgłoszeń). Bez reakcji pozostawiono 258 materiałów (25 proc. zgłoszonych treści).

Meta

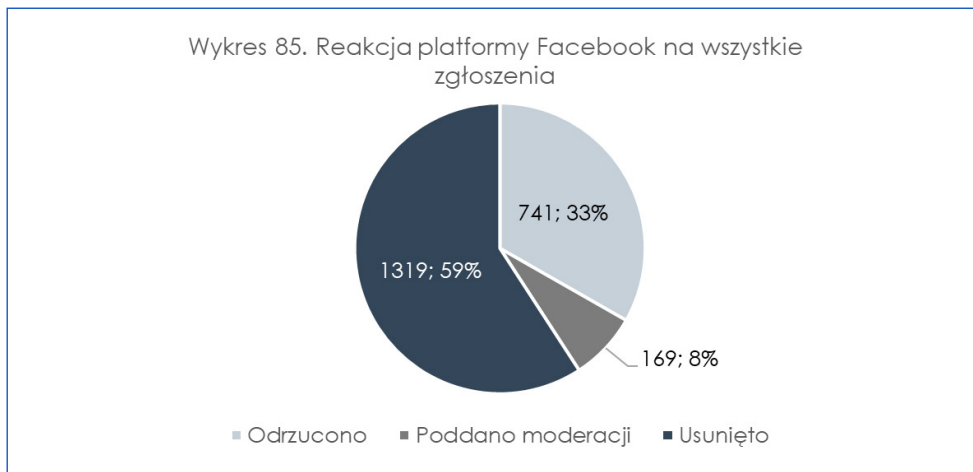


WYKRES 84. Reakcja platformy Meta na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszonych do administracji platform firmy Meta treści pochodzących z serwisów Facebook i Instagram. Administracja usunęła łącznie 59 proc. zgłoszonych treści, czyli 1399 materiałów, 7 proc., czyli 169 materiałów, poddano moderacji, a 34 proc. zgłoszeń, czyli 788 przypadków, zostało odrzuconych.

Facebook



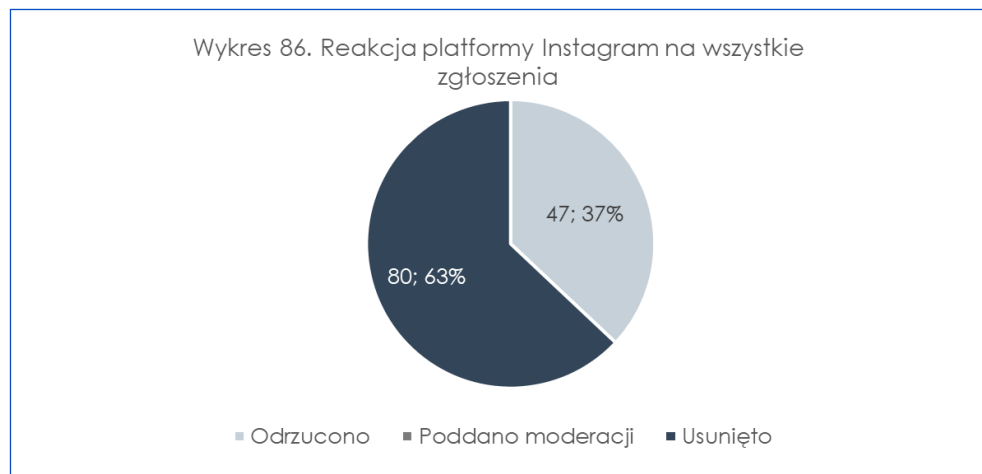
WYKRES 85. Reakcja platformy Facebook na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

W analizowanym okresie platforma Facebook usunęła 59 proc. zgłoszonych treści, czyli 1319 materiałów, a 8 proc. zgłoszonych treści, czyli 169 zgłoszeń,

zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 33 proc. zgłoszeń, czyli 741 przypadków.

Instagram

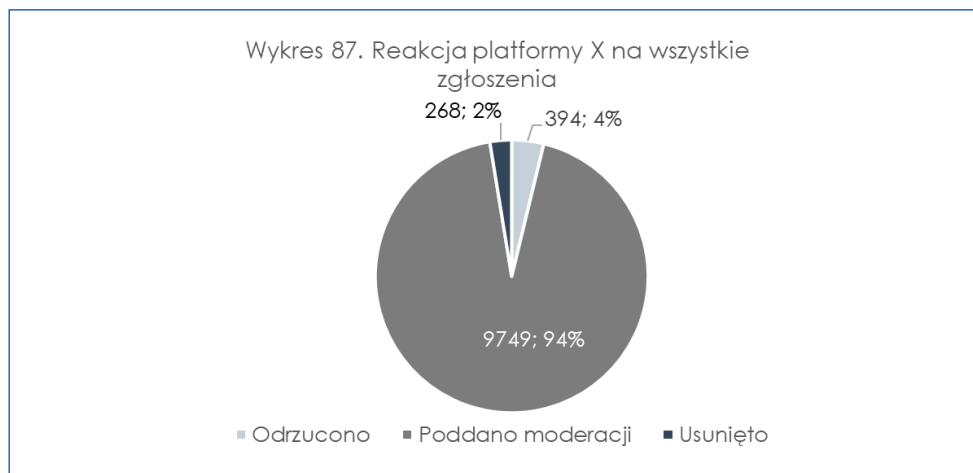


WYKRES 86. Reakcja platformy Instagram na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram zdecydowała o usunięciu 63 proc. zgłoszonych treści, czyli 80 materiałów. Nie usunięto natomiast 37 proc. zgłoszonych treści, czyli 46 materiałów.

X.com

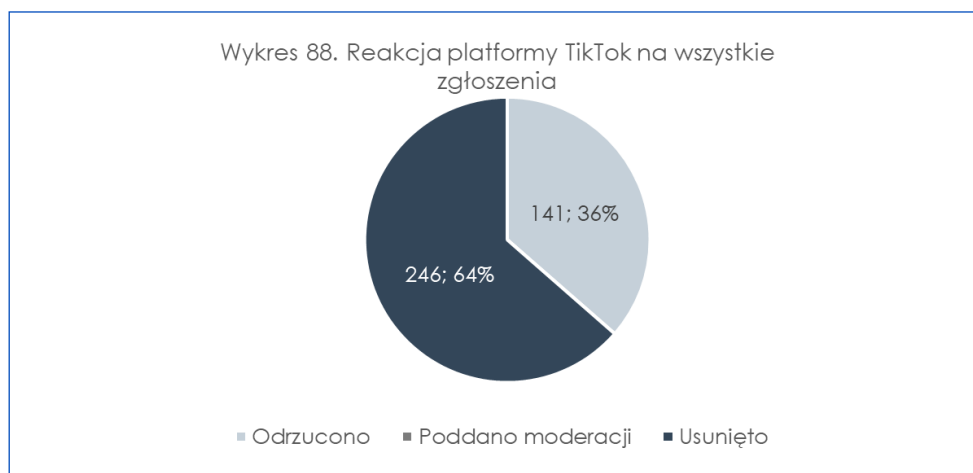


WYKRES 87. Reakcja platformy X na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 2 proc. zgłoszonych treści, czyli 268 materiałów, a 94 proc. zgłoszonych treści, czyli 9749 zgłoszeń, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 4 proc. zgłoszeń, czyli 394 przypadków.

TikTok

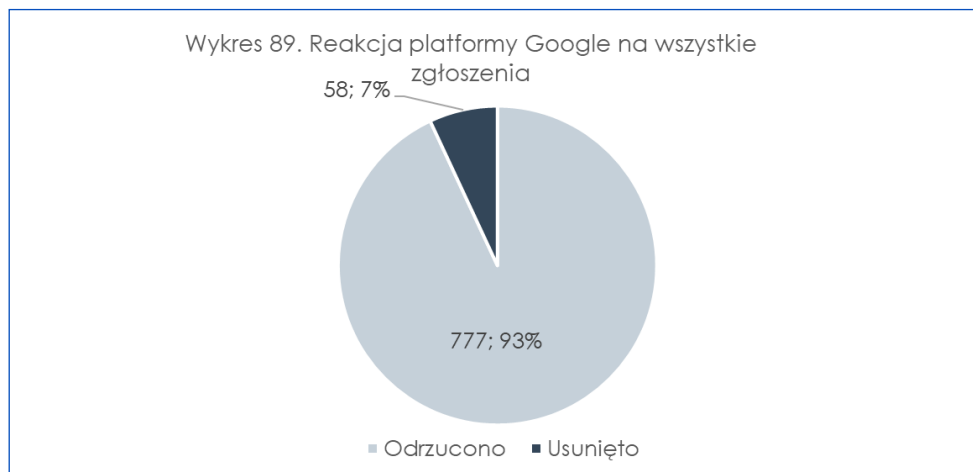


WYKRES 88. Reakcja platformy TikTok na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok zdecydowała o usunięciu 64 proc. zgłoszonych treści, czyli 246 materiałów. Nie usunięto natomiast 36 proc. zgłoszonych treści, czyli 141 materiałów.

Google (YouTube i Google Ads)



WYKRES 89. Reakcja platformy Google na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 7 proc. zgłoszonych treści, czyli 58 materiałów. Nie usunięto natomiast 93 proc. zgłoszonych treści, czyli 777 materiałów.

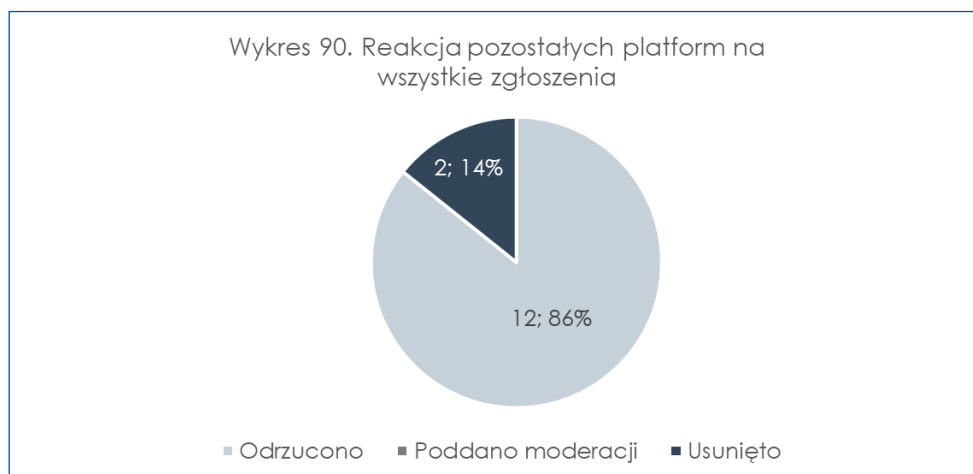
Telegram

Administracja platformy Telegram odrzuciła 1 przesłane do niej zgłoszenie.

Discrod

Administracja platformy Discrod odrzuciła 1 przesłane do niej zgłoszenie.

Pozostałe platformy

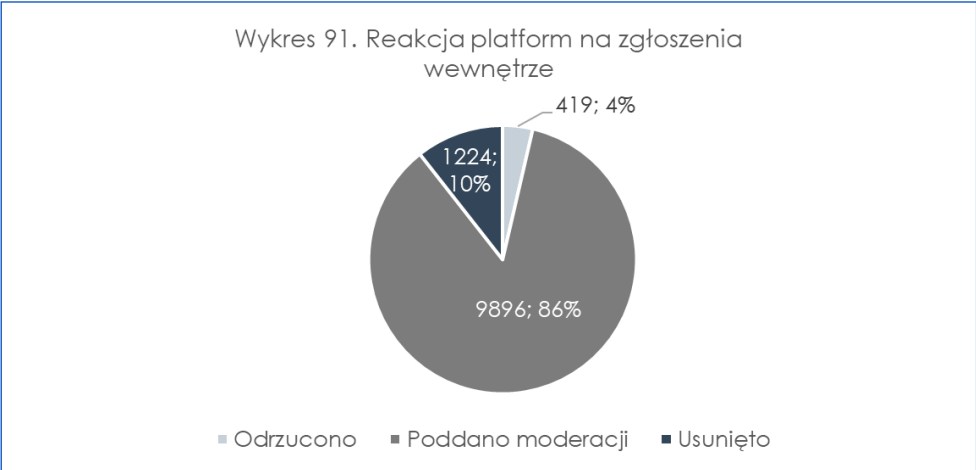


WYKRES 90. Reakcja pozostałych platform na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Pozostałe platformy usunęły 14 proc. zgłoszonych treści, czyli 2 materiały. Nie usunięto natomiast 86 proc. zgłoszonych treści, czyli 12 materiałów.

Zgłoszenia wewnętrzne



WYKRES 91. Reakcja platform na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku zgłoszeń wewnętrznych (pochodzących z bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej) w III kwartale 2025 roku z platform społecznościowych usunięto 10 proc. zgłoszonych treści, czyli 1224 wpisy. Dalszych 86 proc. zgłoszonych treści, czyli 9896 wpisów, zostało poddanych moderacji, natomiast 4 proc. zgłoszeń, czyli 419 przypadków, nie spotkało się z żadną widoczną reakcją ze strony moderacji.

Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń usuniętych	3	1078	143
Liczba zgłoszeń odrzuconych	31	345	43
Liczba zgłoszeń poddanych moderacji	0	9872	24
Całkowita liczba zgłoszeń	34	11295	210

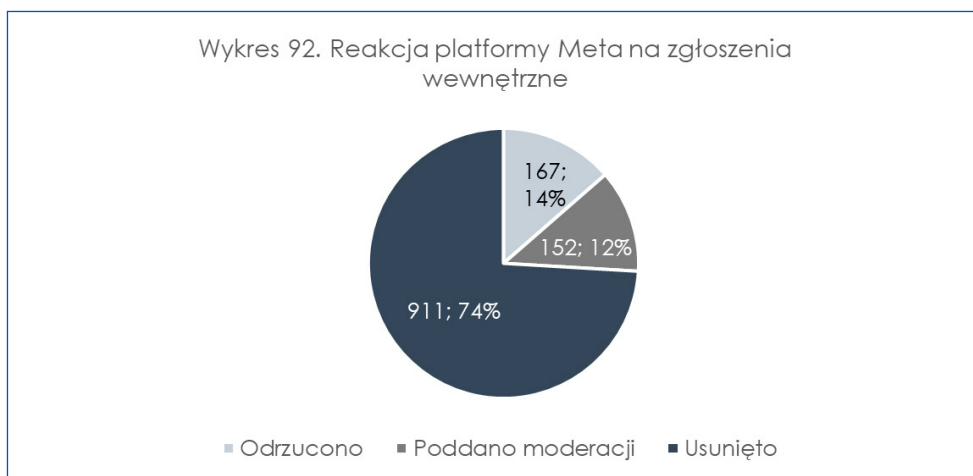
TABELA 9. Rodzaj przypisanego naruszenia w zgłoszeniach wewnętrznych

ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń. Wśród wszystkich zgłoszeń zidentyfikowano:

- **1,4 proc. (259 przypadków)** wpisów zawierających fałszywe informacje;
- **1,2 proc. (217 przypadków)** wpisów stanowiących możliwe naruszenie ciszy wyborczej;
- **93,7 proc. (17 517 przypadków)** wpisów zawierających manipulacje;
- **3,7 proc. (694 przypadków)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

Meta

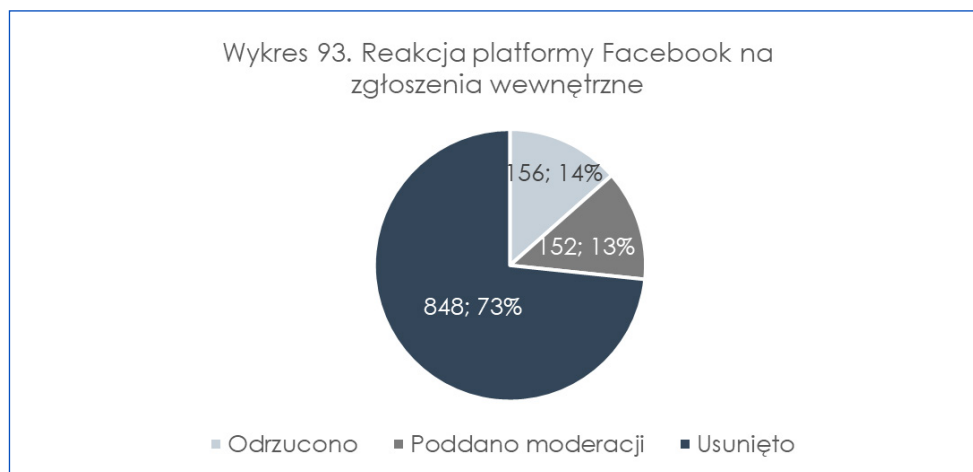


WYKRES 92. Reakcja platformy Meta na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszeń wewnętrznych przekazanych do administracji platform firmy Meta w związku z treściami pochodzącymi z serwisów Facebook i Instagram. Administracja usunęła łącznie 74 proc. zgłoszonych treści, czyli 911 materiałów, 12 proc., czyli 152 materiały, poddano moderacji, a 14 proc. zgłoszeń, czyli 167 przypadków, zostało odrzuconych.

Facebook



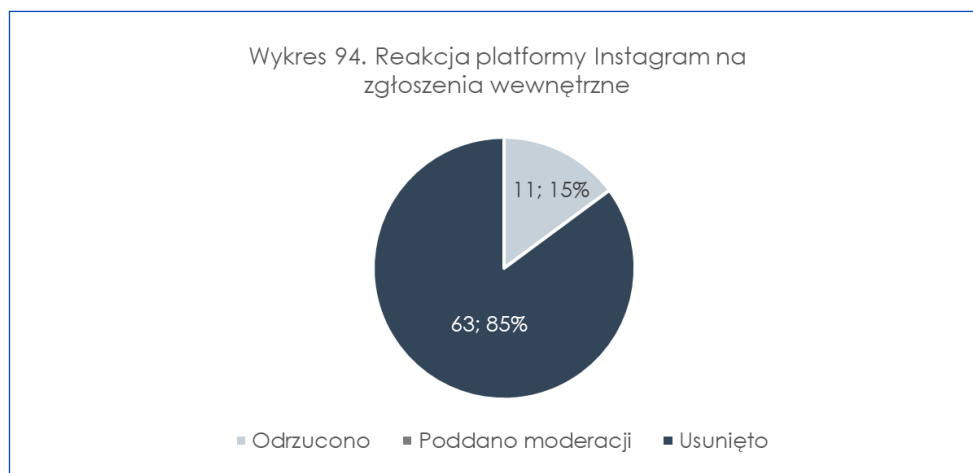
WYKRES 93. Reakcja platformy Facebook na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Facebook usunęła 73 proc. zgłoszonych treści, czyli 848 materiałów, a 13 proc. zgłoszonych treści, czyli 152 zgłoszenia, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 14 proc. zgłoszeń, czyli 156 przypadków.

Instagram

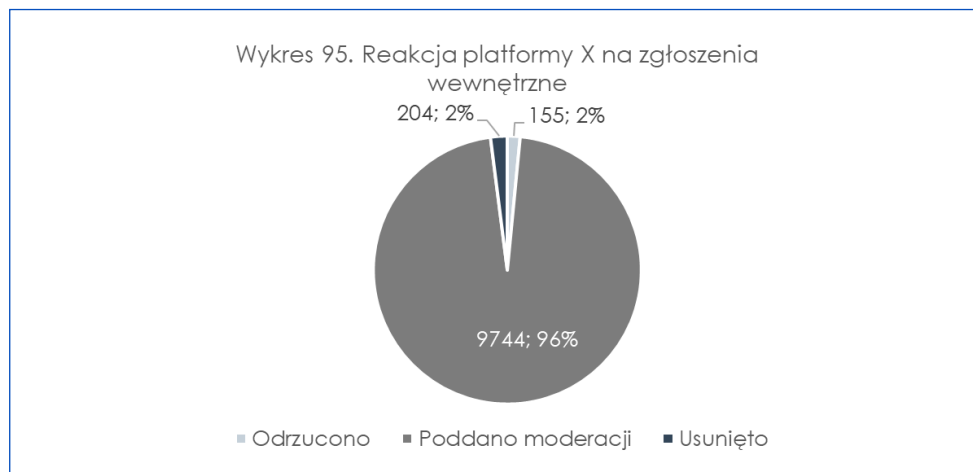
WYKRES 94. Reakcja platformy Instagram na zgłoszenia wewnętrzne



ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram zdecydowała o usunięciu 85 proc. zgłoszonych treści, czyli 63 materiałów. Nie usunięto natomiast 15 proc. zgłoszonych treści, czyli 11 materiałów.

X.com

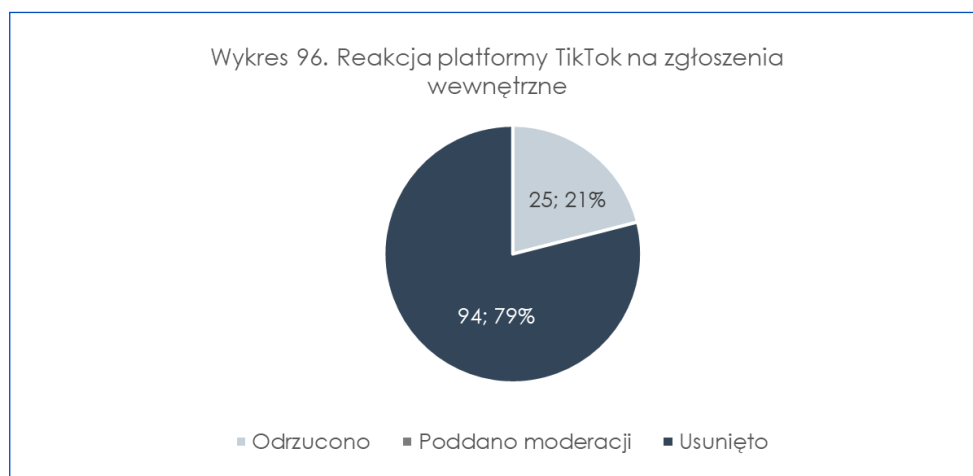


WYKRES 95. Reakcja platformy X na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 2 proc. zgłoszonych treści, czyli 204 materiały, a 96 proc. zgłoszonych treści, czyli 9744 zgłoszenia, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 2 proc. zgłoszeń, czyli 155 przypadków.

TikTok

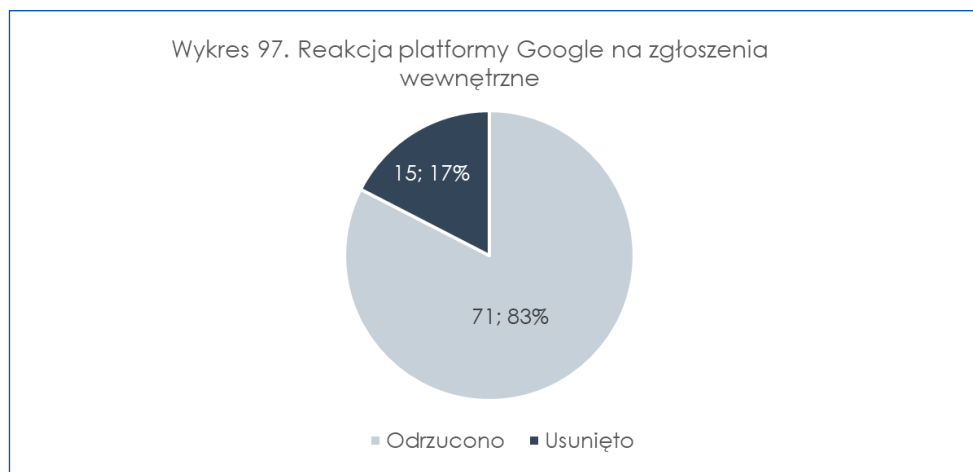


WYKRES 96. Reakcja platformy TikTok na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok zdecydowała o usunięciu 79 proc. zgłoszonych treści, czyli 94 materiałów. Nie usunięto natomiast 21 proc. zgłoszonych treści, czyli 25 materiałów.

Google (YouTube i Google Ads)



WYKRES 97. Reakcja platformy Google na zgłoszenia wewnętrzne

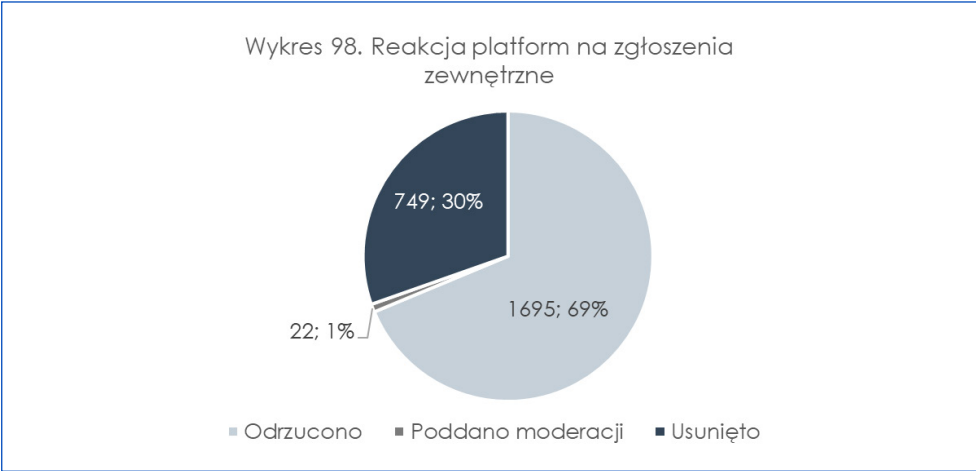
ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 17 proc. zgłoszonych treści, czyli 15 materiałów. Nie usunięto natomiast 83 proc. zgłoszonych treści, czyli 71 materiałów.

Pozostałe platformy

Administratorzy pozostałych platform odrzucili 1 przesłane do nich zgłoszenie.

Zgłoszenia zewnętrzne



WYKRES 98. Reakcja platform na zgłoszenia zewnętrzne
ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku zgłoszeń zewnętrznych (przekazywanych do Zespołu Szybkiego Reagowania i Wykrywania Dezinformacji przez obywateli i obywatelki, podmioty publiczne oraz inne instytucje) w III kwartale 2025 roku z platform społecznościowych usunięto 30 proc. zgłoszonych treści, czyli 749 wpisów. Dalszy 1 proc. zgłoszonych treści, czyli 22 wpisy zostały poddane moderacji, natomiast 69 proc. zgłoszeń, czyli 1695 przypadków, nie spotkało się z żadną widoczną reakcją ze strony moderacji. 1427 materiałów zgłoszonych do OAD NASK zostało ocenionych jako niezasadne, czyli takie, które nie stanowiły naruszenia regulaminów bądź innych regulacji prawnych.

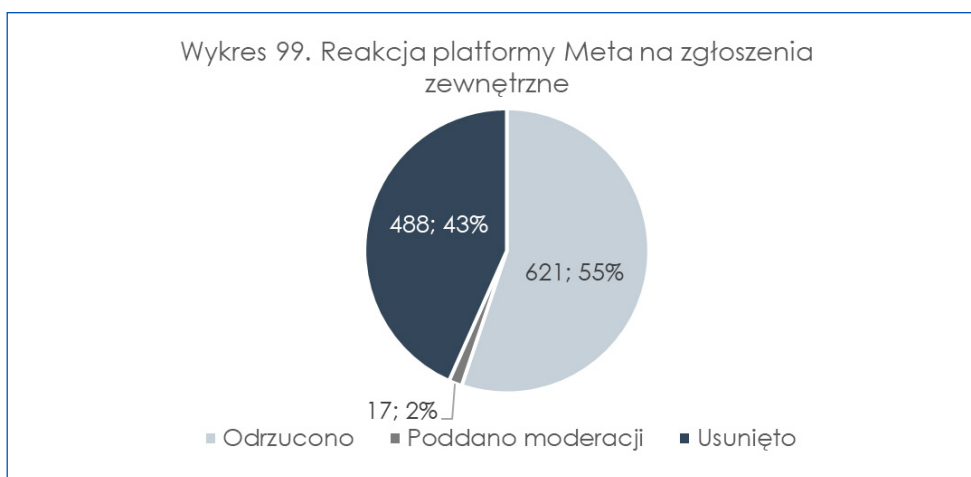
Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń usuniętych	14	143	592
Liczba zgłoszeń odrzuconych	90	1390	215
Liczba zgłoszeń poddanych moderacji	3	10	9
Całkowita liczba zgłoszeń	107	1543	816

TABELA 10. Rodzaj naruszenia w zgłoszeniach zewnętrznych
ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszenia¹⁹. Wśród zgłoszeń zewnętrznych zidentyfikowano:

- 4 proc. (107 przypadków) wpisów zawierających fałszywe informacje;
- 63 proc. (1543 przypadki) wpisów zawierających manipulacje;
- 33 proc. (816 przypadków) wpisów związanych z podszywaniem się pod osobę lub instytucję.

Meta



WYKRES 99.

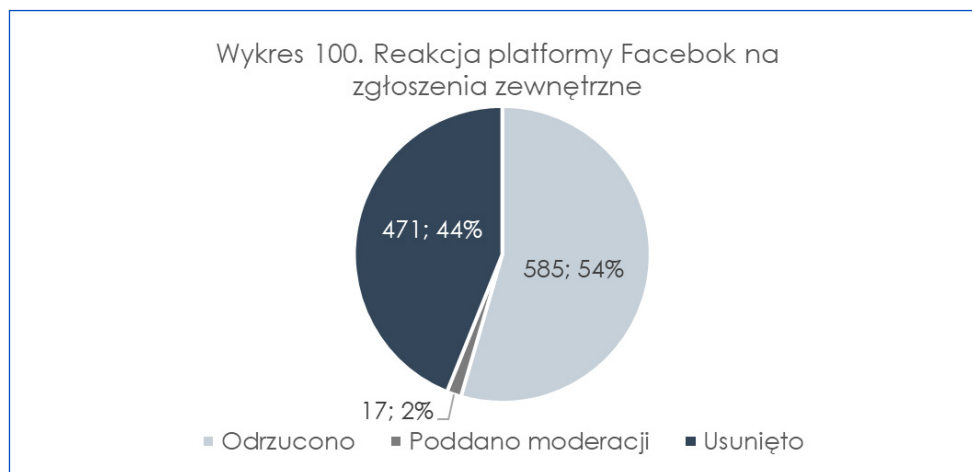
Reakcja platformy Meta na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszeń zewnętrznych przekazanych do administracji platform firmy Meta w związku z treściami pochodzącymi z serwisów Facebook i Instagram. Administracja usunęła łącznie 43 proc. zgłoszonych treści, czyli 488 materiałów, 2 proc., czyli 17 materiałów, poddano moderacji, a 55 proc. zgłoszeń, czyli 621 przypadków, zostało odrzuconych.

¹⁹ Podział na wskazanej kategorii został dokonany wewnętrznie przez Zespół Szybkiego Reagowania i Wykrywania Dezinformacji w celu uproszczenia procedury realizacji zgłoszeń.

Facebook

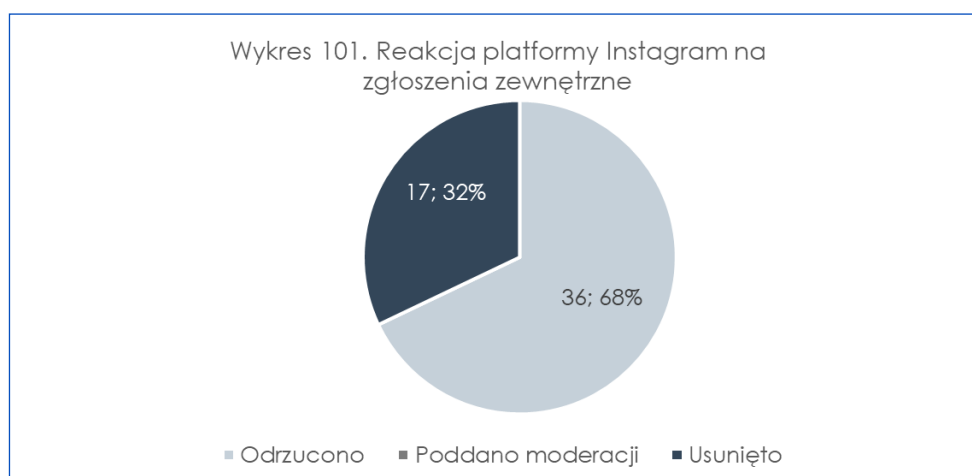


WYKRES 100. Reakcja platformy Facebook na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Facebook usunęła 44 proc. zgłoszonych treści, czyli 471 materiałów, a 2 proc. zgłoszonych treści, czyli 17 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 54 proc. zgłoszeń, czyli 585 przypadków.

Instagram

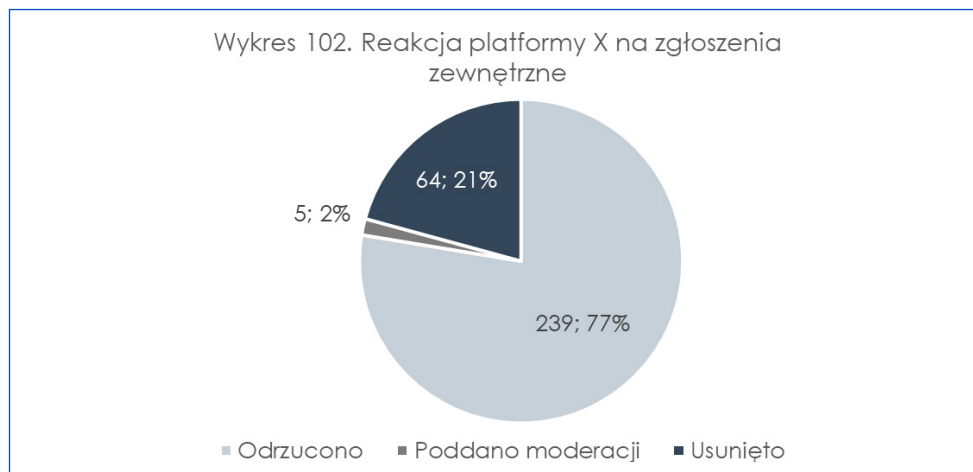


WYKRES 101. Reakcja platformy Instagram na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram zdecydowała o usunięciu 17 proc. zgłoszonych treści, czyli 32 materiałów. Nie usunięto natomiast 68 proc. zgłoszonych treści, czyli 36 materiałów.

X.com

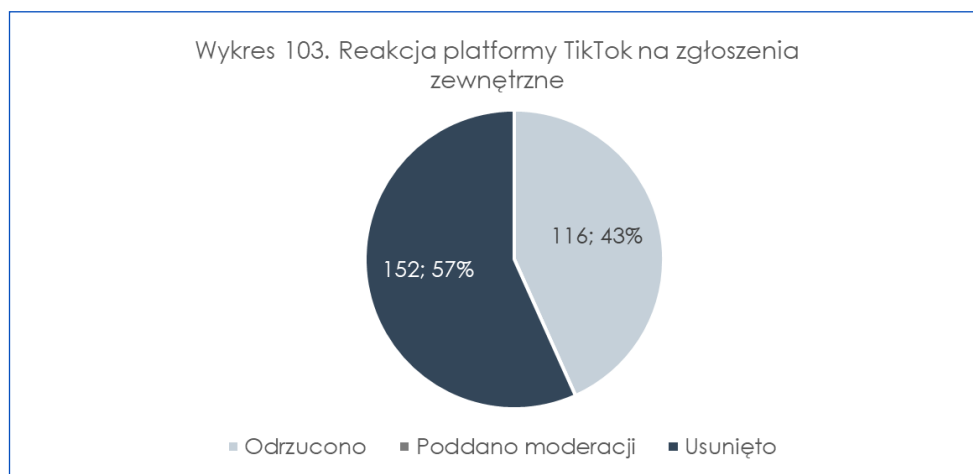


WYKRES 102. Reakcja platformy X na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 21 proc. zgłoszonych treści, czyli 64 materiały, a 2 proc. zgłoszonych treści, czyli 5 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 77 proc. zgłoszeń, czyli 239 przypadków.

TikTok

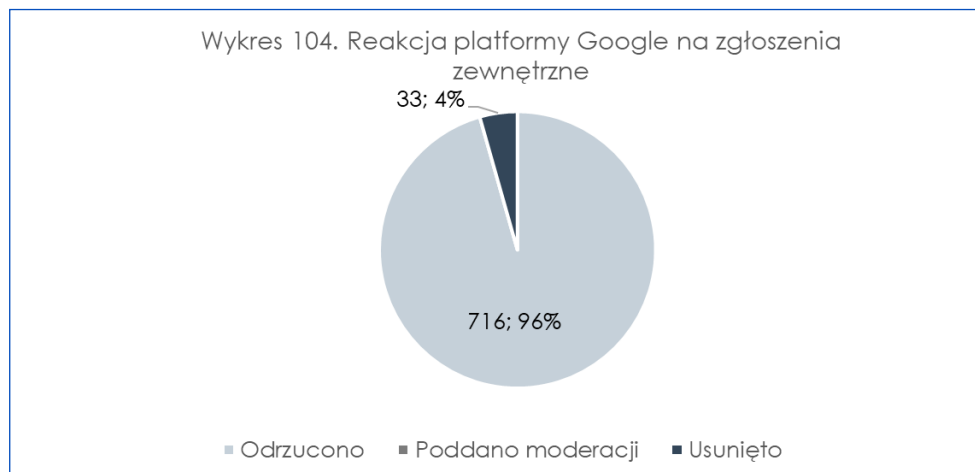


WYKRES 103. Reakcja platformy TikTok na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok zdecydowała o usunięciu 57 proc. zgłoszonych treści, czyli 152 materiałów. Nie usunięto natomiast 43 proc. zgłoszonych treści, czyli 116 materiałów.

Google (YouTube i Google Ads)



WYKRES 104. Reakcja platformy Google na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 4 proc. treści, co stanowi 33 materiały. Nie usunięto natomiast 96 proc. treści, czyli 716 materiałów.

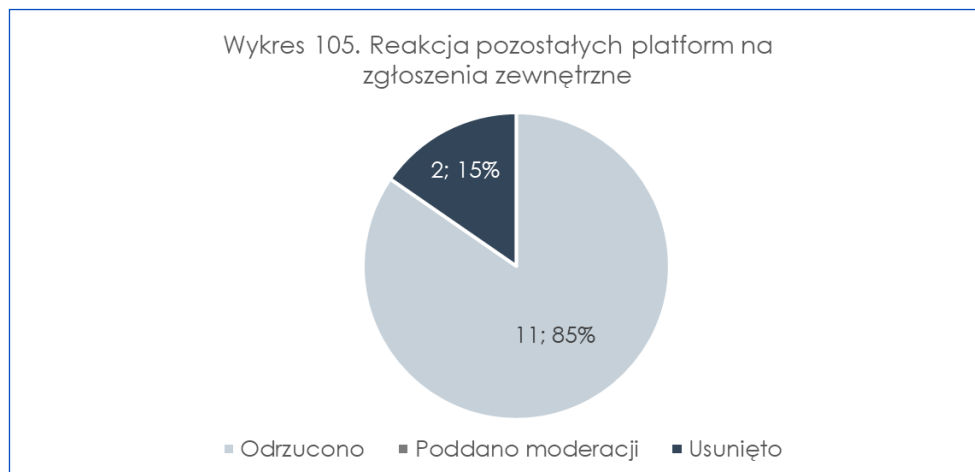
Telegram

Platforma Telegram odrzuciła 1 przesłane do niej zgłoszenie.

Discrod

Platforma Discrod odrzuciła 1 przesłane do niej zgłoszenie.

Pozostałe platformy



WYKRES 105. Reakcja pozostałych platform na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Pozostałe platformy usunęły 15 proc. zgłoszonych treści, czyli 2 materiały. Nie usunięto natomiast 85 proc. zgłoszonych treści, czyli 11 materiałów.

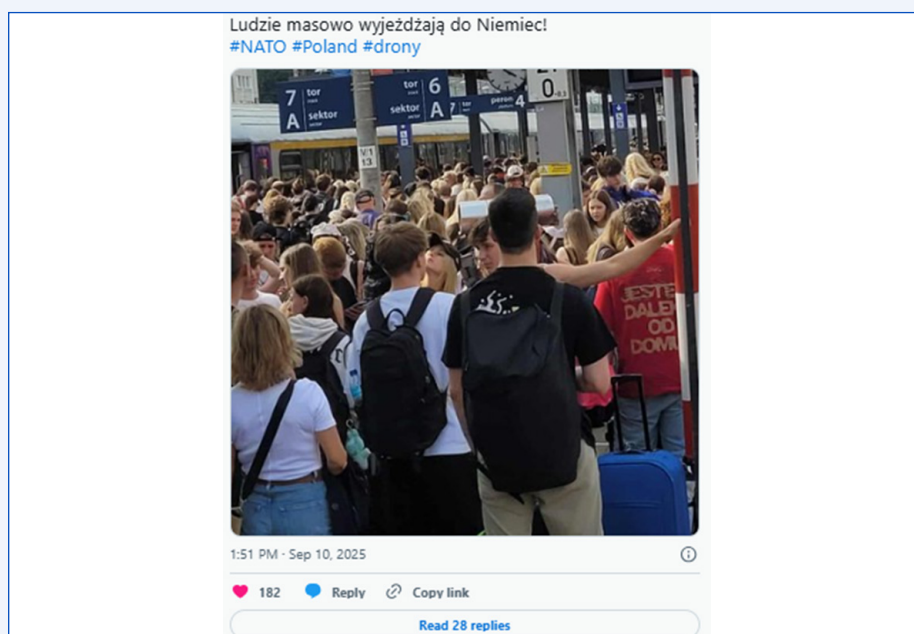
Case studies

Dezinformacja na temat incydentu z dronami

Wnocy z 9 na 10 września 2025 roku miał miejsce incydent naruszenia polskiej przestrzeni powietrznej przez obce drony. W związku z tym wydarzeniem w polskojęzycznej infosferze pojawiło się wiele narracji dezinformacyjnych.

1. Narracje przedstawiające Polskę jako terytorium zbrojnych działań wojennych

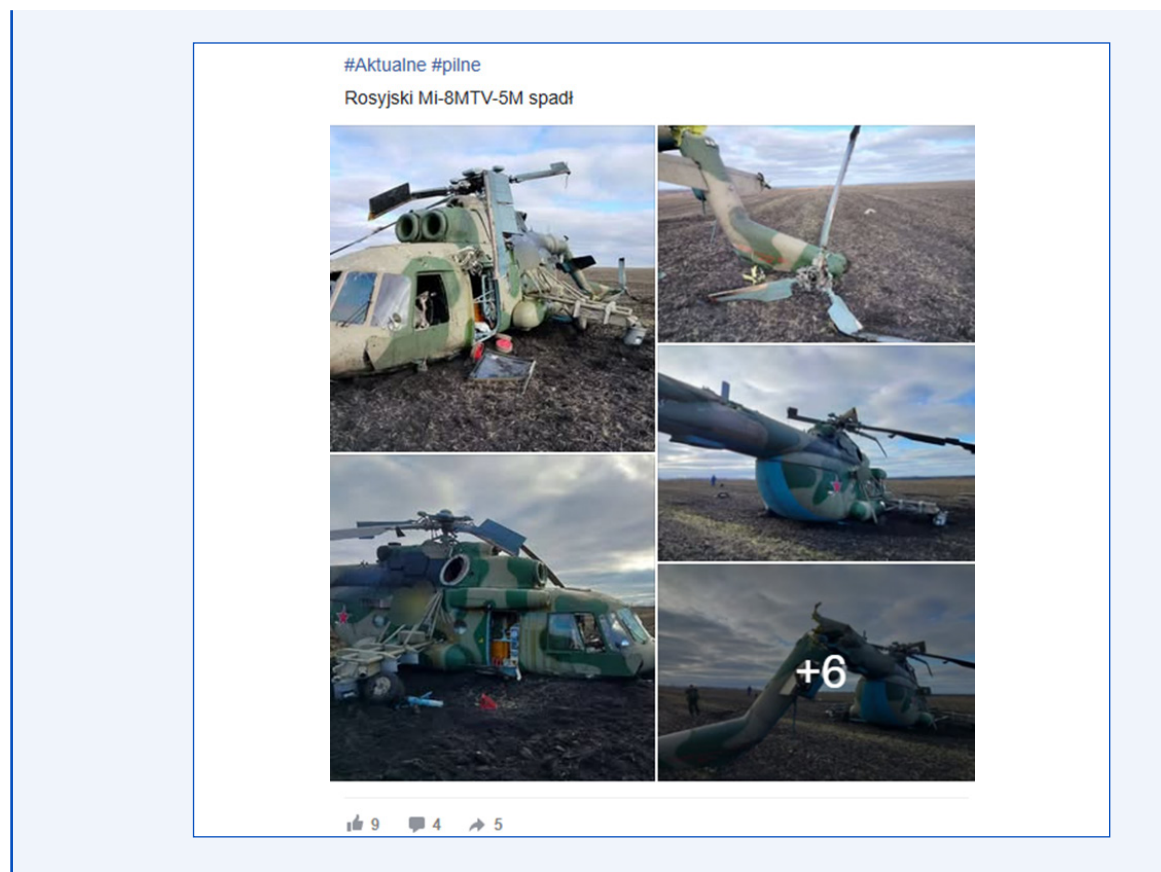
Do szczególnie niebezpiecznych treści należały te, które sugerowały, że w Polsce rozpoczęła się wojna, i inne fałszywe informacje na temat bieżącej sytuacji militarnej. Mimo zgłoszeń dokonanych przez zespół OAD NASK do administracji portali społecznościowych wiele publikacji pozostawało dostępnych lub poddanych moderacji jedynie w ograniczonym zakresie.



OBRAZ 63.

Wpis przedstawiający Polskę jako terytorium zbrojnych działań wojennych (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM



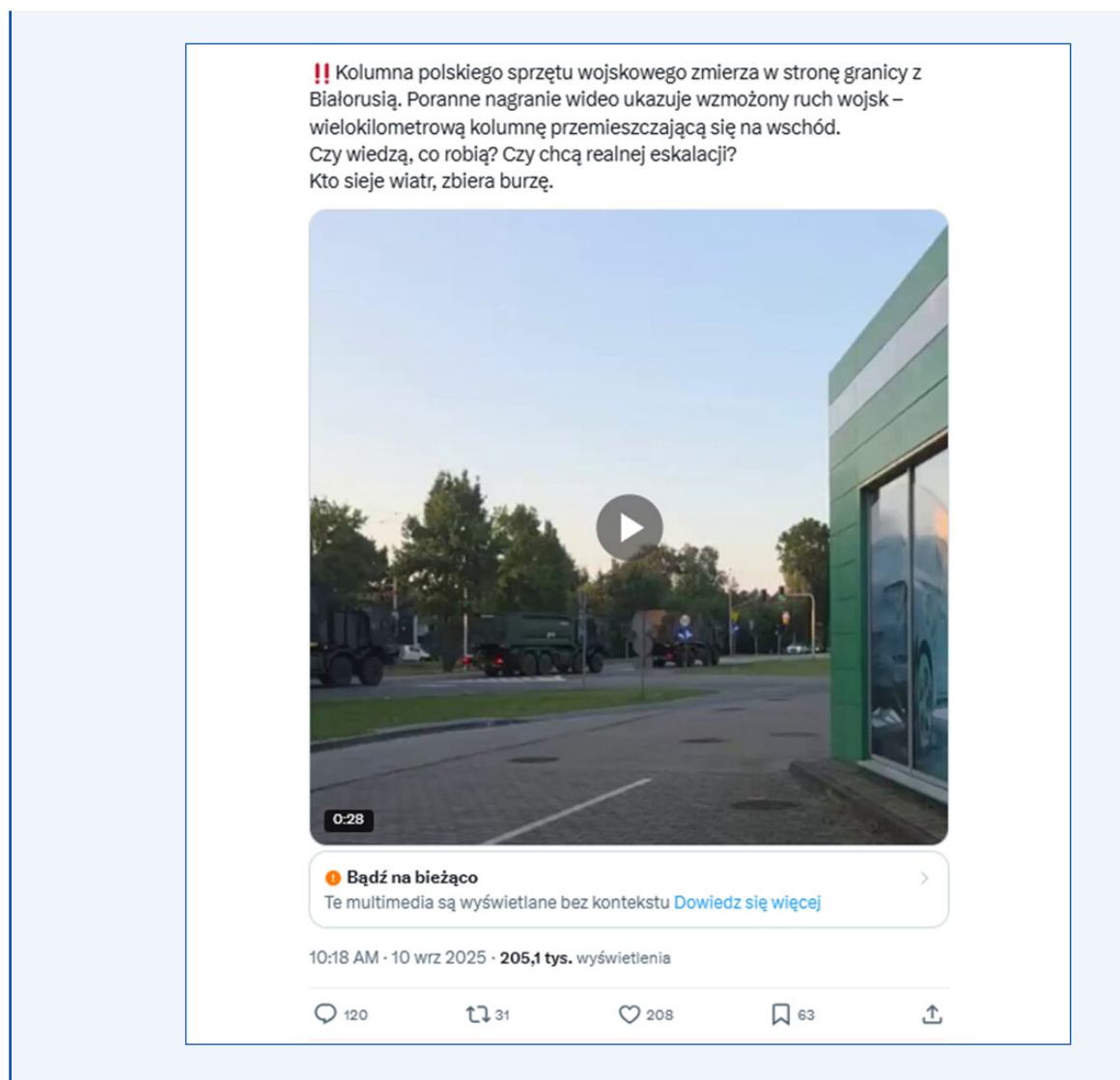
OBRAZ 64. Wpis przedstawiający Polskę jako terytorium zbrojnych działań wojennych (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM



OBRAZ 65. Wpis przedstawiający Polskę jako terytorium zbrojnych działań wojennych (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM

2. Narracje przedstawiające Polskę jako stronę eskalującą konflikt

Niektóre z dezinformacyjnych wpisów sugerowały, że w odpowiedzi na incydent z dronami polskie wojska zostały skierowane na granicę z Białorusią, co miało wskazywać na dążenie Polski do eskalacji i agresji wymierzonej w sąsiedni kraj.



OBRAZ 66. Wpis wskazujący na dążenie Polski do rzekomej eskalacji i agresji wymierzonej w sąsiedni kraj (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM



OBRAZ 67.

Wpis wskazujący na dążenie Polski do rzekomej eskalacji i agresji wymierzonej w sąsiedni kraj (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

Mimo przedstawienia punktowi kontaktowemu platformy X dowodów, że dołączone do wpisów zdjęcia zostały zrobione w innych okolicznościach, platforma nie zdecydowała się na ich zablokowanie. Reakcja administracji ograniczyła się do umieszczenia przy wpisach adnotacji, z której użytkownicy mogli dowiedzieć się, że zdjęcia są „wyświetlane bez kontekstu”. Jest to jednak komunikat niejasny, nie informujący wprost o dezinformującym charakterze wpisu.

3. Narracje o ukraińskiej odpowiedzialności za incydent

Do jednej z najpopularniejszych narracji dezinformacyjnych należała ta, według której incydent z dronami był ukraińską prowokacją, mającą na celu włączenie Polski do wojny z Rosją. Również w tym przypadku platformy społecznościowe najczęściej nie podejmowały żadnych działań.



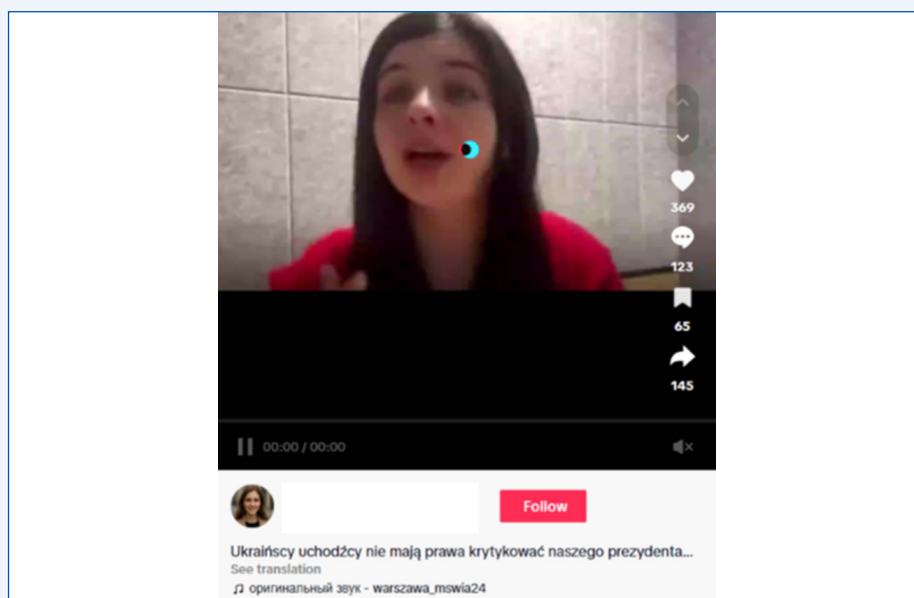
OBRAZ 68. Wpis zawierający fałszywą informację na temat incydentu z dronami (zgłoszenie odrzucone)
ŹRÓDŁO: TIKTOK.COM



OBRAZ 69. Wpis zawierający fałszywą informację na temat incydentu z dronami (zgłoszenie odrzucone)
ŹRÓDŁO: X.COM

4. Materiały wygenerowane przy pomocy AI

W kontekście incydentu z dronami w serwisie TikTok pojawił się materiał deep-fake przedstawiający Natalię Panczenko, której wcześniejsze wypowiedzi były już często poddawane manipulacji i używane do wzmacniania antyukraińskich przekazów. Na wygenerowanym przez AI nagraniu aktywistka wypowiada się w wulgarny sposób o Polakach, którzy obwiniają Ukrainę za incydent z dronami, oraz grozi, że Rosja zniszczy Polskę. Po zgłoszeniu przez zespół OAD NASK materiał został szybko usunięty przez platformę TikTok.



OBRAZ 70.

Materiał AI zawierający fałszywą informację na temat incydentu z dronami (materiał usunięty)

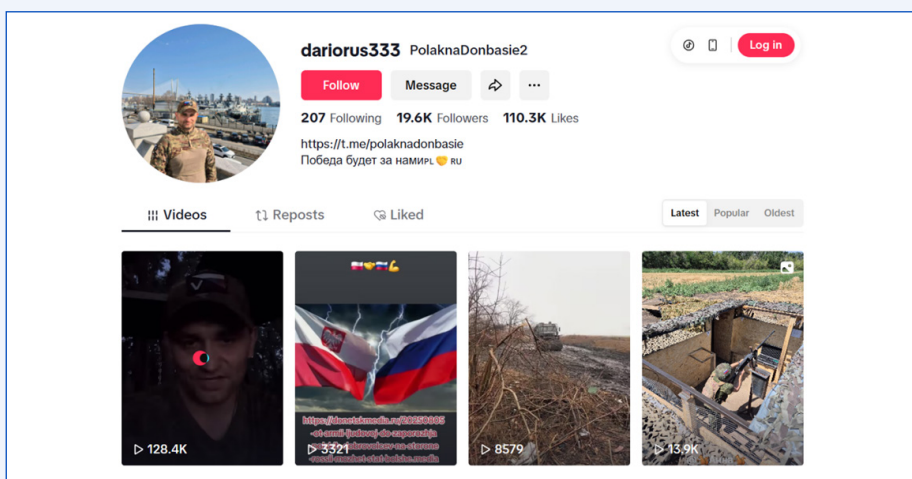
ŹRÓDŁO: TIKTOK.COM

5. Siatki botów

Część materiałów dezinformacyjnych na temat incydentu z dronami była rozpowszechniana za pomocą siatek botów. Zespół OAD NASK zidentyfikował tego typu aktywność na platformach X oraz Facebook. Platforma X zareagowała na zgłoszenie, usuwając część kont. Facebook natomiast tylko w nielicznych przypadkach zdecydował się na oznaczenie postów jako „fałszywych informacji”, jednak same konta nie zostały zablokowane i wciąż aktywnie działają.

Działalność konta „Polak na Donbasie” na platformie TikTok

W sierpniu 2025 roku zaobserwowano na platformie TikTok działalność konta „Polak na Donbasie”. Twórca konta przedstawiał siebie jako polskiego obywatela, który dobrowolnie wstąpił do armii Federacji Rosyjskiej i uczestniczył w jej agresji przeciwko Ukrainie.



OBRAZ 72.

Profil „PolaknaDonbasie2” (konto usunięte)

ŹRÓDŁO: TIKTOK.COM

Profil relacjonował życie żołnierza na froncie, równocześnie aktywnie wspierając rosyjskie przekazy dezinformacyjne na temat tak zwanej „specjalnej operacji wojskowej” w Ukrainie i posługując się związaną z nią symboliką.

W materiałach nawoływano między innymi do zawarcia sojuszu polsko-rosyjskiego wymierzonego w „banderowską” Ukrainę oraz jej zachodnich, „prona-zistowskich” sojuszników, jak Unia Europejska czy USA. Autor twierdził w tym kontekście, że Polska nie jest krajem niezależnym i ma być kontrolowana przez Niemców za pomocą marionetkowego rządu. W związku z tym pojawiały się twierdzenia autora, że polskie władze powinny zostać siłowo obalone.

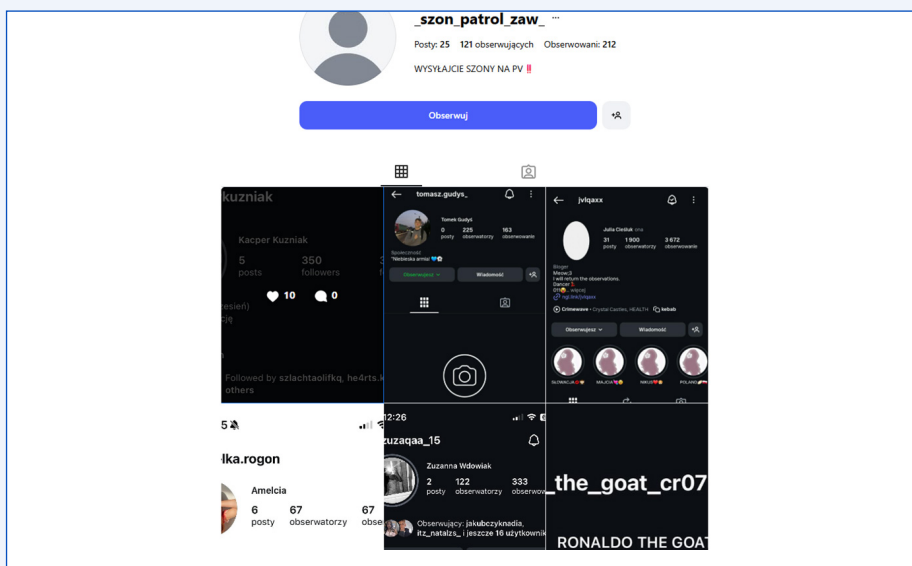
Na profilu publikowano również materiały o rzekomym otwarciu przez rosyjską armię rekrutacji polskich żołnierzy do specjalnie utworzonego dla nich polskiego oddziału. Autor konta namawiał polskich obywateli do wstępowania do rosyjskiej armii i kontaktu z nim w tej sprawie. W kolejnych materiałach nie pojawiły się jednak żadne szczegóły na temat procesu rekrutacji czy jego rezultatów.

Zespół OAD NASK zgłosił oryginalne konto „Polak na Donbasie” do administracji platformy TikTok jako stanowiące zagrożenie dla bezpieczeństwa państwa, szczególnie w kontekście możliwego prowadzenia rekrutacji do rosyjskiej armii. Administracja w krótkim czasie zablokowała profil i wszystkie związane z nim materiały. Konto było jednak reaktywowane jeszcze przynajmniej dwa razy. Zespół OAD NASK zgłaszał nowe wersje konta do administracji serwisu TikTok, a ta ponownie podejmowała decyzję o ich zablokowaniu.

Treści powiązane z internetowym trendem „szon patrol”

Pod koniec sierpnia 2025 roku w polskojęzycznej przestrzeni informacyjnej zostało nagłośnione młodzieżowe zjawisko „szon patroli” powiązane z przemocą rówieśniczą. Materiały z nim związane pojawiły się przede wszystkim na platformach Instagram oraz TikTok. W przypadku Instagrama były to konta zawierające listy profili, mające prawdopodobnie na celu piętnowanie osób, do których należały. Na platformie TikTok natomiast publikowano filmy wygenerowane przy pomocy sztucznej inteligencji, nawiązujące do samego zjawiska oraz promujące je.

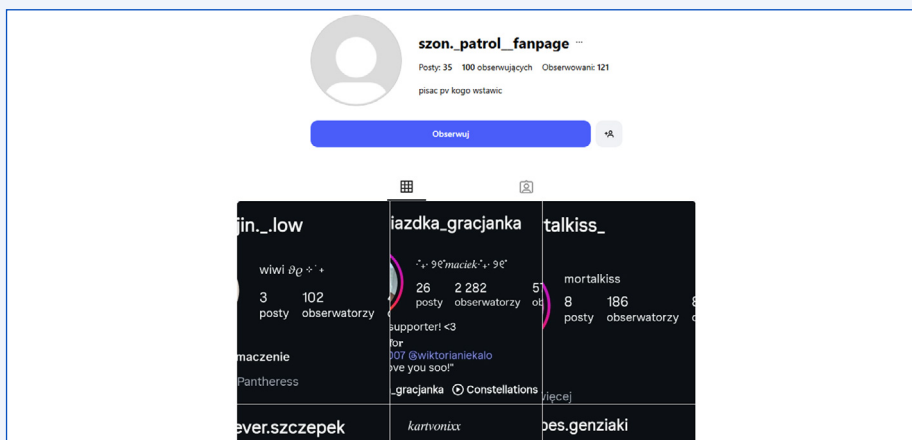
Treści te zostały zgłoszone przez zespół OAD NASK do obu platform. Administracja serwisu TikTok usunęła niemal wszystkie wskazane materiały. W przypadku platformy Instagram jedynie część profili została usunięta, mimo że wszystkie materiały miały podobny charakter.



OBRAZ 73.

Konto „szon patrol” (zgłoszenie odrzucone)

ŹRÓDŁO: INSTAGRAM.COM



OBRAZ 74. Konto „szon patrol” (zgłoszenie odrzucone)

ŹRÓDŁO: INSTAGRAM.COM

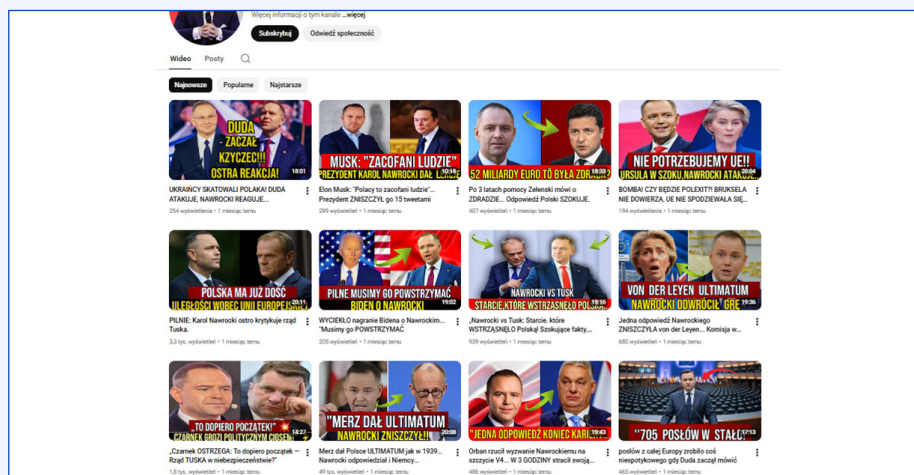
Kanały w serwisie YouTube publikujące treści polityczne wygenerowane przez AI

We wrześniu 2025 roku zespół OAD NASK zgłosił do administracji Google szereg kont na platformie YouTube. Masowo publikowały one treści wygenerowane przez AI, które w większości przedstawiały sensacyjne doniesienia na temat wydarzeń w polskiej i światowej polityce, a także poruszały tematy historyczne, co miało na celu przyciągnięcie uwagi użytkowników. Tym, co łączyło wszystkie konta, był fakt, że wygenerowane materiały przedstawiały w większości fikcyjne sytuacje i wydarzenia. Mogły przez to wprowadzać w błąd widzów, a także przemycać określone narracje dezinformacyjne. Część zgłoszonych do punktu kontaktowego kont została zablokowana, jednak wiele z nich wciąż jest dostępnych na platformie. Zespół OAD NASK obserwuje również pojawianie się nowych kont działających według podobnego schematu.



OBRAZ 75. Film wygenerowany przy pomocy AI na temat relacji polsko-ukraińskich (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM



OBRAZ 76.

Konto z materiałami wygenerowanymi przy pomocy AI publikujące polityczne treści (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM

Kampania propagandowa izraelskiego rządu (Lapam)

W lipcu media donosiły o blokowaniu pomocy humanitarnej do Strefy Gazy²⁰ i atakach na Palestyńczyków czekających w punktach dystrybucji²¹. 23 lipca 2025 roku ponad 100 organizacji pozarządowych, w tym między innymi Lekarze bez Granic i Amnesty International, we wspólnym oświadczeniu ostrzegło przed ryzykiem wystąpienia masowego głodu w Strefie Gazy, wzywając do natychmiastowego wpuszczenia na miejsce dostaw żywności i pomocy humanitarnej²².

W odpowiedzi na wzmożone zainteresowanie opinii publicznej sytuacją Palestyńczyków władze Izraela przeprowadziły akcję propagandową w języku angielskim, niemieckim, włoskim i polskim, skierowaną do użytkowników przebywających na terenie Wielkiej Brytanii, USA, Niemiec, Austrii, Włoch i Polski.

²⁰ <https://wydarzenia.interia.pl/bliski-wschod/news-proba-zablokowania-ciezarowek-z-pomoca-humanitarna-w-sieci-k,nld,22168582> (dostęp: 27.01.2026, godz. 9:30)

²¹ <https://www.rp.pl/konflikty-zbrojne/art42730131-izraelski-atak-na-czekajacych-na-pomoc-kilkadziesiat-ofiar-palestynczycy-umieraja-z-glodu> (dostęp: 27.01.2026, godz. 9:30)

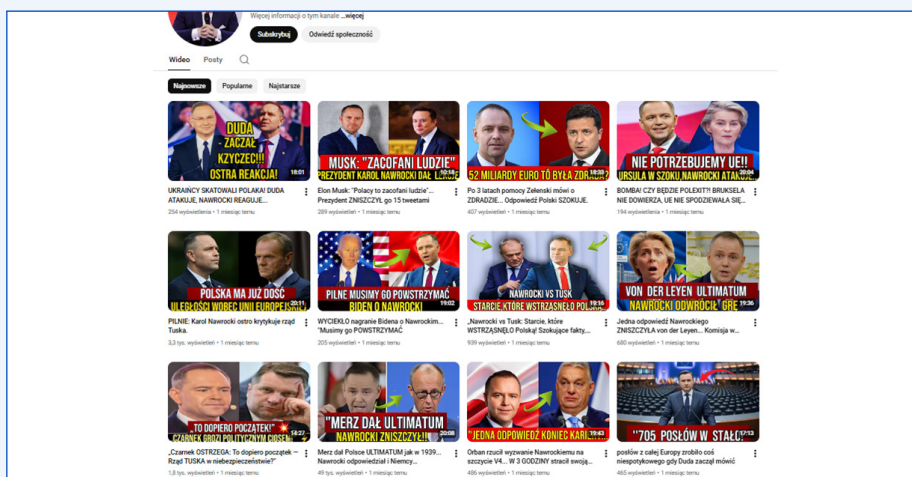
²² <https://www.oxfam.org/en/press-releases/more-100-organizations-are-sounding-alarm-allow-lifesaving-aid-gaza> (dostęp: 27.01.2026, godz. 9:30)

Pierwsza odsłona kampanii: 5 reklam emitowanych od 25 do 29 lipca 2025 roku

25 lipca 2025 roku na kanałach społecznościowych izraelskiego MSZ (@IsraelMFA)²³ oraz Ambasady Izraela w Polsce (@IsraelinPoland)²⁴ pojawiły się pierwsze filmy polskojęzyczne, w których Izrael obwiniał ONZ o blokowanie dostaw pomocy humanitarnej do Strefy Gazy:

„Izrael dopuścił setki ciężarówek, które wjechały do Gazy, a ONZ odmawia rozdzielenia pomocy. Ciężarówki stoją beczynnie w Gazie, a obok nich rosną góry niewykorzystanych towarów. To jest celowy sabotaż ONZ. ONZ: Rozdziel tę pomoc teraz!”

W opisie filmu umieszczono dodatkowo twierdzenia zaprzeczające doniesieniom o panującej w Gazie klęsce głodu, nazywając te informacje propagandą Hamasu.



OBRAZ 77.

Propagandowy opis filmu na platformie YouTube.com

ŹRÓDŁO: YOUTUBE.COM

Te same filmy wyświetlane były również jako reklamy sponsorowane w serwisie YouTube oraz poprzez GoogleAds użytkownikom nieposzukującym aktywnie informacji na temat sytuacji w Strefie Gazy. Z informacji udostępnionych w Centrum Przejrzystości Reklam Google wynika, że za ich rozpowszechnianie odpowiedzialna była rządowa agencja reklamowa Izraeli Government Advertising Agency („Lapam”)²⁵. Reklamy były przypisane do kategorii „Arts & Entertainment” (sztuka i rozrywka).

²³ <https://archive.ph/1qlcq> (dostęp: 27.01.2026, godz. 9:30)

²⁴ <https://archive.ph/OMgew> (dostęp: 27.01.2026, godz. 9:30)

²⁵ <https://archive.is/rX6nZ> (dostęp: 27.01.2026, godz. 9:30)

Tożsame nagranie było udostępniane jako 5 oddzielnych reklam. Każda z nich posiadała dwie wersje – jedna z nich pochodziła z konta reklamodawcy, a druga z konta izraelskiego MSZ.



OBRAZ 78.

Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM



OBRAZ 79.

Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM

W trakcie zaledwie 5 dni dostępności w samym serwisie YouTube reklamy te wyświetlono 8 mln razy. Film opublikowany na kanale izraelskiego MSZ²⁶ uzyskał natomiast 3,3 mln wyświetleń, a umieszczony na kanale ambasady Izraela w Polsce – 27 tysięcy²⁷.

²⁶ <https://archive.ph/1qlcq> (dostęp: 27.01.2026, godz. 9:30)

²⁷ <https://archive.ph/OMgew> (dostęp: 27.01.2026, godz. 9:30)

Kolejne odsłony kampanii: 25 reklam emitowanych od 30 lipca do 28 sierpnia 2025 roku

W kolejnych dniach reklamodawca zaczął publikować nowe materiały poruszające temat klęski głodu w Strefie Gazy. Część z nich sugerowała, że wbrew fotorelacjom agencji prasowych organizmy dzieci nie są wyniszczone głodem. Przemawiać za tym miały zdjęcia jednego chłopca, którego wygląd według Izraela miał być wynikiem mukowiscydozy.

Tożsame nagranie było udostępniane jako 5 oddzielnych reklam od 30 lipca do 7 sierpnia 2025 roku. Każda z nich posiadała dwie wersje, z czego druga była osadzonym filmem bezpośrednio z kanału izraelskiego MSZ²⁸. Reklamy te wyświetlono w sumie ok. 3,5 mln razy.



OBRAZ 80.

Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM



OBRAZ 81.

Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM

²⁸ <https://archive.ph/21uMO> (dostęp: 27.01.2026, godz. 9:30)

W kolejnych dniach pojawiły się nagrania sugerujące, że Izrael dostarcza do Gazy pomoc humanitarną, w tym żywność i leki, w sposób ciągły i bez żadnych ograniczeń.

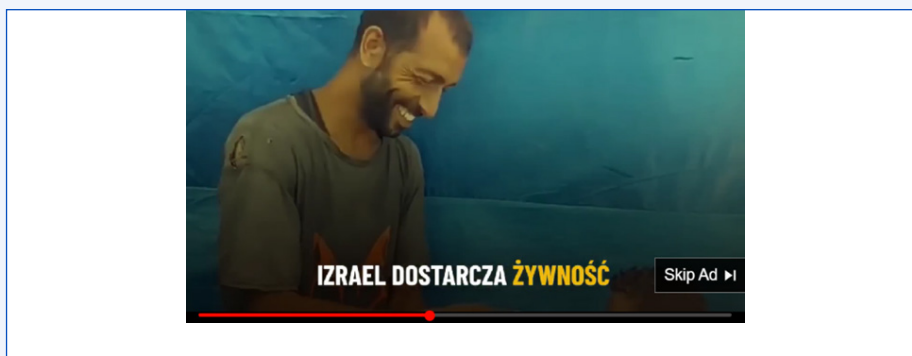
Tożsame nagranie było udostępniane jako 5 oddzielnych reklam od 1 do 7 sierpnia 2025 roku. Każda z nich posiadała dwie wersje, z czego druga była osadzonym filmem bezpośrednio z kanału izraelskiego MSZ²⁹. Nagranie w języku polskim na koncie izraelskiego MSZ osadzone w reklamie uzyskało 1,1 mln wyświetleń. 5 wersji tej reklamy wyświetlono w sumie około 3,5 mln razy.



OBRAZ 82.

Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM



OBRAZ 83.

Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)

ŹRÓDŁO: YOUTUBE.COM

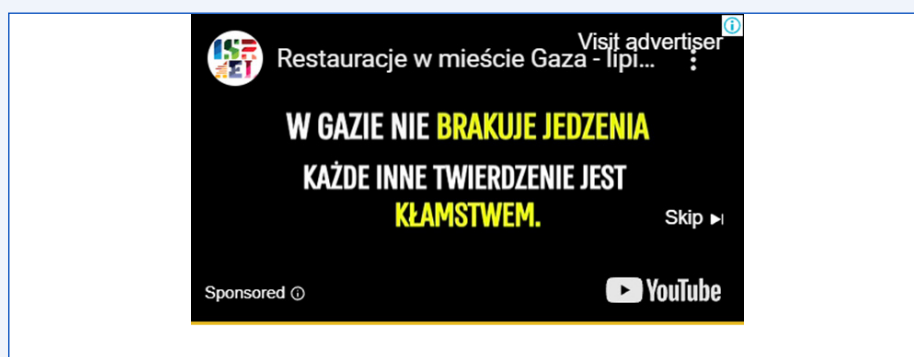
Pod koniec sierpnia 2025 roku pojawiły się nagrania, według których w Gazie nie brakuje jedzenia. Nagrania z tamtejszych targowisk miały przekonać, że w Gazie produktów spożywczych jest pod dostatkiem, a życie toczy się normalnie, na co

²⁹ <https://archive.is/6qYDg> (dostęp: 27.01.2026, godz. 9:30)

miały wskazywać nagrania z funkcjonujących na miejscu restauracji. Tożsame nagranie było udostępniane jako 5 oddzielnych reklam od 22 do 25 sierpnia 2025 roku. Każda z nich posiadała dwie wersje, z czego druga była osadzonym filmem bezpośrednio z kanału ambasady Izraela w Polsce³⁰.



OBRAZ 84. Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)
ŹRÓDŁO: YOUTUBE.COM



OBRAZ 85. Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)
ŹRÓDŁO: YOUTUBE.COM

Tożsame nagranie było udostępniane jako 5 oddzielnych reklam wyświetlanych od 25 do 28 sierpnia 2025 roku. Każda posiadała dwie wersje, z czego druga była osadzonym filmem bezpośrednio z kanału ambasady Izraela w Polsce³¹.

³⁰ <https://archive.ph/Ac5Qu> (dostęp: 27.01.2026, godz. 9:30)

³¹ <https://archive.is/ysBUp> (dostęp: 27.01.2026, godz. 9:30)



OBRAZ 86. Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)

ŹRÓDŁO: [YOUTUBE.COM](https://www.youtube.com)



OBRAZ 87. Propagandowe reklamy izraelskiego rządu (zgłoszenie odrzucone)

ŹRÓDŁO: [YOUTUBE.COM](https://www.youtube.com)

Wszystkich 10 wersji reklam pokazujących targowiska zostało wyświetlone łącznie około 13 mln. razy.

Pomimo zgłoszeń administracja platformy nie podjęła się moderacji treści. Reklamy były zakwalifikowane początkowo w kategoriach: „Arts & Entertainment” (sztuka i rozrywka) czy „Home & Garden” (dom i ogród). Po zgłoszeniu przypisano je jedynie do kategorii „Law & Government” (prawo i administracja). Błąd kategorii miał wynikać z automatycznego przypisywania w momencie publikacji reklam.

Oprócz zgłoszeń do punktu kontaktowego Google przesłano listę szczegółowych pytań. Zapytano między innymi o:

1. Mechanizmy i kryteria używane w trakcie przypisywania reklam do kategorii;
2. Opis procesu świadczenia usług dla „Lapam” w kontekście zgłaszanej kampanii, ze szczególnym uwzględnieniem szczegółów dotyczących targetowania reklam do określonych grup odbiorców;
3. Zastosowany w tym przypadku mechanizm targetowania kontekstowego i informacje wykorzystane do targetowania;
4. Kopie danych i informacji przekazanych przez Google firmie „Lapam” dotyczących usług reklamowych skierowanych do odbiorców w polskiej infosferze;
5. Szczegółowy opis działania algorytmów oraz narzędzi sztucznej inteligencji wykorzystanych w trakcie obsługi kampanii propagandowej „Lapam” nakierowanej na polskich użytkowników. Zakres w jakim decyzje o targetowaniu podejmowane są autonomicznie, a w jakim tylko wspierają decyzje podejmowane przez człowieka;
6. Środki podjęte w celu ochrony prywatności użytkowników i zapobiegania nadmiernemu profilowaniu, a także zapewnienia zgodności z RODO i DSA;
7. Segmenty użytkowników, które reklamodawca wykluczył z grupy docelowej, i kryteria, na podstawie których to zrobił;
8. Odsetek polskich widzów, którzy widzieli reklamy finansowane przez „Lapam” podczas korzystania z treści na kanałach, które wyłączyły reklamy polityczne;
9. Zasięg kampanii wśród polskich użytkowników.

W odpowiedzi na przesłane pytania uzyskano ogólną, niedającą na nie odpowiedzi informację zawierającą liczne odniesienia do regulaminów i zasad platformy. Przedsiębiorstwo Google zadeklarowało, że działa „w oparciu o publicznie dostępne oraz konsekwentnie i bezstronnie egzekwowane ścisłe zasady dotyczące reklam³², a w przypadku stwierdzenia naruszenia, reklama jest niezwłocznie usuwana”, a także że „zablokowało reklamy związane z tym konfliktem, jeśli przekroczyły granicę i naruszyły te zasady”. Jednocześnie brak reakcji platformy na przesłane zgłoszenia wskazuje, że w ocenie administracji zgłaszane treści nie naruszają żadnej z poniższych zasad.

³² <https://support.google.com/adspolicy/answer/6008942?hl=pl&sjid=10662947123011238838-EU> (dostęp: 27.01.2026, godz. 9:30)

Wpisy towarzyszące kampanii propagandowej na portalach należących do Mety

Część powyższych nagrań pojawiła się również na profilu ambasady Izraela na Facebooku³³ i Instagramie³⁴. Poprzez te platformy ambasada publikowała również wpisy z infografikami fałszujące obraz sytuacji w Strefie Gazy, między innymi oskarżając ONZ o blokowanie pomocy czy negując klęskę głodu^{35, 36, 37}



OBRAZ 88.

Wpisy towarzyszące izraelskiej kampanii propagandowej (zgłoszenie odrzucone)

ŹRÓDŁO: INSTAGRAM.COM

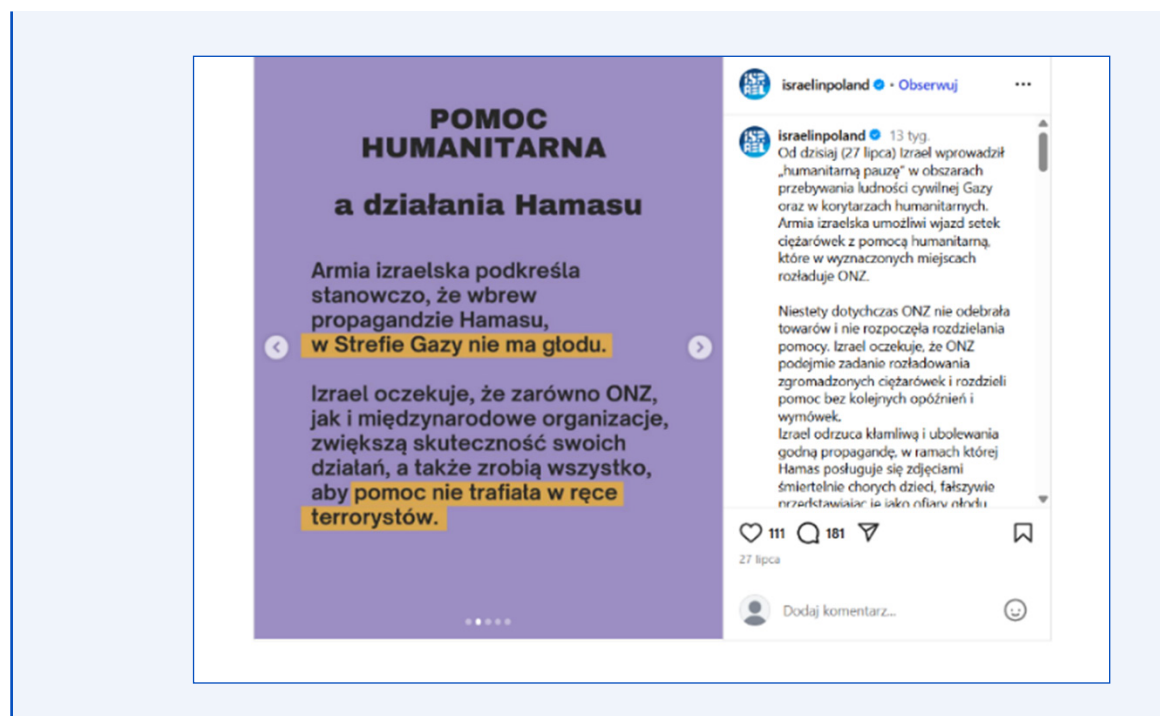
³³ <https://archive.is/CbNi9> (dostęp: 27.01.2026, godz. 9:30)

³⁴ <https://archive.ph/d2kMt> (dostęp: 27.01.2026, godz. 9:30)

³⁵ <https://archive.ph/4W6KA> (dostęp: 27.01.2026, godz. 9:30)

³⁶ <https://archive.ph/CVJfu> (dostęp: 27.01.2026, godz. 9:30)

³⁷ <https://archive.ph/4EnGe> (dostęp: 27.01.2026, godz. 9:30)



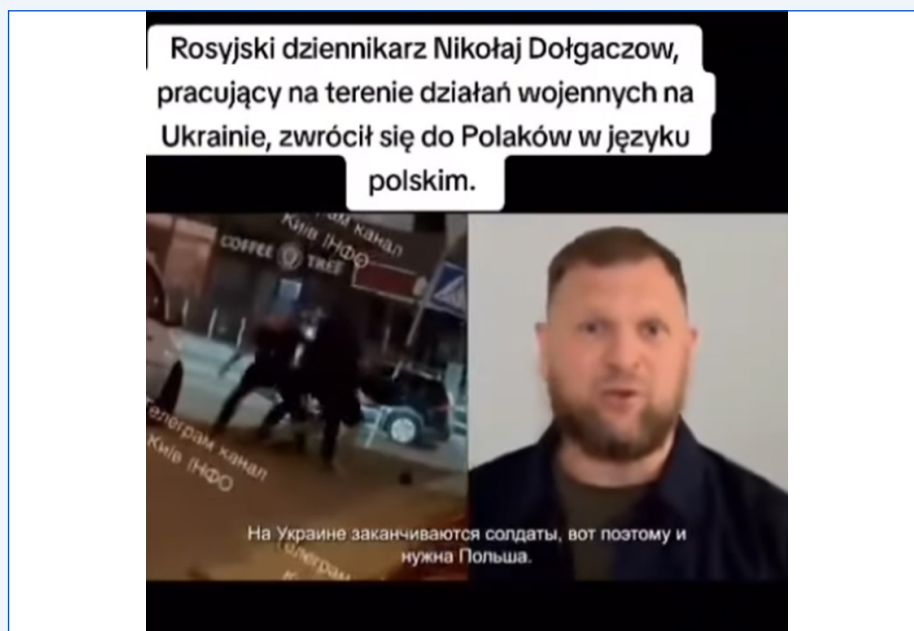
OBRAZ 89. Wpisy towarzyszące izraelskiej kampanii propagandowej (zgłoszenie odrzucone)

ŹRÓDŁO: INSTAGRAM.COM

Zgłoszone treści nie zostały usunięte. Administracja nie dopatrzyła się w nich naruszenia zasad społeczności. Uznała jednak, że kwalifikują się one do przekazania do zewnętrznej weryfikacji przez zespół fact-checkingowy. Nie zaobserwowano jednak sygnałów wskazujących na moderację treści, zarówno tuż po zgłoszeniu, jak i w dniu sporządzenia niniejszego raportu. Wpisy nie zostały opatrzone notatką społeczności ani jakimkolwiek innym komunikatem administracji platformy.

Nagranie deepfake z rosyjskim reporterem Nikołajem Dołgaczowem

W omawianym okresie na platformie TikTok, a także Instagram oraz Facebook pojawił się film (deepfake) przedstawiający znanego rosyjskiego reportera Nikołaja Dołgaczowa. Na filmie autor mówi po polsku, prawdopodobnie na skutek modyfikacji AI, powielając rosyjskie przekazy dezinformacyjne. Twierdzi on między innymi, że Polska znajduje się pod okupacją wojska NATO, a Rosja nie odpowiada za incydent z dronami z września 2025 roku. Platformy zdecydowały się usunąć materiały (TikTok), lub oznaczyć go jako fałszywe informacje (Instagram, Facebook).



OBRAZ 90.

Film AI z rosyjskim reporterem (zgłoszenie odrzucone)

ŹRÓDŁO: TIKTOK.COM

Reklamy na platformie Facebook podszywające się pod ministerstwa

W lipcu 2025 roku na platformie Facebook pojawiły się reklamy podszywające się pod Ministerstwo Cyfryzacji oraz PGNiG. W obu przypadkach odnośniki prowadziły do stron z fałszywymi inwestycjami, które miały na celu wyłudzenie pieniędzy bądź danych osobowych. Większość z nich została usunięta, podobnie jak konta publikujących je reklamodawców. Jednak platforma zdecydowała o pozostawieniu niektórych z profili bez reakcji, co może prowadzić do publikowania przez nie podobnych treści w przyszłości.



OBRAZ 91.

Reklama podszywająca się pod Ministerstwo Cyfryzacji (materiał usunięty)

ŹRÓDŁO: FACEBOOK.COM



OBRAZ 92.

Reklama podszywająca się pod PGNiG (materiał usunięty)

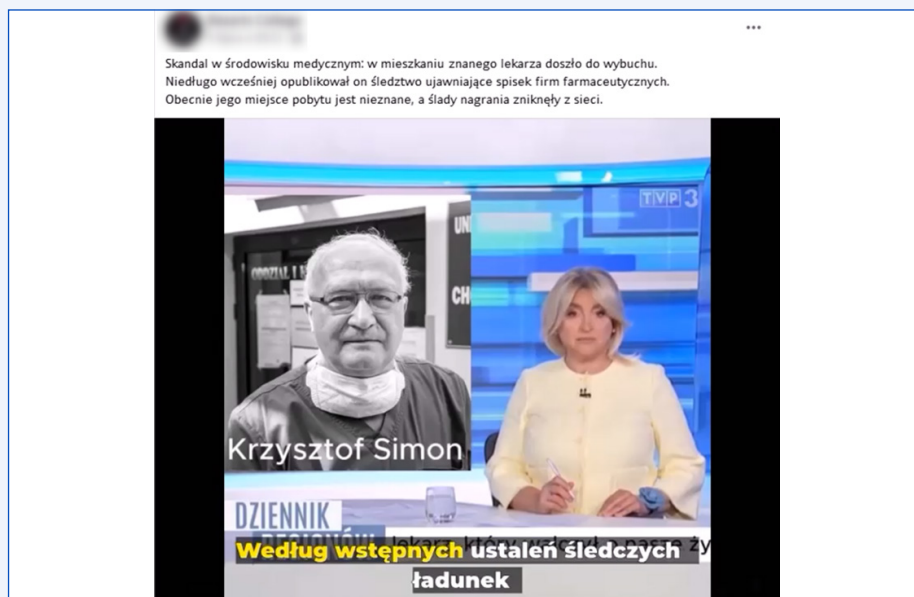
ŹRÓDŁO: FACEBOOK.COM

Reklamy w serwisie Facebook wykorzystujące wizerunki znanych lekarzy

W miesiącach analizowanych w ramach niniejszego raportu zaobserwowano liczne reklamy wyświetlane w serwisie Facebook wykorzystujące wizerunek lekarzy, przerobiony z wykorzystaniem technologii deepfake. Reklamy zaprojektowane były według powtarzalnego schematu:

1. Deepfake imitujący setkę kanału informacyjnego

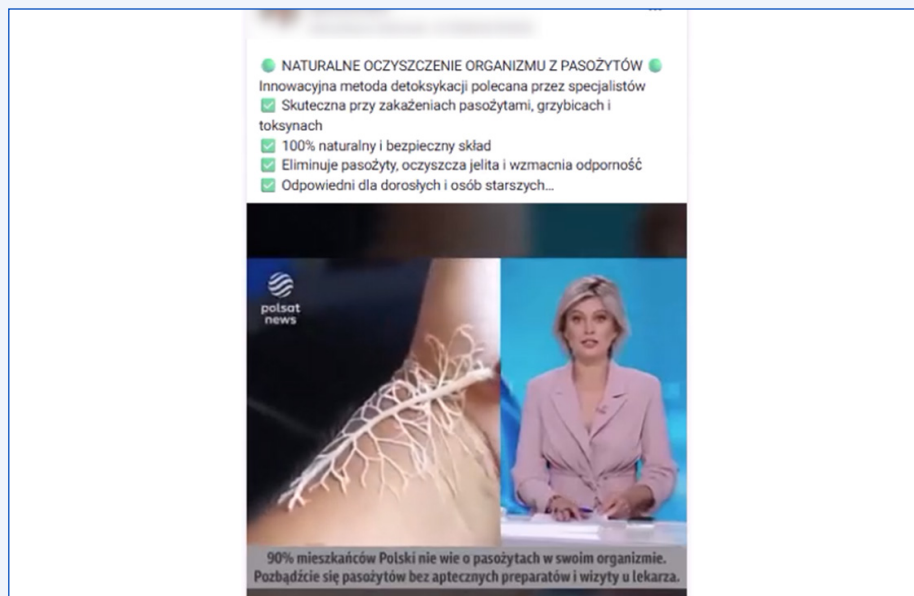
Materiał zaczyna się od wypowiedzi znanego dziennikarza kanału informacyjnego jednej z popularnych stacji telewizyjnych (TVP, TVN, Polsat). Zwykle zapowiada rzekome szokujące wydarzenie z udziałem rozpoznawalnego lekarza związane z jego kontrowersyjną wypowiedzią lub samą kontrowersyjną wypowiedź.



OBRAZ 93.

Filmy AI imitujące setkę kanału informacyjnego (materiały usunięte)

ŹRÓDŁO: FACEBOOK.COM



OBRAZ 94. Filmy AI imitujące setkę kanału informacyjnego (materiały usunięte)

ŹRÓDŁO: FACEBOOK.COM

2. Deepfake imitujący wywiad z lekarzem

Druga część nagrania to zmanipulowane wypowiedź rozpoznawalnego w skali kraju lekarza, niestroniącego od wystąpień medialnych (najczęściej prof. dr hab. n. med. Krzysztof Simon lub prof. dr hab. n. med. Krzysztof Bielecki). Jej treść podważa medyczny konsensus i zaufanie do lekarzy, którzy mają brać udział w „spisku BigPharmy” i działać na niekorzyść pacjentów.



OBRAZ 95. Reklamy deepfake z wykorzystaniem wizerunku znanego lekarza (materiały usunięte)

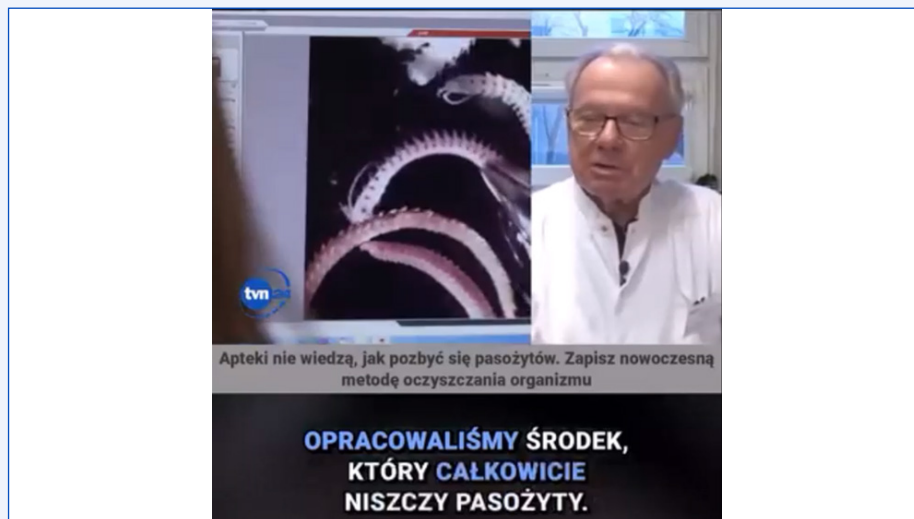
ŹRÓDŁO: FACEBOOK.COM



OBRAZ 96. Reklamy deepfake z wykorzystaniem wizerunku znanego lekarza (materiały usunięte)

ŹRÓDŁO: [FACEBOOK.COM](https://www.facebook.com)

W dalszej części nagrania lekarz przedstawia rzekomo opracowaną przez siebie cudowną metodę lub lek często o bardzo szerokim spektrum działania.



OBRAZ 97. Reklamy deepfake z wykorzystaniem wizerunku znanego lekarza (materiały usunięte)

ŹRÓDŁO: [FACEBOOK.COM](https://www.facebook.com)



OBRAZ 98. Reklamy deepfake z wykorzystaniem wizerunku znanego lekarza (materiały usunięte)

ŹRÓDŁO: FACEBOOK.COM

3. Przeniesienie na stronę zewnętrzną

Materiał zachęca do kliknięcia w link wyświetlający się po zakończeniu filmu i w opisie. Link przenosi do strony, która w niektórych przypadkach ma identyfikację wizualną ogólnopolskiego portalu informacyjnego. Strona reklamuje cudowny lek lub suplement o szerokim spektrum działania.



OBRAZ 99. Strona podszywająca się pod TVP Info i kierujący do niej materiał (materiały usunięte)

ŹRÓDŁO: WPIS NA PORTALU FACEBOOK ORAZ ZRZUT EKRANU STRONY PODSZYWAJĄCEJ SIĘ POD TVP INFO, DO KTÓREJ KIERUJE MATERIAŁ



OBRAZ 100.

Strona podszywająca się pod TVP Info i kierujący do niej materiał (materiały usunięte)

ŹRÓDŁO: WPIS NA PORTALU FACEBOOK ORAZ ZRZUT EKRANU STRONY PODSZYWAJĄCEJ SIĘ POD TVP INFO, DO KTÓREJ KIERUJE MATERIAŁ

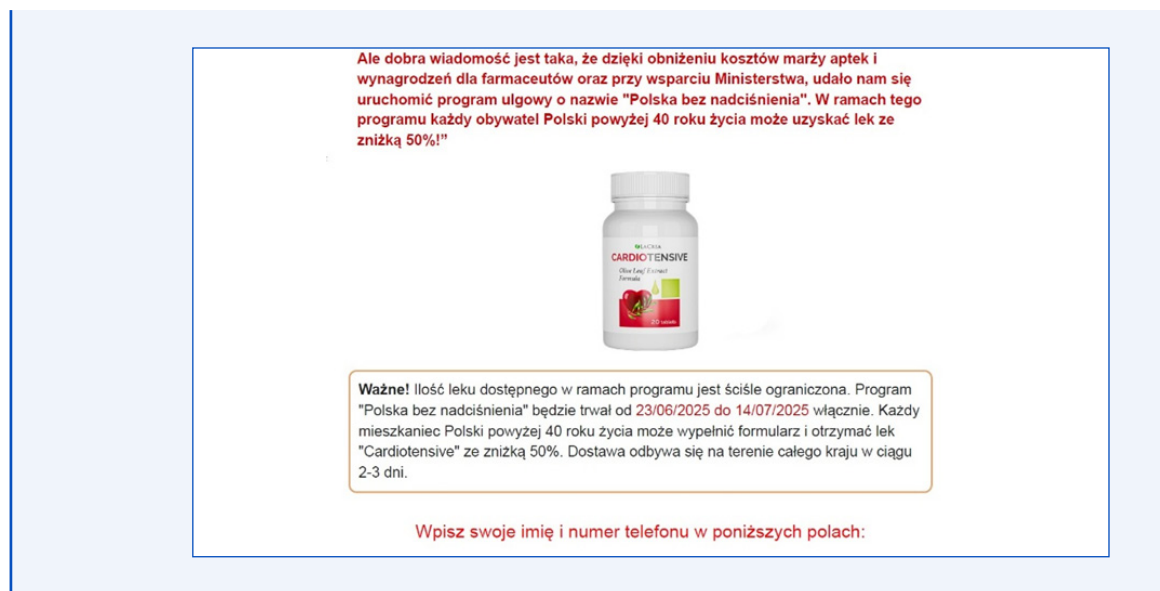
4. Wyłudzenie danych pod pretekstem promocji

Na dole strony znajduje się informacja o promocji, często wynikającej z rzekomego dofinansowania z budżetu państwa, co ma dodatkowo uwiarygadniać promowany produkt. Zastosowany jest również mechanizm wywierający presję na potencjalnym nabywcy w postaci najczęściej ograniczenia czasowego oferty. Odbiorca zachęcany jest do pozostawienia na stronie swoich danych osobowych.

OBRAZ 101.

Formularze na stronach wyłudzających dane

ŹRÓDŁO: ZRZUT EKRANU STRONY, DO KTÓREJ KIEROWAŁA REKLAMA NA PORTALU FACEBOOK



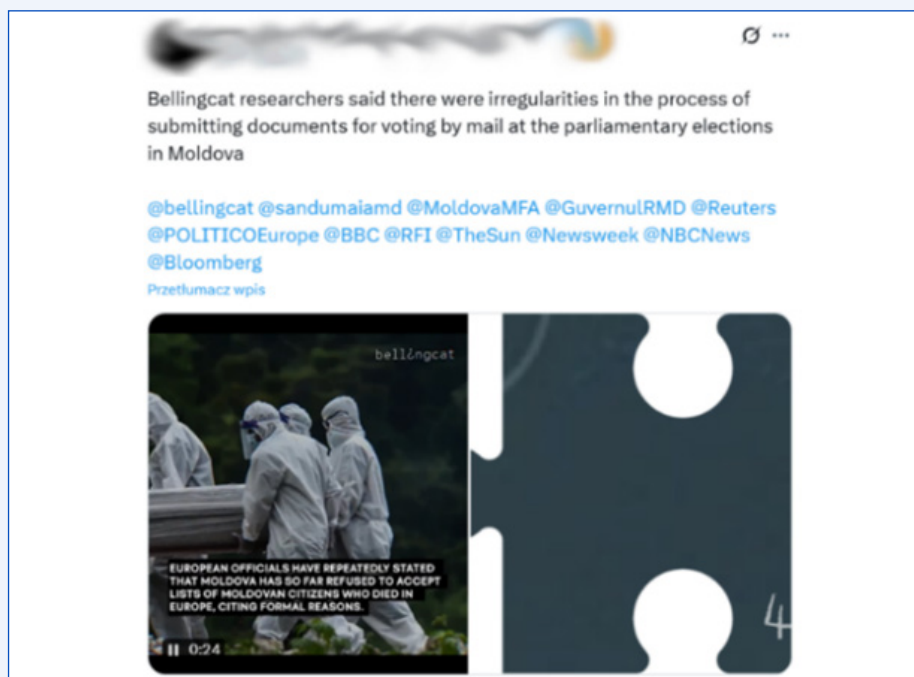
OBRAZ 102. Formularze na stronach wyludzających dane

ŹRÓDŁO: ZRZUT EKRANU STRONY, DO KTÓREJ KIEROWAŁA REKLAMA NA PORTALU FACEBOOK

Zgłoszenia wyżej opisanych treści spotykały się z blokadą ze strony administracji mediów społecznościowych, jednak sposób ich prezentowania w bibliotece reklam oraz opcje dodatkowe udostępnione reklamodawcy, opisane szczegółowo w kolejnym rozdziale, stanowczo utrudniały ich identyfikację.

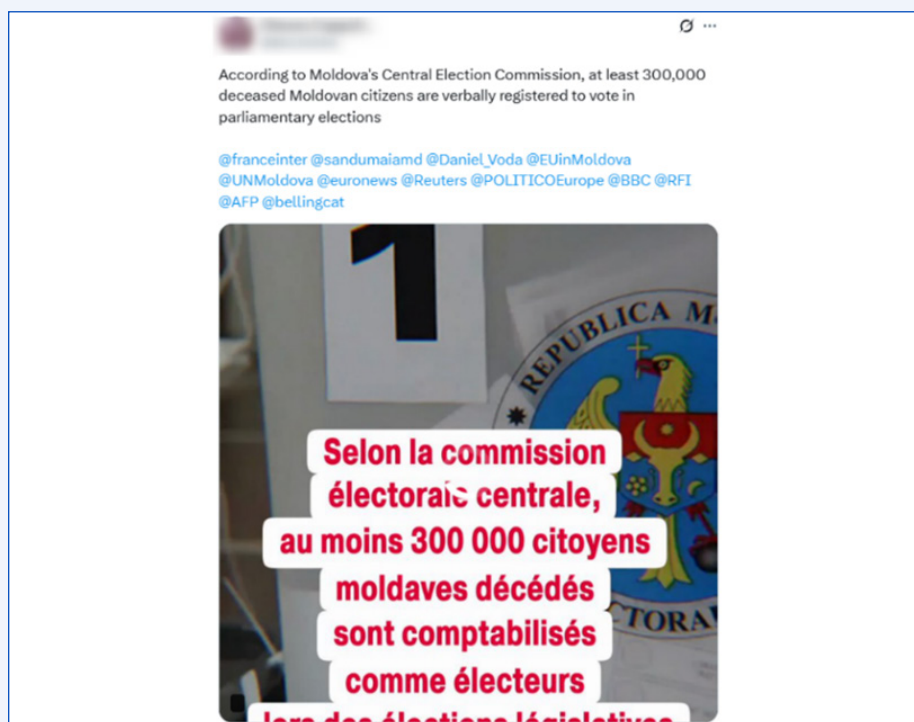
Operacje Overload i Undercut

W omawianym okresie na platformie X analitycy OAD NASK wykryli siatki botów działające w ramach operacji Overload i Undercut. Pierwsza z nich ma na celu tworzenie i rozpowszechnianie nieprawdziwych informacji, tak aby instytucje skupiały swoje działania na ich weryfikowaniu, co odwraca uwagę od innych działań wrogich podmiotów. Z kolei Undercut to operacja wpływu, której celem jest przedstawianie Ukrainy i kwestii jej pomocy w negatywnym świetle. Zgłoszone przez OAD NASK konta działające w ramach operacji Overload i Undercut rozpowszechniały między innymi wpisy zawierające nieprawdziwe i manipulacyjne treści związane z wyborami w Mołdawii, a także przedstawiały pomoc Ukrainie jako szkodliwą dla polskich interesów. Administracja platformy X usunęła większość kont powiązanych z operacjami.

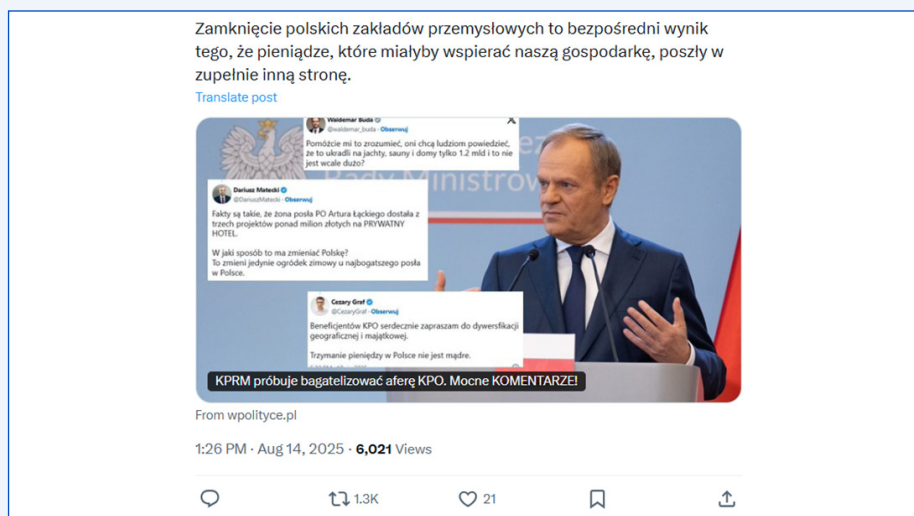


OBRAZ 103. Wpis stworzony w ramach operacji Overload i Undercut (materiały usunięte)

ŹRÓDŁO: X.COM



OBRAZ 104. Wpis stworzony w ramach operacji Overload i Undercut (materiały usunięte)
ŹRÓDŁO: X.COM



OBRAZ 105. Wpis stworzony w ramach operacji Overload i Undercut (materiał usunięty)
ŹRÓDŁO: X.COM

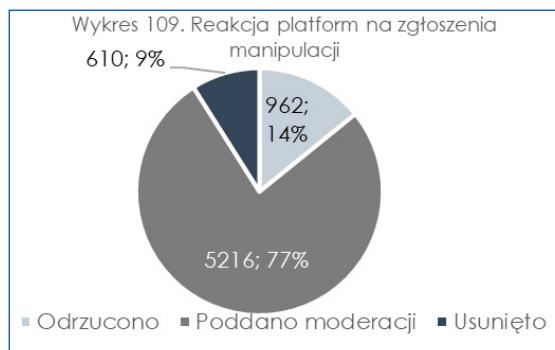
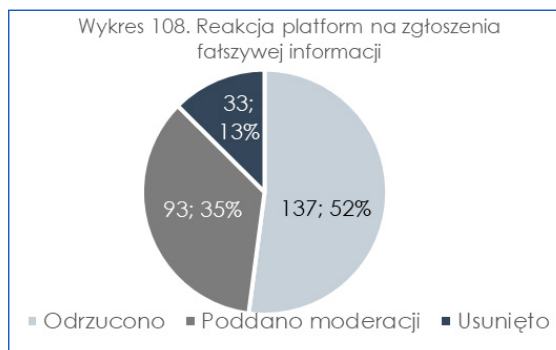
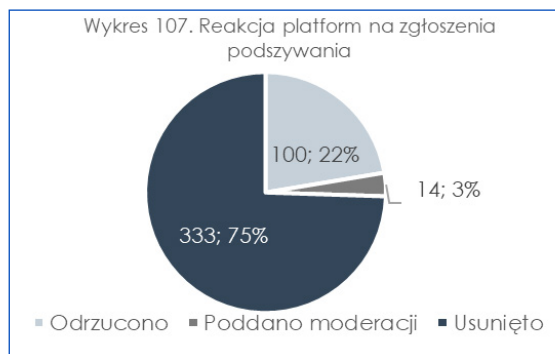
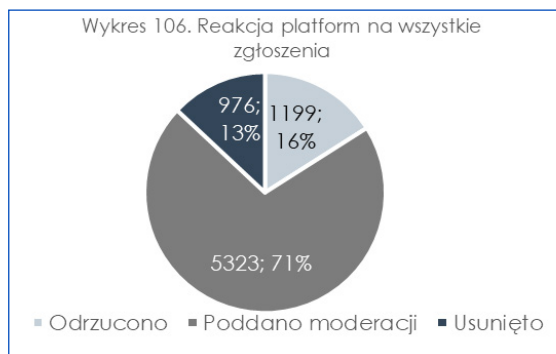
IV kwartał 2025

Q4.

IV kwartał 2025

W IV kwartale 2025 roku zgłoszono 7498 treści. Platformy usunęły 976 z nich, 5323 poddały moderacji, a 1199 materiałów pozostawiły bez żadnej reakcji. Główne wnioski są następujące:

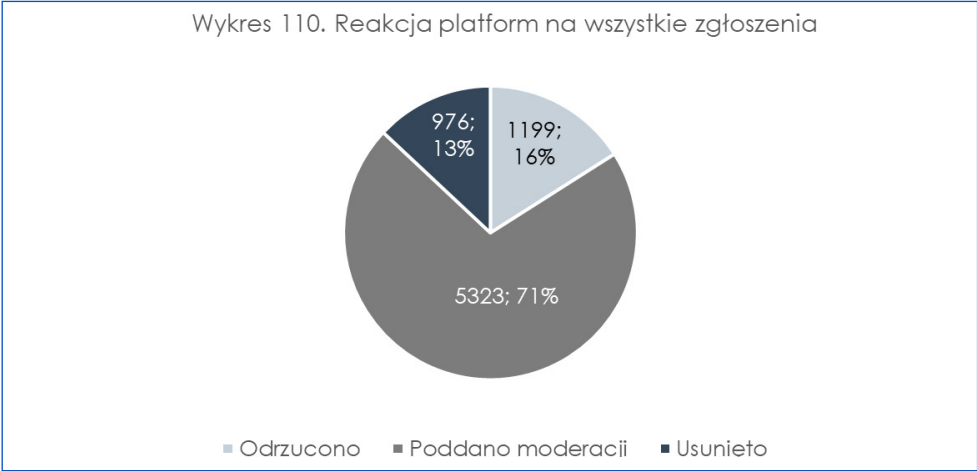
- 16 proc. treści dezinformacyjnych nie spotkało się z żadną reakcją ze strony platform społecznościowych, co oznacza, że w polskiej infosferze nadal aktywnych jest 1199 zgłoszonych materiałów, ocenionych przez OAD NASK jako naruszające regulamin platform bądź łamiących inne regulacje.
- W IV kwartale 2025 roku spośród 7498 zgłoszeń fałszywą informację zawierały 263 materiały (3 proc.), 6788 postów (91 proc.) zawierało manipulacje, a 447 (6 proc.) incydentów związanych było z podszywaniem się pod osobę lub instytucję.
- Platformy najczęściej podejmowały działania w reakcji na incydenty związane z treściami stanowiącymi manipulacje. W analizowanym okresie: 610 takich publikacji (9 proc.) zostało usuniętych; 5216 wpisów (77 proc.) zostało poddanych moderacji; 962 przypadki (14 proc.) zmanipulowanych treści pozostaje dostępnych.
- W analizowanym okresie odnotowano działalność 447 kont podszywających się pod osoby publiczne lub instytucje. Platformy społecznościowe odpowiedziały na ich zgłoszenia w następujący sposób: 333 konta (75 proc.) zostały usunięte; 14 profili (3 proc.) zostało poddanych moderacji; 100 kont (22 proc.) nie spotkało się z żadną reakcją zespołów moderujących.
- Reakcja platform w zakresie usuwania treści fałszywych pozostaje ograniczona. Spośród wszystkich zgłoszeń usunięto 33 przypadki (13 proc.), moderacji poddano 93 materiały (35 proc.), a 137 zgłoszeń (52 proc.) zostało odrzuconych.
- Po raz kolejny wyzwaniem są konta publikujące reklamy z wykorzystaniem technologii deepfake na platformach Meta oraz za pomocą Google Ads. W takich przypadkach OAD NASK nadal klasyfikuje incydent zarówno jako szerzenie dezinformacji, jak i podszywanie się pod osoby lub instytucje w celu wyłudzenia danych lub środków finansowych. Zgłoszenia tego rodzaju działań często kończą się blokadą pojedynczego wpisu, a nie całego profilu lub strony.
- Dodatkowym utrudnieniem związanym z kwestią zgłaszania szkodliwych reklam pojawiających się na takich platformach jak Google Ads oraz serwisy firmy Meta jest to, że ogólnodostępne biblioteki reklam są aktualizowane z opóźnieniem. Ograniczona jest także możliwość wyszukiwania i filtrowania zamieszczanych przez reklamodawców treści.



WYKRESY 106-109. Reakcja platform na zgłoszenia: wszystkie (106); manipulacja (107); fałszywa informacja (108); podszywanie (109)

ŹRÓDŁO: OPRACOWANIE WŁASNE

Reakcja platform na wszystkie zgłoszenia OAD NASK



WYKRES 110. Reakcja platform społecznościowych na wszystkie zgłoszenia
ŹRÓDŁO: OPRACOWANIE WŁASNE

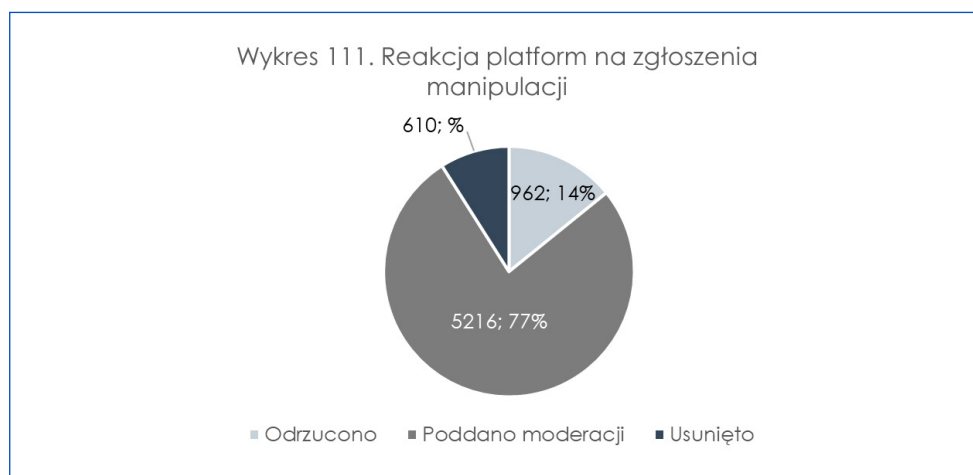
W IV kwartale 2025 roku z platform społecznościowych usunięto 13 proc. zgłoszonych przez OAD NASK treści, czyli 976 wpisów. Dalszych 71 proc. zgłoszonych treści, czyli 5323 wpisy, zostało poddanych moderacji, natomiast 16 proc. zgłoszeń, czyli 1199 przypadków, nie spotkało się z żadną widoczną reakcją ze strony moderacji.

Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń odrzuconych	137	962	100
Liczba zgłoszeń poddanych moderacji	93	5216	14
Liczba zgłoszeń usuniętych	33	610	333
Całkowita liczba zgłoszeń	263	6788	447

TABELA 11. Rodzaj naruszenia w przekazanych zgłoszeniach
ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszenia. Wśród wszystkich zgłoszeń zidentyfikowano:

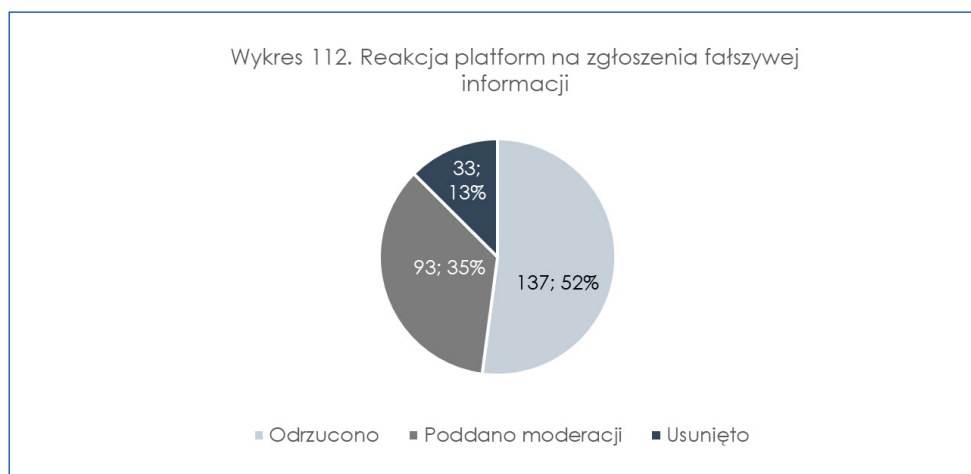
- **3,5 proc. (263 przypadki)** wpisów zawierających fałszywe informacje;
- **90,5 proc. (6788 przypadków)** wpisów zawierających manipulację;
- **6,0 proc. (447 przypadków)** wpisów związanych z podszywaniem się pod osobę lub instytucję.



WYKRES 111. Reakcja platform na zgłoszenie wystąpienia manipulacji

ŹRÓDŁO: OPRACOWANIE WŁASNE

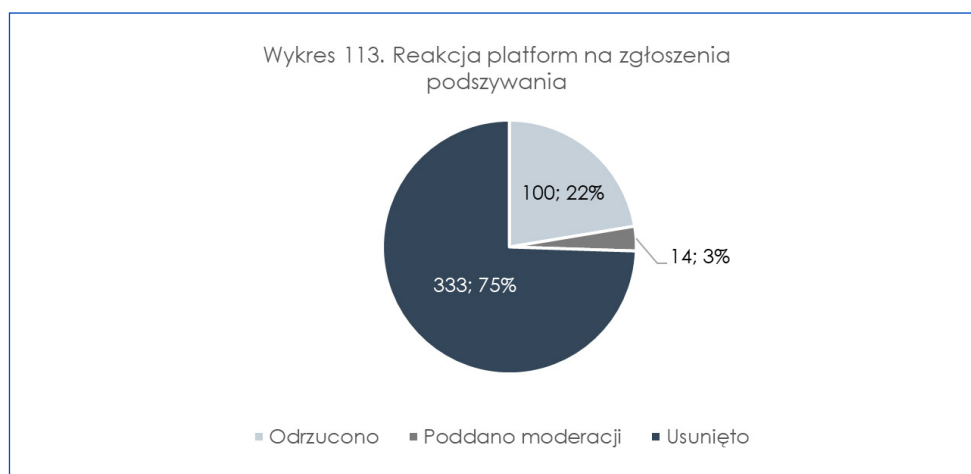
Analiza zgłoszonych treści wykazała, że platformy społecznościowe najczęściej reagowały na incydenty polegające na publikowaniu zmanipulowanych informacji. Usunięto 610 takich materiałów, co odpowiada 9 proc. zgłoszonych treści. Moderacji poddano z kolei 5216 wpisów, czyli 77 proc. zgłoszonych treści. Pozostałe 962 zgłoszenia (14 proc.) zostały odrzucone. Jako incydent wystąpienia manipulacji zakwalifikowane zostały między innymi treści publikowane przez konta-boty, wykazujące oznaki automatycznego działania.



WYKRES 112. Reakcja platform na zgłoszenie fałszywej informacji

ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku wystąpienia fałszywych informacji usunięto 33 materiały (13 proc.), a odrzucono 137 zgłoszeń (52 proc.), zaś 93 zgłoszenia (35 proc.) zostały poddane moderacji.

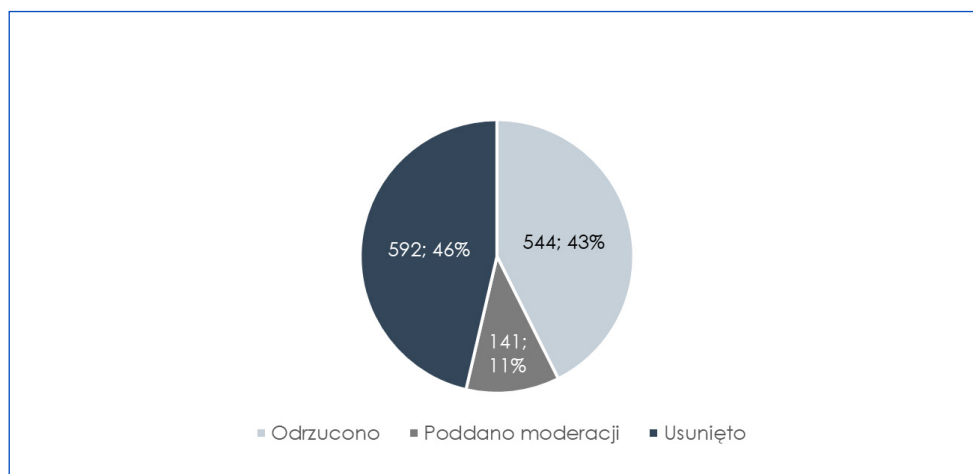


WYKRES 113. Reakcja platform na zgłoszenie podszywania się pod osobę/institucję

ŹRÓDŁO: OPRACOWANIE WŁASNE

Jeśli chodzi o zgłoszenia podszywania pod osoby publiczne lub instytucje, usunięto 333 materiały (75 proc. zgłoszeń); 14 incydentów (3 proc.) zostało poddanych moderacji, a 100 zgłoszeń (22 proc.) odrzucono.

Meta

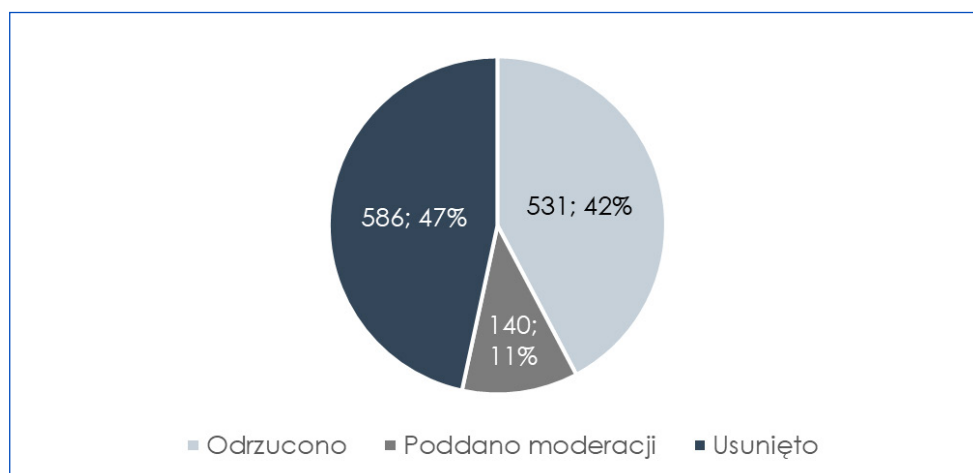


WYKRES 114. Reakcja platformy Meta na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszonych do administracji platform firmy Meta treści pochodzących z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 46 proc. zgłoszonych treści, czyli 592 materiały, 11 proc., czyli 141 materiałów, poddano moderacji, a 43 proc. zgłoszeń, czyli 544 przypadki, zostało odrzuconych.

Facebook



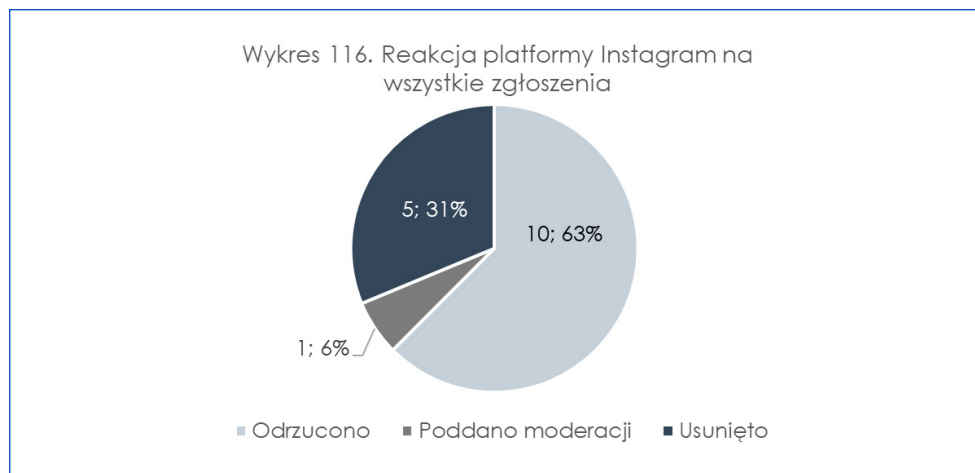
WYKRES 115. Reakcja platformy Facebook na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

W analizowanym okresie platforma Facebook usunęła 47 proc. zgłoszonych treści, czyli 586 materiałów, a 11 proc. zgłoszonych treści, czyli 140 materiałów,

zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 42 proc. zgłoszeń, czyli 531 przypadków.

Instagram

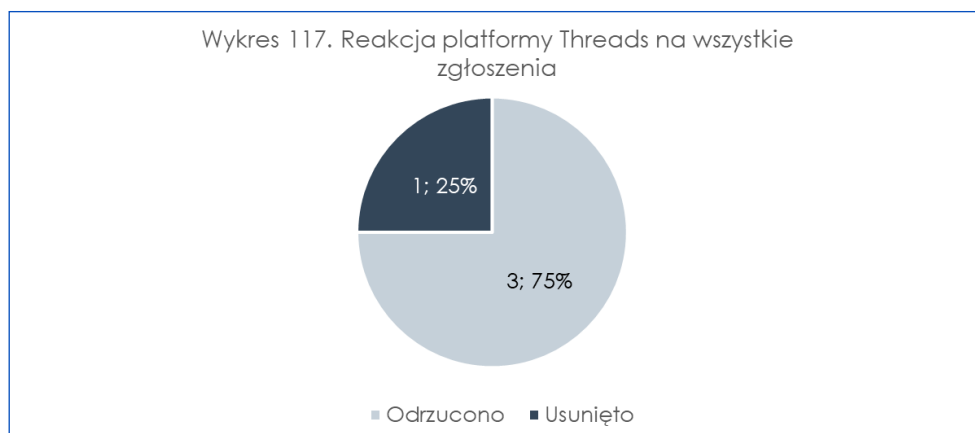


WYKRES 116. Reakcja platformy Instagram na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram usunęła 31 proc. zgłoszonych treści, czyli 5 materiałów, a 6 proc. zgłoszonych treści, czyli 1 materiał został poddany moderacji. Jednocześnie moderatorzy odrzucili 63 proc. zgłoszeń, czyli 10 przypadków.

Threads

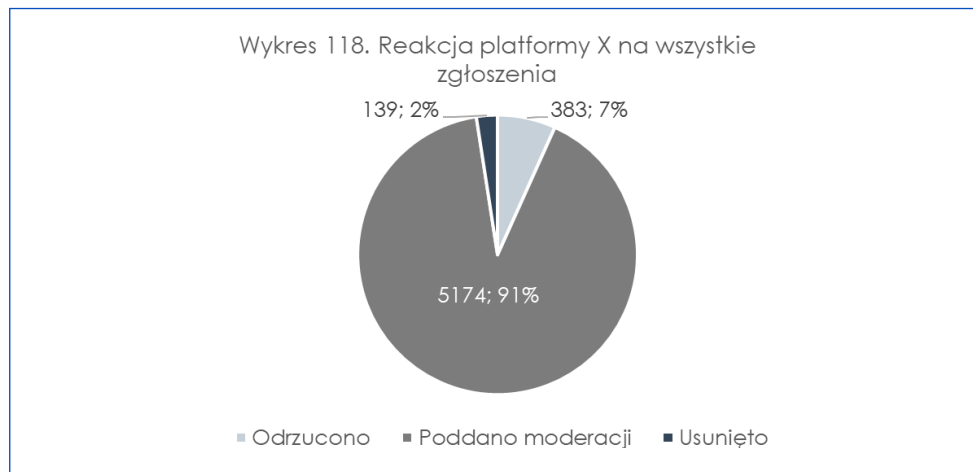


WYKRES 117. Reakcja platformy Instagram na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

W analizowanym okresie do platformy Threads zostały zgłoszone 4 materiały, z których 1 został usunięty, co stanowi 25 proc. wszystkich treści, a pozostałe 3 przypadki, czyli 75 proc. zgłoszeń, zostały odrzucone.

X.com

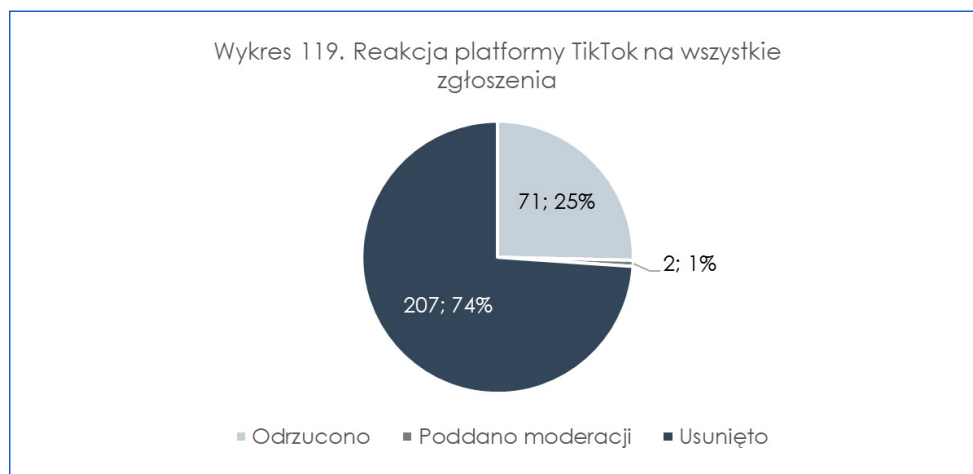


WYKRES 118. Reakcja platformy X na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 2 proc. zgłoszonych treści, czyli 139 materiałów, a 91 proc. zgłoszonych treści, czyli 5174 materiały, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 7 proc. zgłoszeń, czyli 383 przypadków.

TikTok

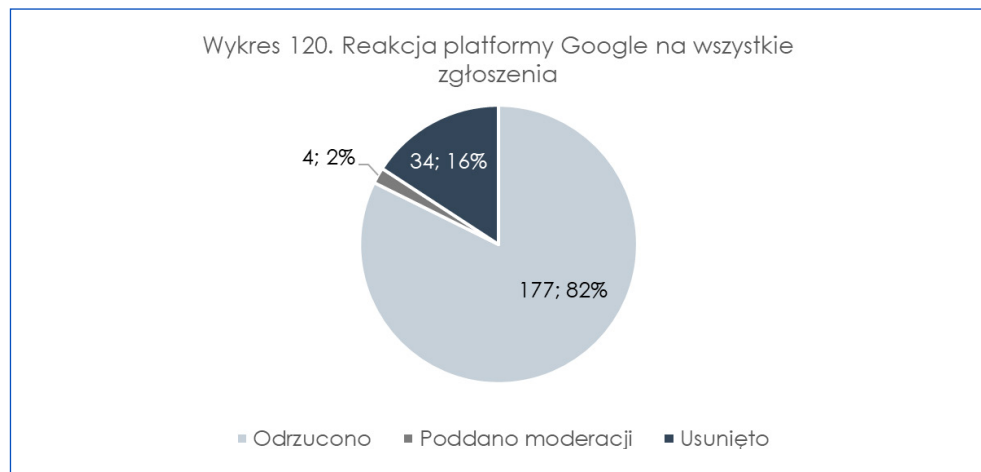


WYKRES 119. Reakcja platformy TikTok na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok usunęła 74 proc. zgłoszonych treści, czyli 207 materiałów, a 1 proc. zgłoszonych treści, czyli 2 materiały zostały poddane moderacji. Jednocześnie moderatorzy odrzucili 25 proc. zgłoszeń, czyli 71 przypadków.

Google (YouTube i Google Ads)



WYKRES 120. Reakcja platformy Google na wszystkie zgłoszenia

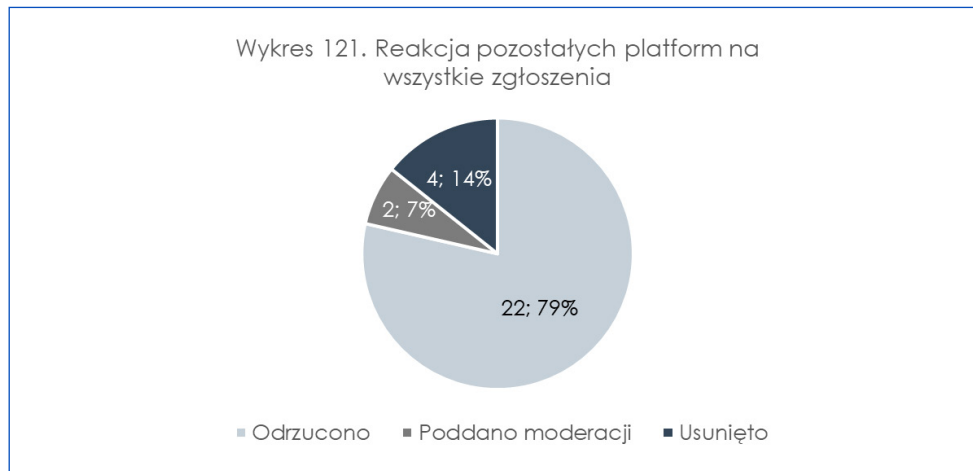
ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 16 proc. zgłoszonych treści, czyli 34 materiałów. Moderacji poddano 2 proc. treści, czyli 4 materiały. Nie usunięto natomiast 82 proc. zgłoszonych treści, czyli 177 materiałów.

Telegram

Platforma Telegram odrzuciła 2 przesłane do niej zgłoszenia.

Pozostałe platformy

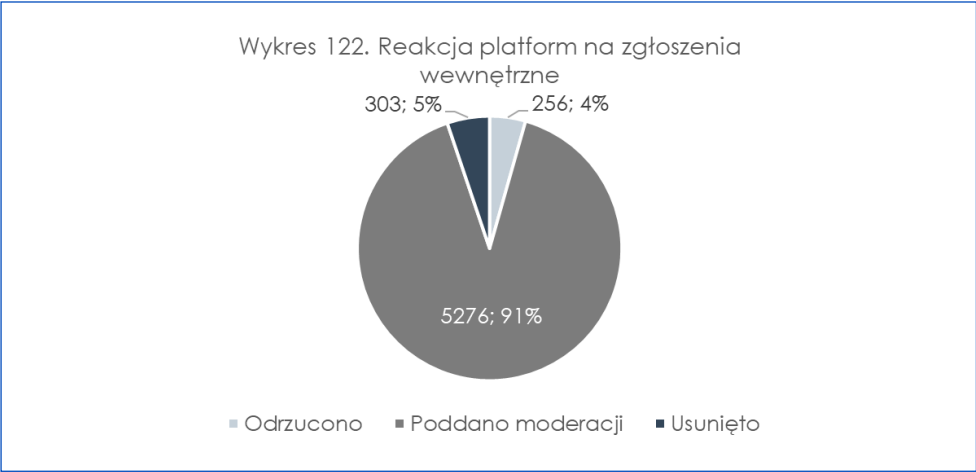


WYKRES 121. Reakcja pozostałych platform na wszystkie zgłoszenia

ŹRÓDŁO: OPRACOWANIE WŁASNE

Pozostałe platformy usunęły 14 proc. zgłoszonych treści, czyli 4 materiały i poddały moderacji 7 proc. zgłoszonych treści, czyli 2 materiały. Odrzucono natomiast 79 proc. zgłoszeń, czyli 22 przypadki.

Zgłoszenia wewnętrzne



WYKRES 122. Reakcja platform na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku zgłoszeń wewnętrznych (pochodzących z bieżącego monitoringu polskojęzycznej przestrzeni informacyjnej) w IV kwartale 2025 roku z platform społecznościowych usunięto 5 proc. zgłoszonych treści, czyli 303 wpisy. Dalszych 91 proc. zgłoszonych treści, czyli 5276 wpisów, zostało poddanych moderacji, natomiast 4 proc. zgłoszeń, czyli 256 przypadków, nie spotkało się z żadną widoczną reakcją ze strony moderacji.

Rodzaj naruszenia	Fałszywa informacja	Manipulacja
Liczba zgłoszeń odrzuconych	30	226
Liczba zgłoszeń poddanych moderacji	86	5180
Liczba zgłoszeń usuniętych	9	294
Całkowita liczba zgłoszeń	125	5700

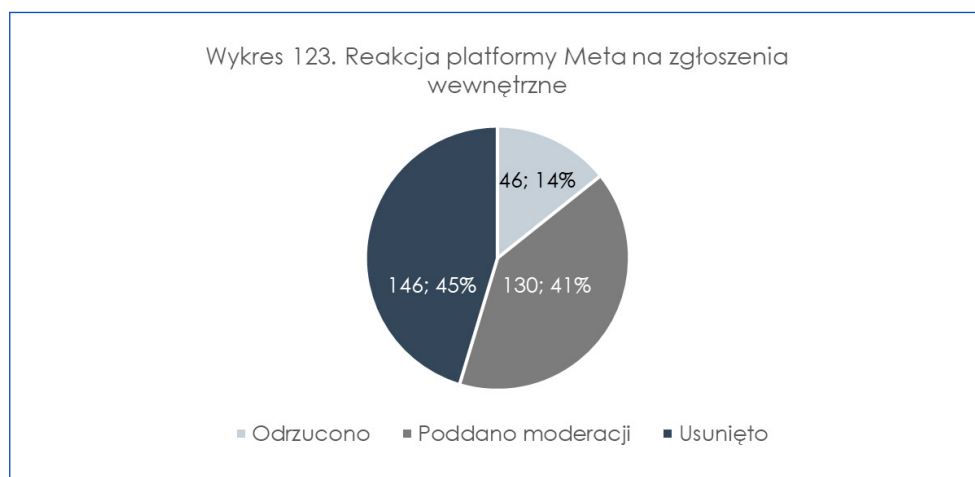
TABELA 12. Rodzaj przypisanego naruszenia w zgłoszeniach wewnętrznych

ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń. Wśród zgłoszeń wewnętrznych zidentyfikowano:

- 2 proc. (125 przypadków) wpisów zawierających fałszywe informacje;
- 98 proc. (5700 przypadków) wpisów zawierających manipulacje.

Meta

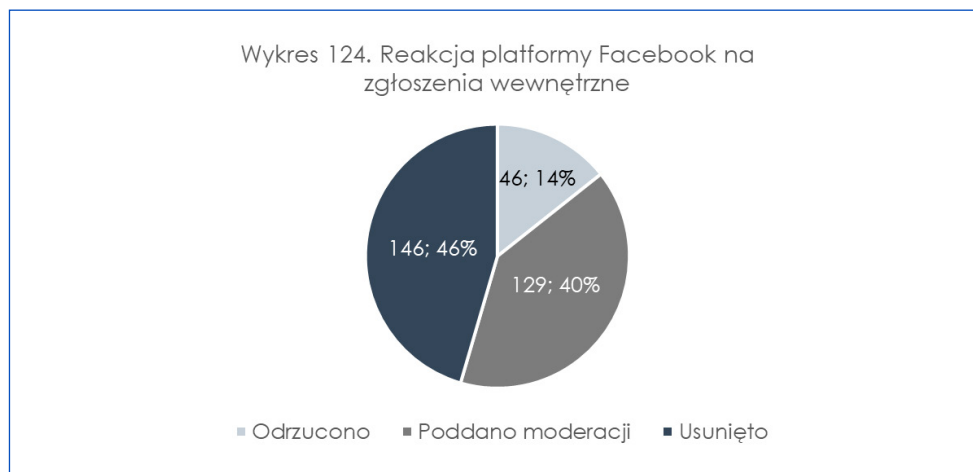


WYKRES 123. Reakcja platformy Meta na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszeń wewnętrznych przekazanych do administracji platform firmy Meta w związku z treściami pochodzącymi z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 45 proc. zgłoszonych treści, czyli 146 materiałów, 41 proc., czyli 130 materiałów, poddano moderacji, a 14 proc. zgłoszeń, czyli 46 przypadków, zostało odrzuconych.

Facebook



WYKRES 124. Reakcja platformy Facebook na zgłoszenia wewnętrzne

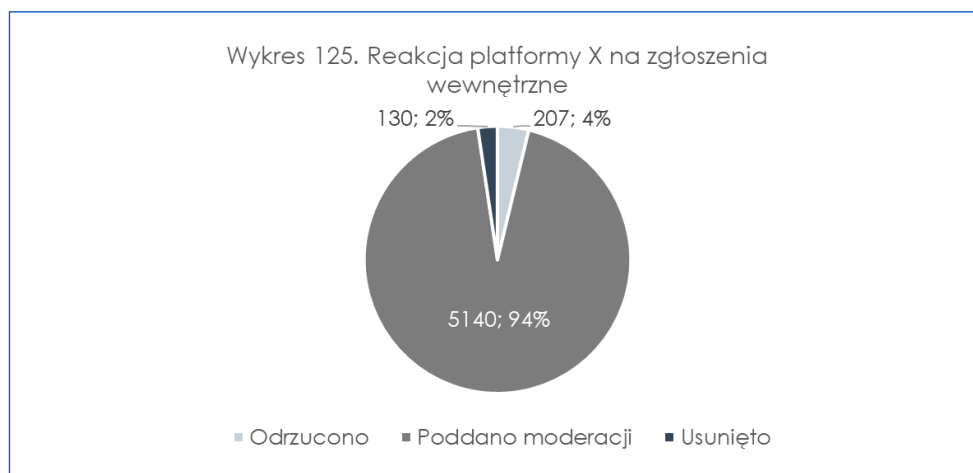
ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Facebook usunęła 46 proc. zgłoszonych treści, czyli 146 materiałów, a 40 proc. zgłoszonych treści, czyli 129 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 14 proc. zgłoszeń, czyli 46 przypadków.

Instagram

Platforma Instagram poddała moderacji [1 zgłoszony materiał](#).

X.com

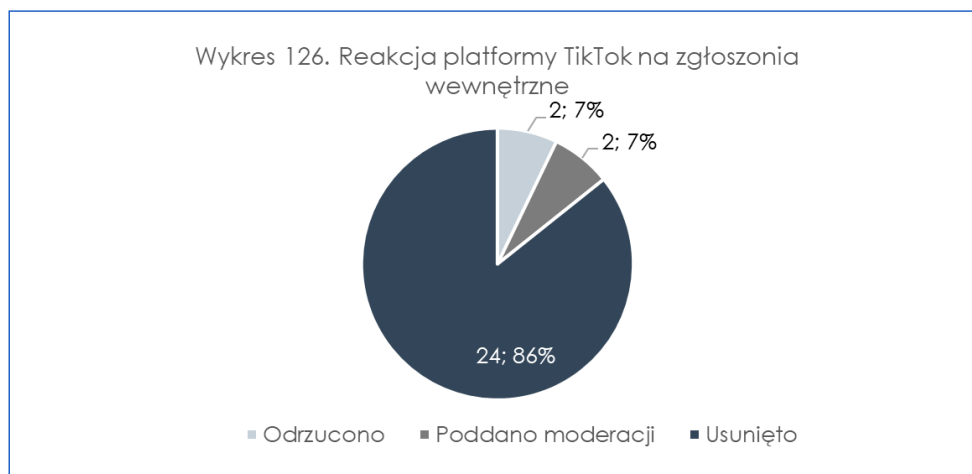


WYKRES 125. Reakcja platformy X na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 2 proc. zgłoszonych treści, czyli 130 materiałów, a 94 proc. zgłoszonych treści, czyli 5140 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 4 proc. zgłoszeń, czyli 207 przypadków.

TikTok



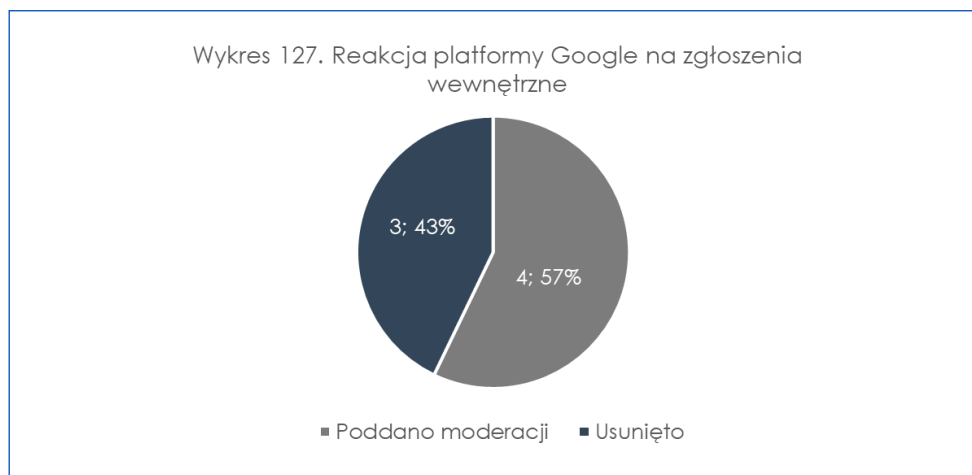
WYKRES 126.

Reakcja platformy TikTok na zgłoszenia wewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok usunęła 86 proc. zgłoszonych treści, czyli 24 materiały, a 7 proc. zgłoszonych treści, czyli 2 materiały zostały poddane moderacji. Jednocześnie moderatorzy odrzucili kolejne 7 proc. zgłoszeń, czyli 2 przypadki.

Google (YouTube i Google Ads)



WYKRES 127.

Reakcja platformy Google na zgłoszenia wewnętrzne

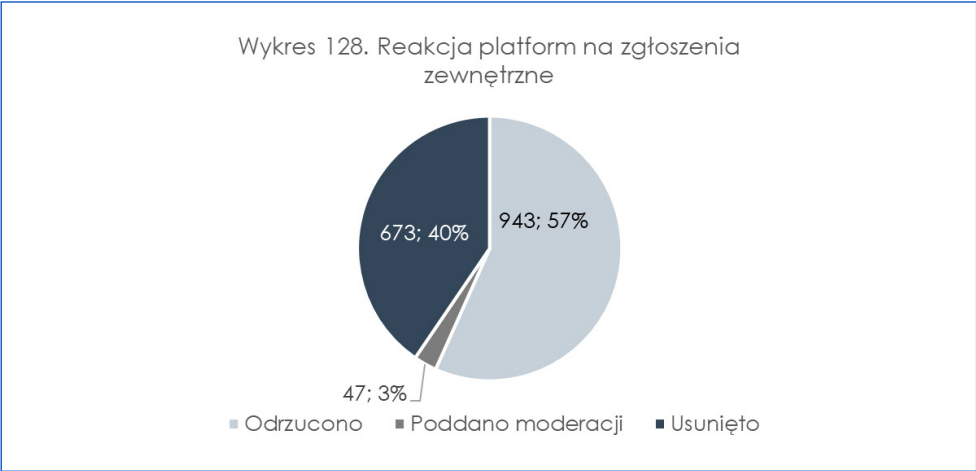
ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 43 proc. zgłoszonych treści, czyli 3 materiałów. Nie usunięto natomiast 57 proc. zgłoszonych treści, czyli 4 materiałów.

Pozostałe platformy

Administratorzy pozostałych platform odrzucili 1 przesłane zgłoszenie.

Zgłoszenia zewnętrzne



WYKRES 128. Reakcja platform na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

W przypadku zgłoszeń zewnętrznych (przekazywanych do Zespołu Szybkiego Reagowania i Wykrywania Dezinformacji przez obywateli i obywatelki, podmioty publiczne oraz inne instytucje) w I kwartale 2025 roku z platform społecznościowych usunięto 40 proc. zgłoszonych treści, czyli 673 wpisy. Dalszych 3 proc. zgłoszonych treści, czyli 47 wpisów, zostało poddanych moderacji, natomiast 57 proc. zgłoszeń, czyli 943 przypadków, nie spotkało się z żadną widoczną reakcją ze strony moderacji. 1708 materiały zgłoszone do OAD NASK zostały ocenione jako niezasadne, czyli takie, które nie stanowiły naruszenia regulaminów bądź innych regulacji prawnych.

Rodzaj naruszenia	Fałszywa informacja	Manipulacja	Podszywanie pod osobę/instytucję
Liczba zgłoszeń odrzuconych	107	736	100
Liczba zgłoszeń poddanych moderacji	7	36	4
Liczba zgłoszeń usuniętych	24	316	333
Całkowita liczba zgłoszeń	138	1088	437

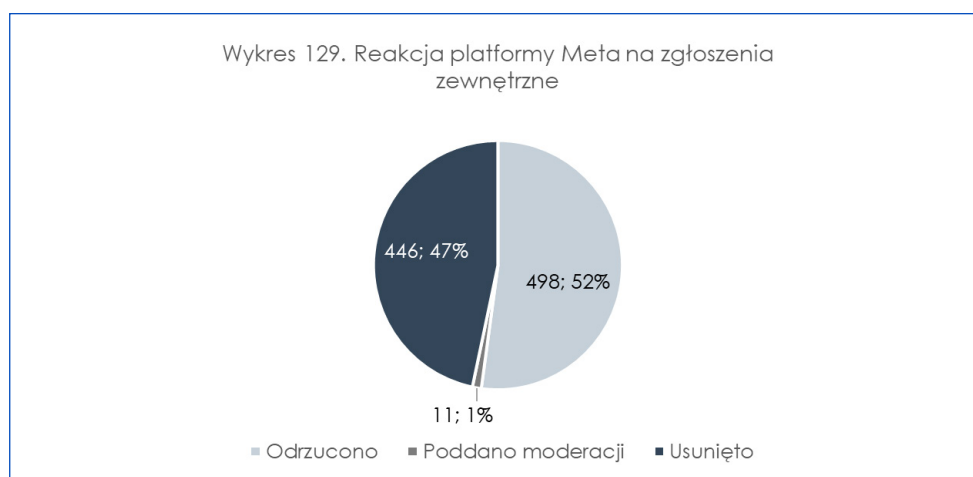
TABELA 13. Rodzaj naruszenia w zgłoszeniach zewnętrznych

ŹRÓDŁO: OPRACOWANIE WŁASNE

Tabela przedstawia podział zgłoszeń według rodzaju naruszeń³⁸. Wśród zgłoszeń zewnętrznych zidentyfikowano:

- **8 proc. (138 przypadki)** wpisów zawierających fałszywe informacje;
- **66 proc. (1088 przypadki)** wpisów zawierających manipulacje;
- **26 proc. (37 przypadki)** wpisów związanych z podszywaniem się pod osobę lub instytucję.

Meta



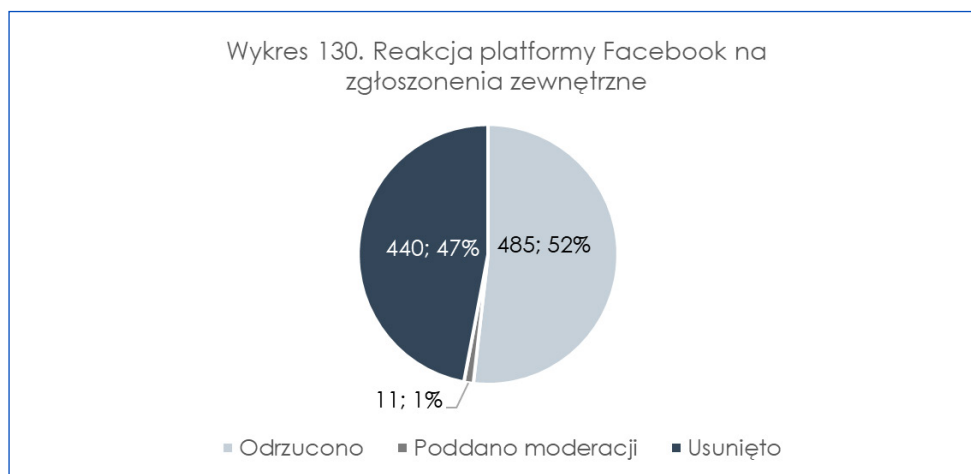
WYKRES 129. Reakcja platformy Meta na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono szczegółowe dane dotyczące wszystkich zgłoszeń zewnętrznych przekazanych do administracji platform firmy Meta w związku z treściami pochodzącymi z serwisów Facebook, Instagram i Threads. Administracja usunęła łącznie 47 proc. zgłoszonych treści, czyli 446 materiałów, 1 proc., czyli 11 materiałów, poddano moderacji, a 52 proc. zgłoszeń, czyli 498 przypadków, zostało odrzuconych.

³⁸ Podział na wskazane kategorie został dokonany wewnętrznie przez Zespół Szybkiego Reagowania i Wykrywania Dezinformacji w celu uproszczenia procedury realizacji zgłoszeń.

Facebook

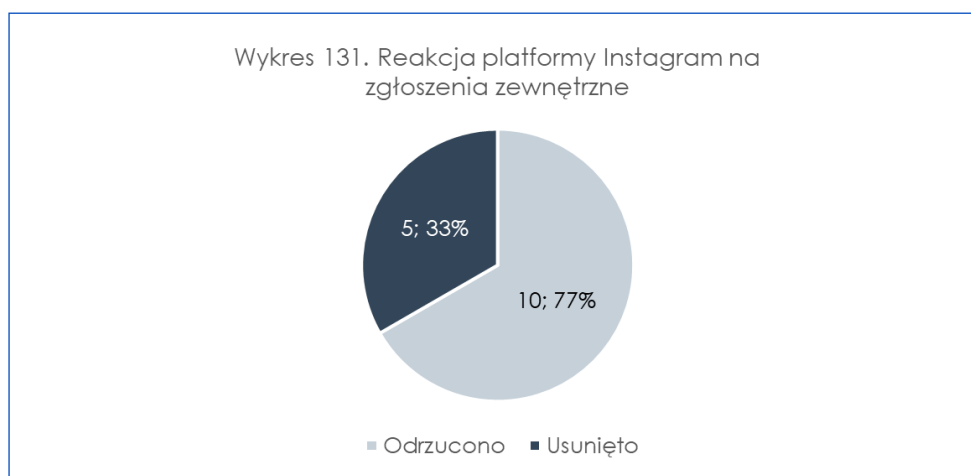


WYKRES 130. Reakcja platformy Facebook na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Facebook usunęła 47 proc. zgłoszonych treści, czyli 440 materiałów, a 1 proc. zgłoszonych treści, czyli 11 materiałów, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 52 proc. zgłoszeń, czyli 485 przypadków.

Instagram

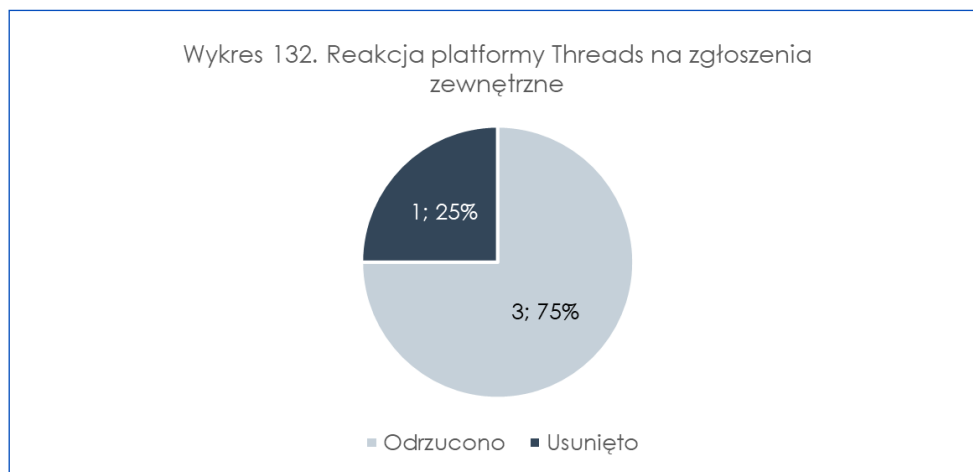


WYKRES 131. Reakcja platformy Instagram na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Instagram zdecydowała o usunięciu 33 proc. zgłoszonych treści, czyli 5 materiałów. Nie usunięto natomiast 77 proc. zgłoszonych treści, czyli 10 materiałów.

Threads

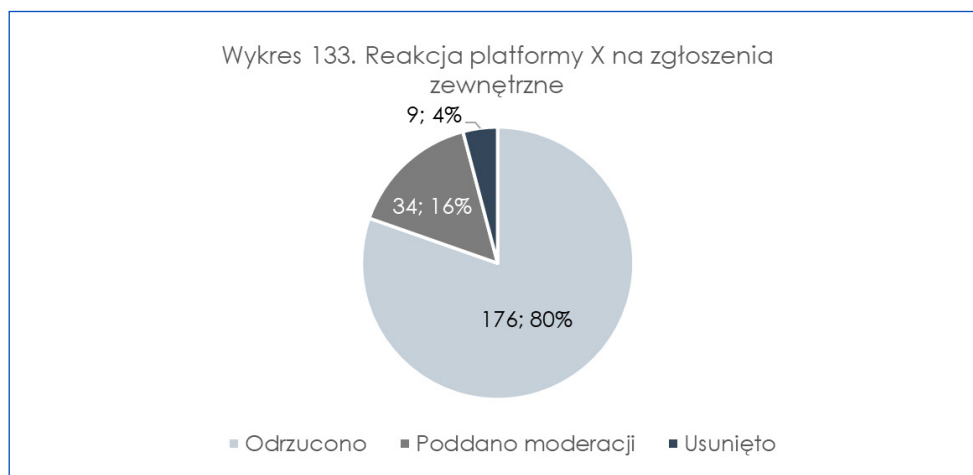


WYKRES 132. Reakcja platformy Instagram na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma Threads zdecydowała o usunięciu 75 proc. zgłoszonych treści, czyli 3 materiałów. Nie usunięto natomiast 25 proc. zgłoszonych treści, czyli 1 materiału.

X.com

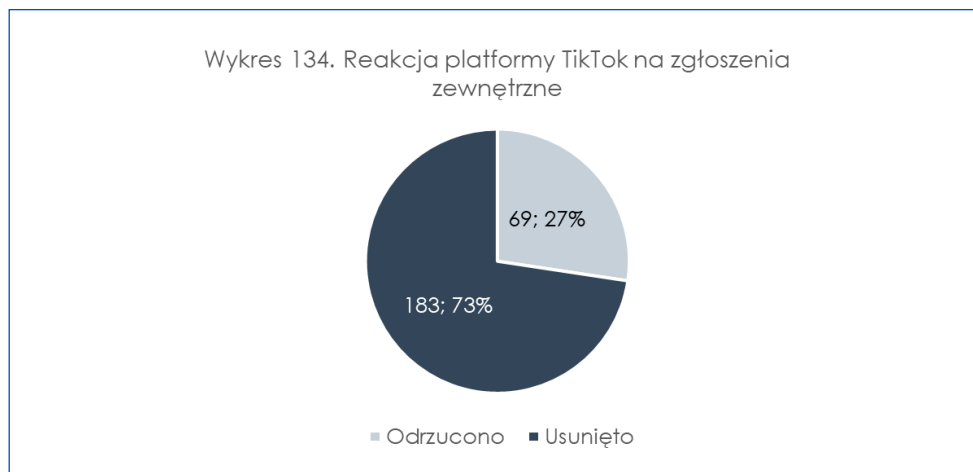


WYKRES 133. Reakcja platformy X na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma X usunęła 4 proc. zgłoszonych treści, czyli 9 materiałów, a 16 proc. zgłoszonych treści, czyli 34 materiały, zostało poddanych moderacji. Jednocześnie moderatorzy odrzucili 80 proc. zgłoszeń, czyli 176 przypadków.

TikTok

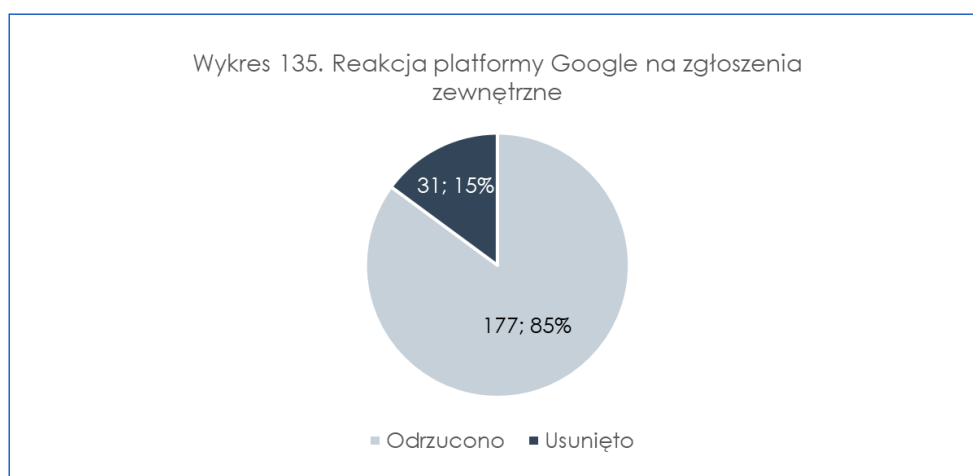


WYKRES 134. Reakcja platformy TikTok na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Platforma TikTok zdecydowała o usunięciu 73 proc. zgłoszonych treści, czyli 183 materiałów. Nie usunięto natomiast 27 proc. zgłoszonych treści, czyli 69 materiałów.

Google (YouTube i Google Ads)



WYKRES 135. Reakcja platformy Google na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Powyżej przedstawiono łączne dane dotyczące zgłoszonych do Google treści pochodzących z serwisów YouTube i Google Ads. Platforma zdecydowała o usunięciu 15 proc. zgłoszonych treści, czyli 31 materiałów. Nie usunięto natomiast 85 proc. zgłoszonych treści, czyli 117 materiałów.

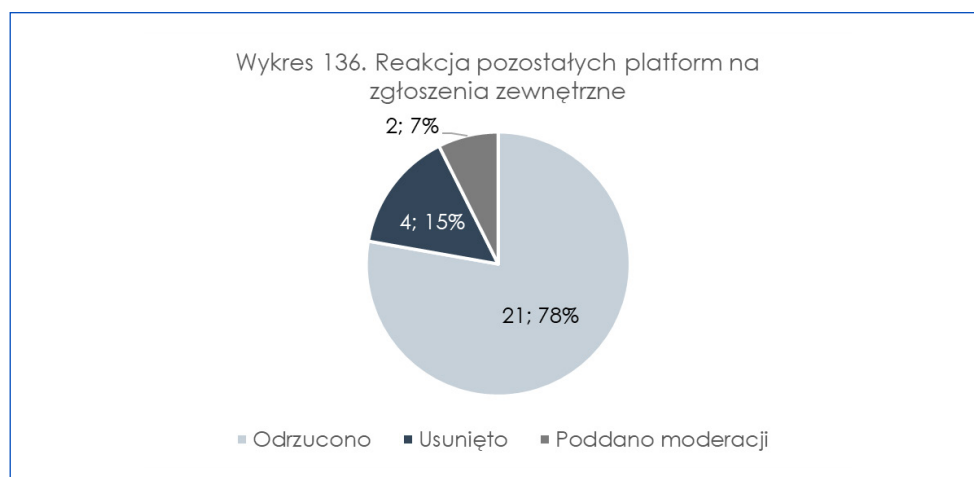
Telegram

Platforma Telegram odrzuciła 1 przesłane do niej zgłoszenie.

Discrod

Platforma Discrod odrzuciła 1 przesłane do niej zgłoszenie.

Pozostałe platformy



WYKRES 136. Reakcja pozostałych platform na zgłoszenia zewnętrzne

ŹRÓDŁO: OPRACOWANIE WŁASNE

Pozostałe platformy usunęły 15 proc. zgłoszonych treści, czyli 4 materiały i poddały moderacji 7 proc. zgłoszonych treści, czyli 2 materiały. Odrzuciły natomiast 78 zgłoszeń, czyli 21 przypadków.

Case studies

Dezinformacja na temat aktu dywersji na tory kolejowe

16 listopada 2025 roku na trasie kolejowej Warszawa–Lublin, w rejonie wsi Mika, doszło do aktu dywersji przeciwko infrastrukturze kolejowej. Z informacji przekazanych przez służby wynikało³⁹, że za zdarzenie odpowiada dwóch obywateli Ukrainy, którzy od dłuższego czasu mieli współpracować z rosyjskimi służbami specjalnymi. Zgodnie z oficjalnymi komunikatami sprawcy opuścili terytorium Rzeczypospolitej Polskiej przez przejście graniczne w Terespolu.

W następstwie tego incydentu w polskiej infosferze odnotowano pojawienie się licznych narracji o charakterze dezinformacyjnym.

1. Narracje oskarżające Ukrainę o chęć wciągnięcia Polski do wojny z Rosją

Do szczególnie szkodliwych należały wpisy podważające wiarygodność państwa i mediów. Treści te sugerowały, że polskie władze i media celowo przedstawiają zdarzenie jako atak ze strony Rosji, podczas gdy w mniemaniu autorów faktycznym sprawcą miała być Ukraina.

Wpisy te insynuowały, że takie przedstawienie wydarzeń ma na celu wywołanie poczucia zagrożenia i wciągnięcie Polski w konflikt zbrojny. Jednocześnie sugerowano, że polski rząd, rzekomo podległy obcym interesom, ukrywa prawdziwe okoliczności zdarzenia i świadomie wprowadza opinię publiczną w błąd.

³⁹ <https://X.com/WarNewsPL1/status/1990763951445975480> (dostęp 11.02.2026 godz. 13.30)

Kolejny akt dywersyjny. Najpierw podpalenia, potem ściąganie informacji na temat naszej infrastruktury, wczoraj tory i dzisiaj kolejne. Ktoś nas bardzo chce wciągnąć do wojny. Nie możemy ulec presji i wojennym strojom rządzących. Wszyscy dobrze wiemy, kto odpowiada za te ataki.

Rządzący już krzyczą o wojnie. Już im śpieszno do posyłania ludzi na front. Proponuję, by w pierwszej kolejności na wojnę szli ci, którzy do niej podżegają. Jeśli tak bardzo śpieszno im się bić za cudzy kraj, niech idą, ale Polaków niech w to nie wciągają. Bo to nie nasza wojna.

Zdjęcie przedstawia największy dziś ośrodek dezinformacyjny w Polsce.



OBRAZ 106.

Wpis zawierający fałszywą informację na temat aktu dywersji na kolei (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

Następny news, jakiego możemy się spodziewać: „Dywersja na Warszawska-Lublin? To Rosja!” 😏 Chyba nikt z rozsądkiem nie wierzy, że Rosja miałaby korzystać z tego. To prowokacja ukraińska – bo to oni chcą nas wciągnąć w ich wojnę. Czy naprawdę musimy to tolerować?

#Tusk #Dywersja



8:05 PM • 17 lis 2025 • 1975 wyświetleń

42 197 3

OBRAZ 107.

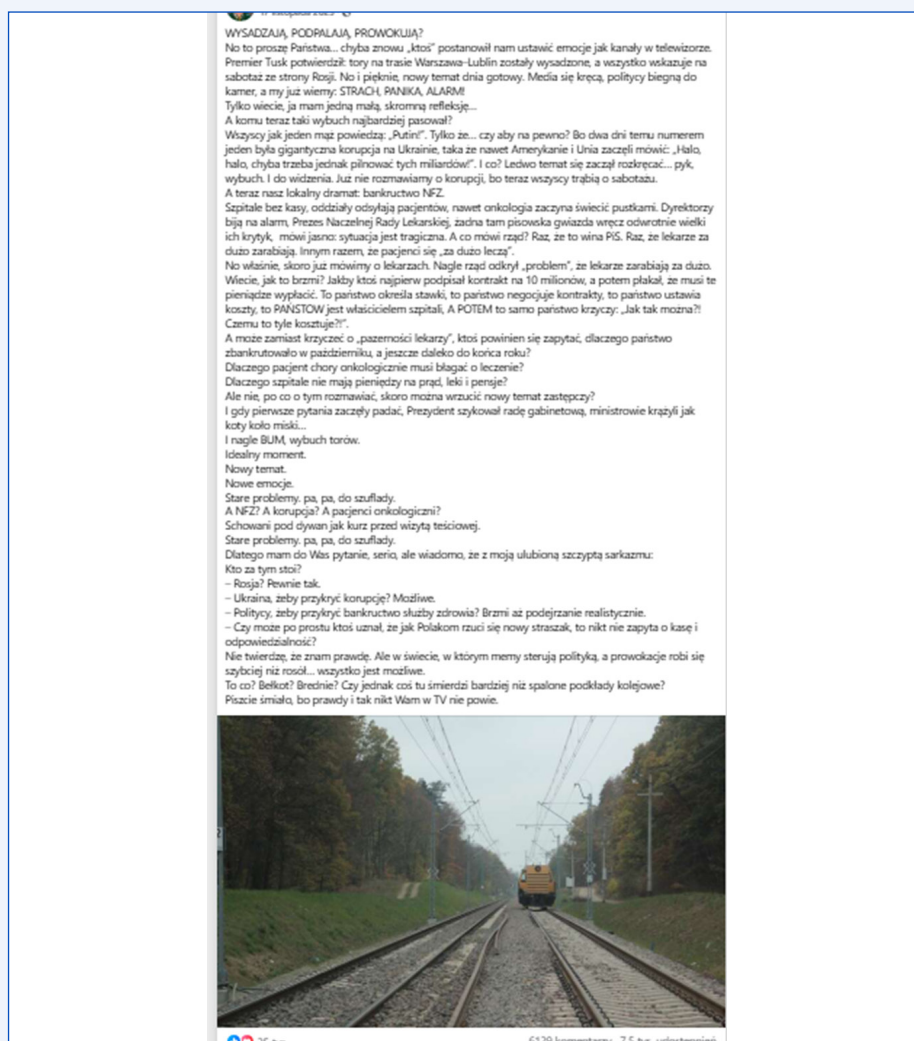
Wpis zawierający fałszywą informację na temat aktu dywersji na kolei (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

2. Narracje sugerujące celowe odciąganie uwagi od afery korupcyjnej w Ukrainie

W kontekście aktu dywersji pojawiały się również wpisy budujące narrację o celowym odwracaniu uwagi opinii publicznej od wcześniej zaistniałej afery korupcyjnej w Ukrainie. Autorzy tych treści sugerowali, że nagłośnienie incydentu kolejowego oraz wskazywanie Rosji jako sprawcy miało służyć przykryciu niewygodnych informacji dotyczących ukraińskich władz.

Również te wpisy podważały oficjalne komunikaty, insynuując, że faktycznym sprawcą aktu dywersji mogła być Ukraina, a cała narracja o rosyjskim sabotażu stanowiła element świadomej manipulacji medialnej.



OBRAZ 108.

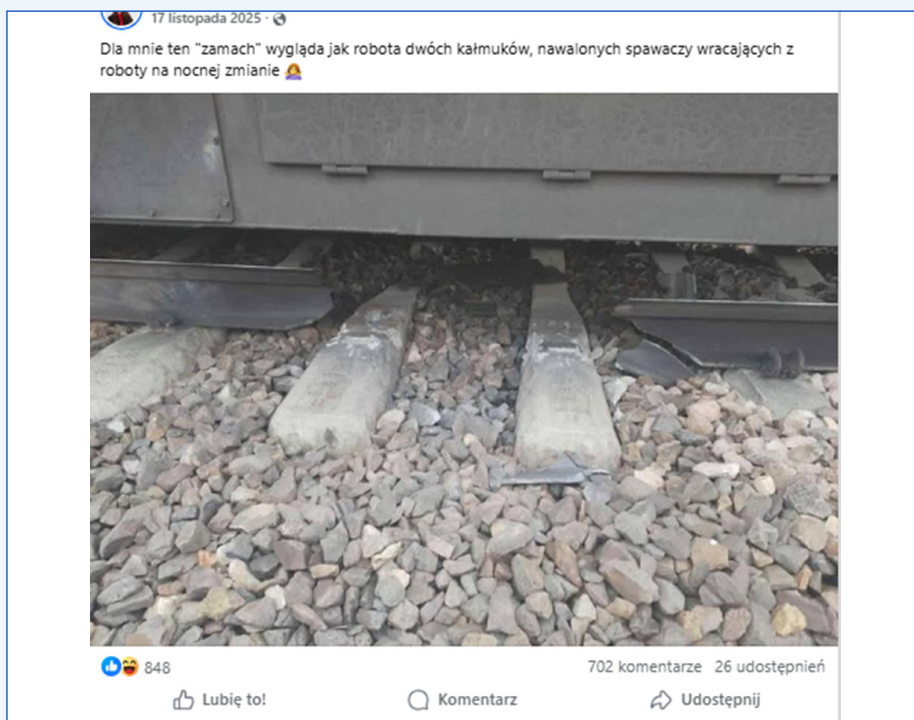
Wpis zawierający fałszywą informację na temat aktu dywersji na kolei (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

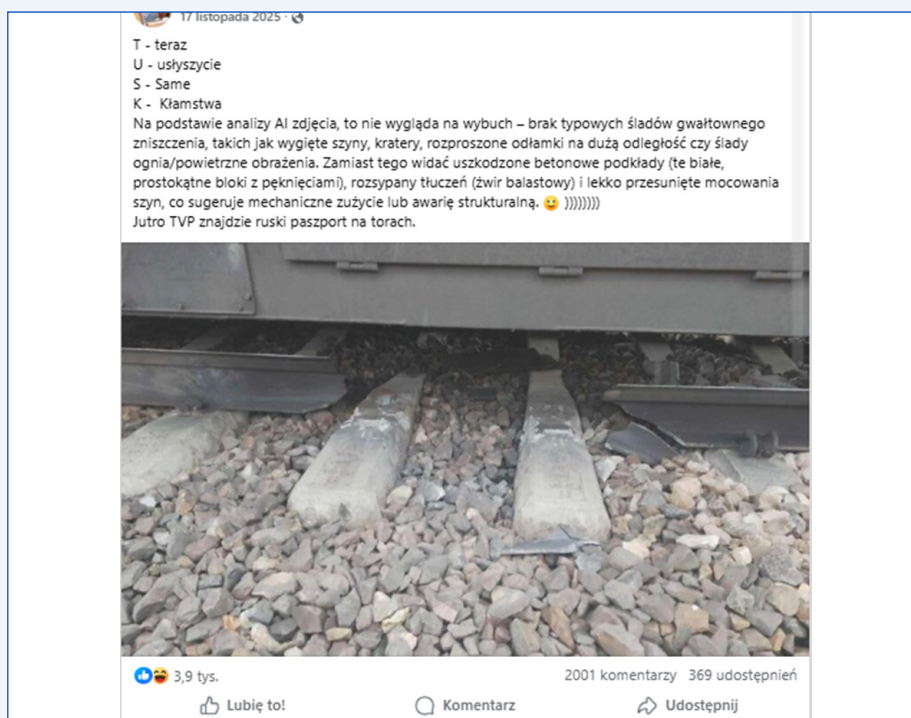
3. Kwestionowanie wystąpienia wybuchu i bagatelizowanie powagi zdarzenia

Pojawiały się także wpisy mające na celu podważenie wiarygodności informacji o akcie dywersji oraz samego faktu użycia materiałów wybuchowych. Autorzy takich treści, często powołując się na rzekome „analizy AI” lub własne interpretacje zdjęć, sugerowali, że nie doszło do wybuchu, a uszkodzenia infrastruktury kolejowej były wynikiem zwykłej awarii technicznej, zużycia elementów torowiska czy też pęknięcia szyny.

Tymi wpisami kwestionowano ustalenia służb, wskazując na brak śladów eksplozji i przedstawiając incydent jako efekt niekompetencji lub błędów nieokreślonych osób, a nie działania o charakterze sabotażowym. Tego rodzaju przekazy miały na celu wzbudzenie wątpliwości co do rzeczywistego przebiegu zdarzenia oraz osłabienie zaufania do oficjalnych komunikatów instytucji państwowych.



OBRAZ 109. Wpis zawierający fałszywą informację na temat aktu dywersji na kolei (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM

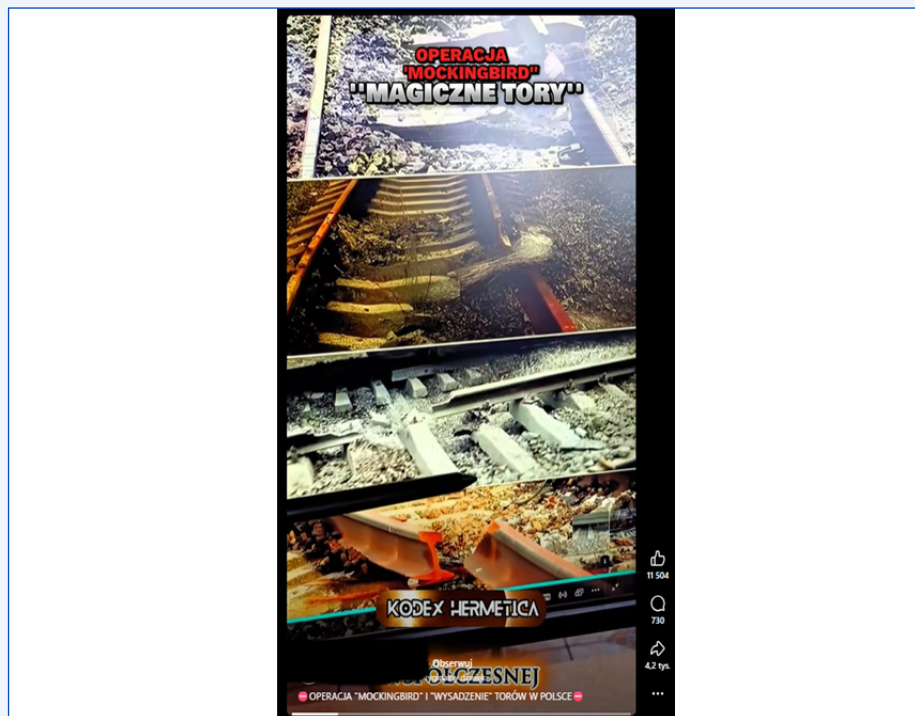


OBRAZ 110. Wpis zawierający fałszywą informację na temat aktu dywersji na kolei (zgłoszenie odrzucone)
ŹRÓDŁO: FACEBOOK.COM

4. Teoria spiskowa Mockingbird

Wśród zgłaszanych treści pojawiały się również wpisy o wyraźnie spiskowym charakterze. Autorzy tych narracji przedstawiali incydent z wysadzeniem torów jako element rzekomej „operacji Mockingbird”, odwołując się do historycznej akcji CIA z lat 50. XX wieku, mającej polegać na manipulowaniu przekazem medialnym.

We wpisach sugerowano, że informacje o domniemanym ataku na infrastrukturę krytyczną są celowo wyolbrzymiane, a dobór i sposób prezentacji szczegółów zdarzenia stanowi formę inżynierii społecznej. W tym ujęciu strach wywoływany w społeczeństwie miał być świadomie wykorzystywany przez władze do zwiększania kontroli nad obywatelami oraz sterowania nastrojami społecznymi. Narracje te podważały autentyczność zagrożenia i przedstawiały incydent nie jako realny akt dywersji, lecz jako element szerszego, ukrytego mechanizmu manipulacji.



OBRAZ 111. Wpis zawierający fałszywą informację na temat aktu dywersji na kolei (zgłoszenie odrzucone)
ŹRÓDŁO: [TIKTOK.COM](https://www.tiktok.com)

Pomimo zgłoszeń kierowanych przez zespół OAD NASK do administracji poszczególnych platform społecznościowych znaczna część wskazanych treści nadal pozostaje publicznie dostępna. W wielu przypadkach działania podejmowane przez serwisy ograniczały się jedynie do częściowej moderacji, która nie prowadziła do trwałego usunięcia dezinformujących przekazów ani do skutecznego ograniczenia ich dalszego rozpowszechniania.

Dezinformacja na temat bezpieczeństwa szczepionek

Wśród zgłaszanych treści znalazły się również materiały z zakresu dezinformacji medycznej, w szczególności dotyczące szczepień.

1. Śmierć 6-miesięcznego chłopca po podaniu szczepionki

Pierwsza z analizowanych narracji dotyczyła przypadku śmierci sześciomiesięcznego chłopca po podaniu szczepionki. Zdarzenie to było wykorzystywane do rozpowszechniania treści sugerujących ogólną szkodliwość szczepień. Z ustaleń wynikało jednak⁴⁰, że incydent był związany z podaniem przeterminowanego preparatu, przy czym informacja ta była w przekazach funkcjonujących w przestrzeni informacyjnej pomijana lub celowo zniekształcana.



OBRAZ 112.

Wpis podważający bezpieczeństwo szczepionek (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

⁴⁰ <https://www.gov.pl/web/wsse-bydgoszcz/interia---21102025-r> (dostęp 11.02.2026 godz. 13.30)

2. Badanie z Korei Południowej wykazało korelację między szczepieniami przeciw COVID-19 a wzrostem liczby zachorowań na nowotwory

Drugą grupę zgłaszanych treści stanowiły materiały odnoszące się do badania przeprowadzonego w Korei Południowej, które w narracjach dezinformacyjnych przedstawiano jako dowód na istnienie związku między szczepieniami przeciwko COVID-19 a częstszymi zachorowaniami na nowotwory.

Podobnie jak w przypadku opisanym powyżej, przekazy te upraszczały i przeinaczały ustalenia naukowe. Należy bowiem podkreślić, że autorzy publikacji jednoznacznie zaznaczyli, iż uzyskane wyniki nie potwierdzają istnienia związku przyczynowo-skutkowego, a jedynie wskazują na obserwowaną korelację wymagającą dalszych badań⁴¹. Również w tym przypadku kluczowa informacja była konsekwentnie pomijana lub zniekształcana w treściach rozpowszechnianych w polskiej infosferze.

Badanie z Korei Południowej na 8,4 mln osób wykazało korelację między szczepieniami na COVID a wzrostem raka: tarczycy +35%, żołądka +33,5%, jelita +28%, płuc +53%, piersi +20%, prostaty +69%. Jeszcze niedawno za takie uwagi represjonowano. Dziś pytam: kto w Polsce za to odpowie?

4:32 PM · 6 lis 2025 · 60,4 tys. wyświetleń

OBRAZ 113.

Wpis podważający bezpieczeństwo szczepionek (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

Obie przedstawione narracje wpisywały się w szersze kampanie antyszczepionkowe, których celem było podważanie zaufania do szczepień oraz do instytucji odpowiedzialnych za ochronę zdrowia publicznego. W obu przypadkach wykorzystywano pojedyncze zdarzenia lub wyniki badań naukowych w sposób wybiórczy i zmanipulowany, pomijając kluczowe informacje oraz kontekst, aby wzmocnić przekaz o rzekomej szkodliwości szczepionek.

Warto zaznaczyć, że pomimo zgłoszeń kierowanych do administratorów platform społecznościowych w większości przypadków nie podjęto działań mających na celu usunięcie lub ograniczenie zasięgu tych treści. W efekcie dezinformacyjne narracje antyszczepionkowe pozostawały dostępne w przestrzeni informacyjnej, nadal oddziałując na odbiorców i utrwalając fałszywe przekonania.

⁴¹ <https://pubmed.ncbi.nlm.nih.gov/41013858/> (dostęp 11.02.2026 godz. 13.30)

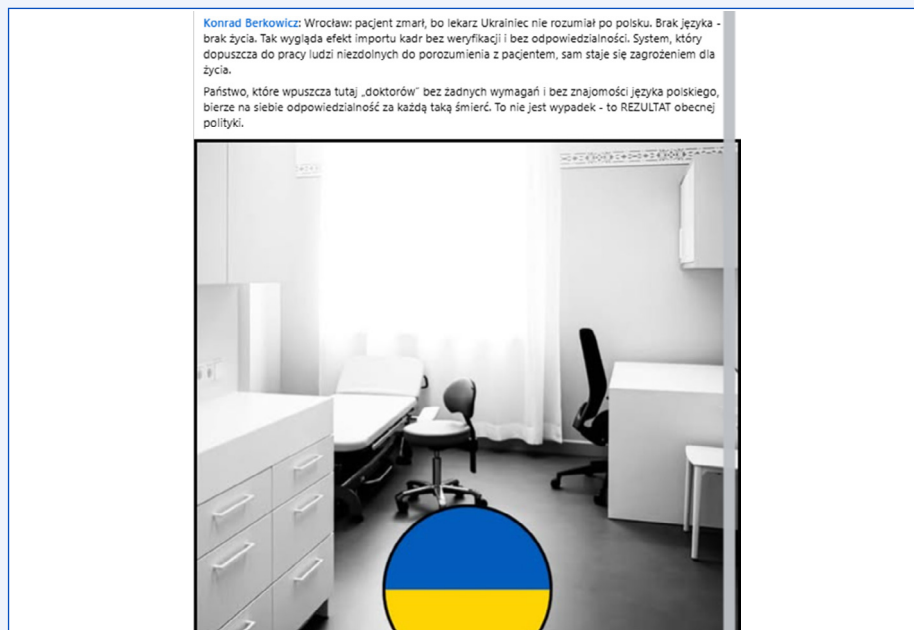
Pacjent zmarł, bo lekarz nie znał języka polskiego

W analizowanym okresie pracownicy OAD NASK otrzymywali także zgłoszenia dotyczące treści pojawiających się w mediach społecznościowych w związku ze śmiercią pacjenta we Wrocławiu podczas badania rezonansem magnetycznym. W przekazach podawano, że zgon był spowodowany brakiem znajomości języka polskiego przez personel placówki diagnostycznej, co miało skutkować niezdiagnozowaniem uczulenia u pacjenta oraz niemożnością wezwania pomocy pogotowia.

Zgłoszenia koncentrowały się na fałszywych twierdzeniach, jakoby bariera językowa była bezpośrednią przyczyną śmierci pacjenta. W rzeczywistości, jak potwierdziła Prokuratura Okręgowa we Wrocławiu oraz rzecznik Naczelnej Izby Lekarskiej⁴², informacje te nie znajdują potwierdzenia w prowadzonym śledztwie. Zespół medyczny w placówce w większości składał się z obywateli Polski, a pozostali członkowie zespołu, choć niepolskiego pochodzenia, biegle posługiwali się językiem polskim. Prokurator osobiście przesłuchał wszystkie osoby z personelu i potwierdził ich znajomość języka.

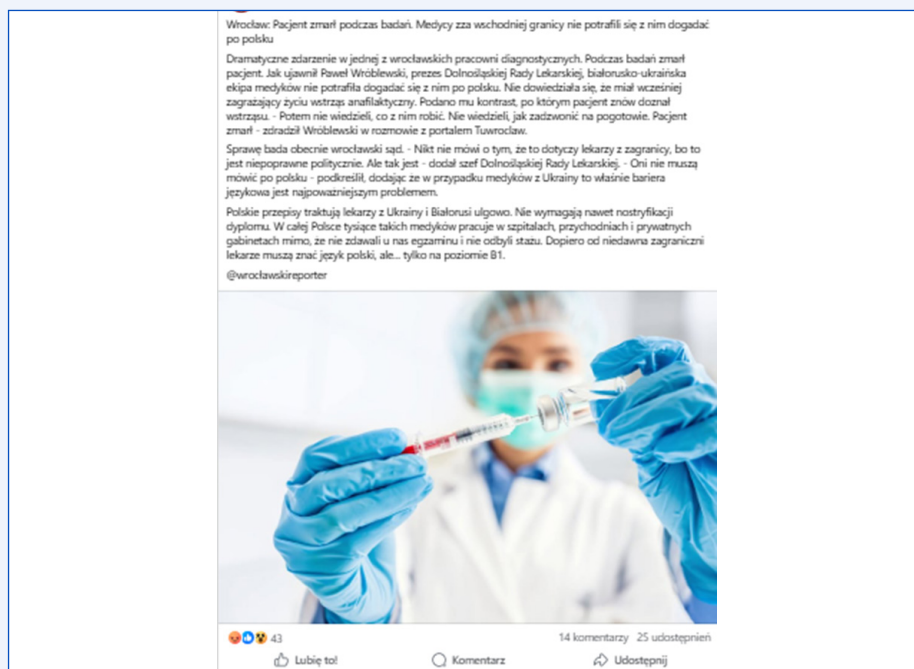
Przekazy w mediach społecznościowych celowo pomijały fakty dotyczące rzeczywistej przyczyny zdarzenia. Według przedstawicieli rodziny pacjenta kluczowymi elementami były brak identyfikacji w ankiecie czynnika wysokiego ryzyka w postaci astmy oraz chaotyczna akcja ratownicza, a nie bariery językowe. Treści zgłaszane do OAD NASK wpisywały się w szerszy nurt dezinformacyjny, mający na celu wprowadzenie odbiorców w błąd poprzez nieprawdziwe przypisywanie odpowiedzialności personelowi placówki.

⁴² <https://www.gov.pl/web/po-wroclaw/oswiadczenie-prokuratury-okregowej-we-wroclawiu-w-zwiazku-z-publicjnymi-prasowymi> (dostęp 11.02.2026 godz. 13:30)



OBRAZ 114. Wpis zawierający fałszywą informację na temat lekarza i jego rzekomej bariery językowej (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



OBRAZ 115. Wpis zawierający fałszywą informację na temat lekarza i rzekomej bariery językowej (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



OBRAZ 116. Wpis zawierający fałszywą informację na temat lekarza i jego rzekomej bariery językowej (zgłoszenie odrzucone)

ŹRÓDŁO: ŹRÓDŁO: X.COM

Pomimo oficjalnych komunikatów ze strony Prokuratury Okręgowej we Wrocławiu oraz szpitala platformy społecznościowe w większości przypadków nie podejmowały działań w celu usunięcia lub moderacji dezinformujących treści. W efekcie fałszywe informacje dotyczące przyczyn śmierci pacjenta pozostawały dostępne w przestrzeni publicznej, nadal mogąc wprowadzać odbiorców w błąd.

Reklamy na platformie YouTube z wykorzystaniem deepfake'a z Prezydentem Karolem Nawrockim

Materiały w serwisie YouTube miały formę reklam wideo wykorzystujących technologię deepfake oraz bezprawnie posługiwały się wizerunkiem Prezydenta Rzeczypospolitej Polskiej Karola Nawrockiego, a także dziennikarza Bogdana Rymanowskiego. W przekazie sugerowano istnienie rzekomego „narodowego programu wsparcia dochodu”.

Reklamy zawierały odnośniki prowadzące do zewnętrznych stron internetowych, które w większości przypadków w momencie analizy były już niedostępne lub nieaktywne. Istotnym problemem w procesie zgłaszania tych treści był brak możliwości pozyskania bezpośredniego, jednoznacznego adresu URL do reklamy. Zgłoszenia często zawierały link do materiału wideo, przed którym wyświetlana była reklama, bądź odnosiły się wyłącznie do strony głównej platformy YouTube, co znacząco utrudniało identyfikację i dalszą analizę treści.



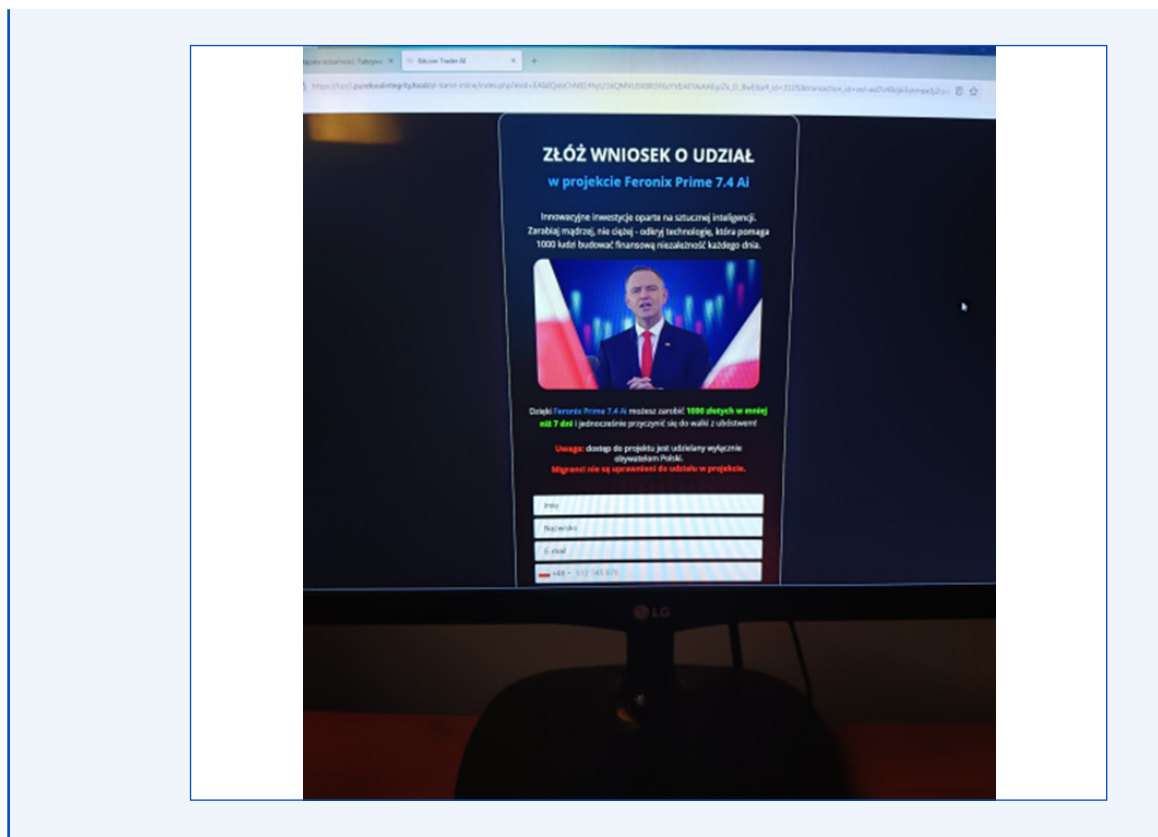
OBRAZ 117. Reklama deepfake z Prezydentem Karolem Nawrockim (reklama zablokowana)

ŹRÓDŁO: NADESŁANE PRZEZ OBYWATELA, YOUTUBE.COM



OBRAZ 118. Reklama deepfake z Prezydentem Karolem Nawrockim (reklama zablokowana)

ŹRÓDŁO: NADESŁANE PRZEZ OBYWATELA, YOUTUBE.COM



OBRAZ 119.

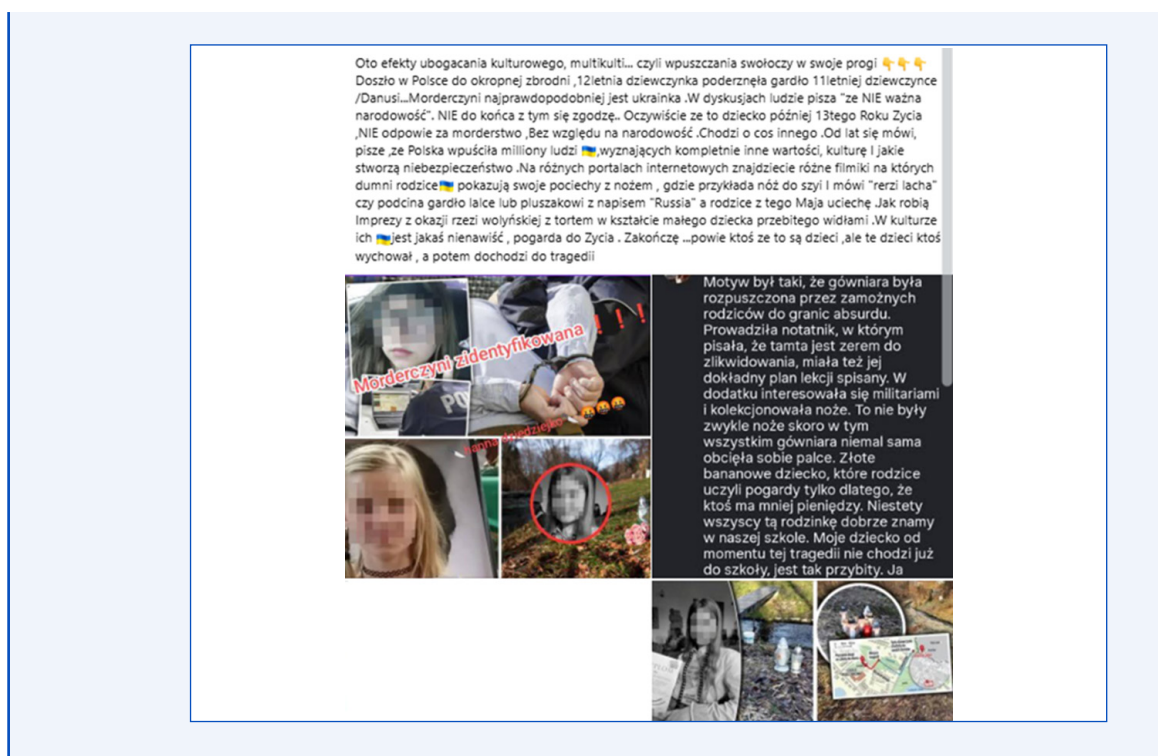
Reklama deepfake z Prezydentem Karolem Nawrockim (reklama zablokowana)

ŹRÓDŁO: NADESŁANE PRZEZ OBYWATELA, YOUTUBE.COM

Analitycy OAD NASK podczas analizy uzyskali bezpośredni link do jednej z reklam, co umożliwiło administracji platformy YouTube jej usunięcie. Na podstawie otrzymanych zrzutów ekranu oraz weryfikacji w bibliotece reklam Google pracownicy Ośrodka nie stwierdzili, aby podobne reklamy były w dalszym ciągu widoczne i dostępne dla użytkowników platformy.

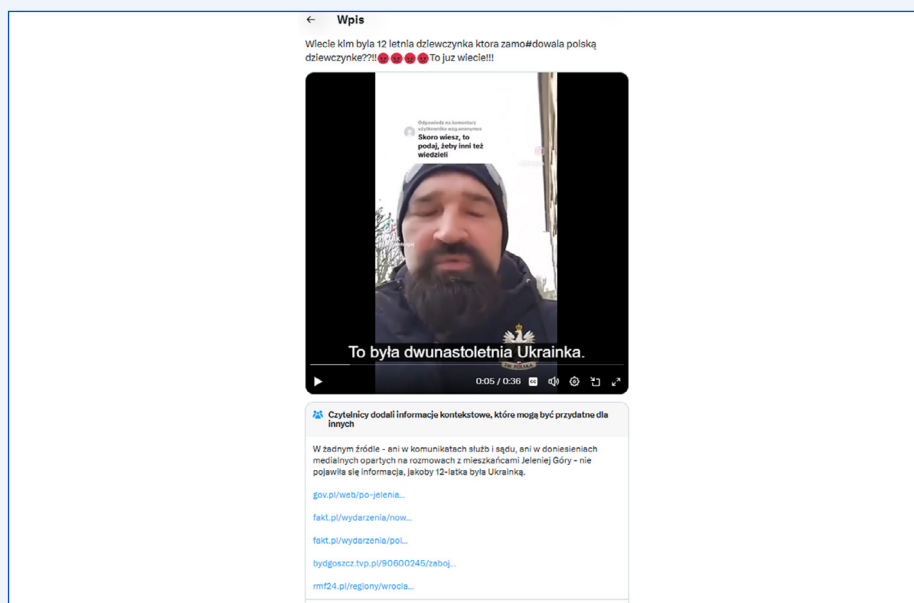
Zabójstwo 11-latki w Jeleniej Górze

15 grudnia 2025 roku, przed godziną 15:00, policja w Jeleniej Górze otrzymała zgłoszenie o odnalezieniu ciała 11-letniej dziewczynki w rejonie jednej ze szkół podstawowych. W związku z tym tragicznym zdarzeniem w mediach społecznościowych nastąpił wzrost liczby treści dezinformacyjnych, w których sugerowano, że sprawczynią miała być osoba narodowości ukraińskiej. Fałszywe informacje dotyczące obywatelstwa podejrzanej były szeroko rozpowszechniane, zwłaszcza w okresie poprzedzającym ukazanie się oficjalnych komunikatów Ministerstwa Spraw Wewnętrznych i Administracji.



OBRAZ 120. Wpis zawierający fałszywą informację na temat narodowości sprawczyni zabójstwa z Jeleniej Góry (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM



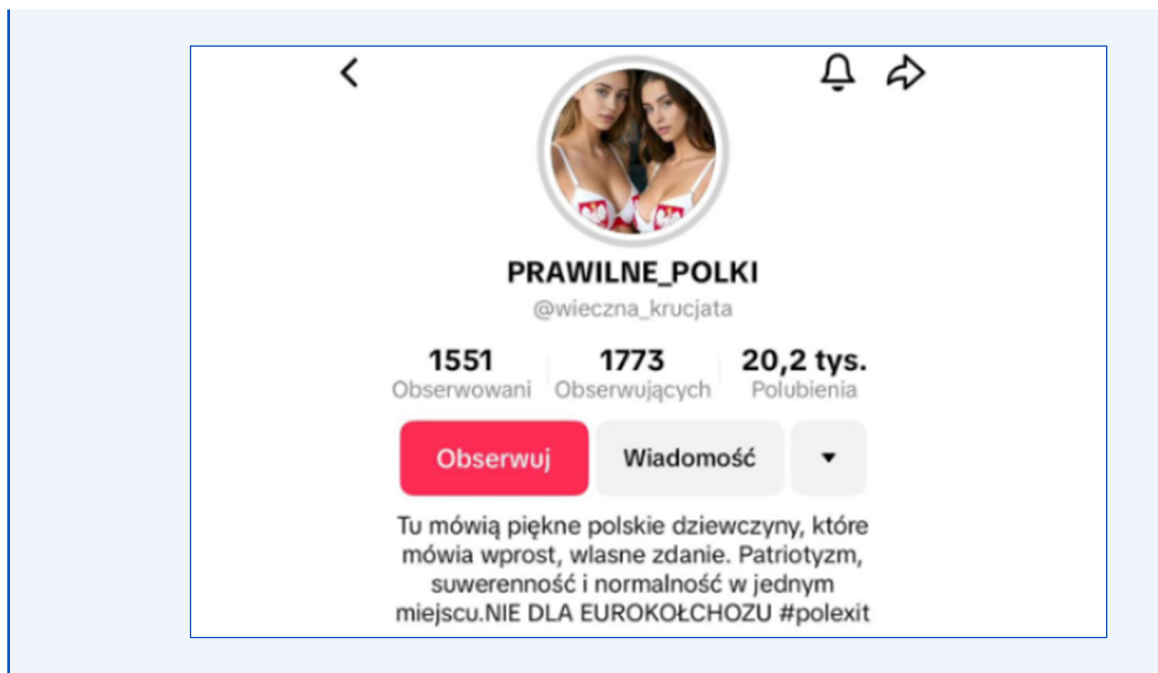
OBRAZ 121. Wpis zawierający fałszywą informację na temat narodowości sprawczyni zabójstwa z Jeleniej Gór (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

Pomimo późniejszego, jednoznacznego odniesienia się do tematu ze strony służb, które potwierdziły, że oskarżona nie była obywatelką Ukrainy, platformy społecznościowe w większości przypadków odrzucały zgłoszenia dotyczące usunięcia tego typu wpisów. Ograniczały się jedynie do minimalnej moderacji, pozostawiając fałszywe treści w przestrzeni publicznej.

Konto „Prawilne_Polki”

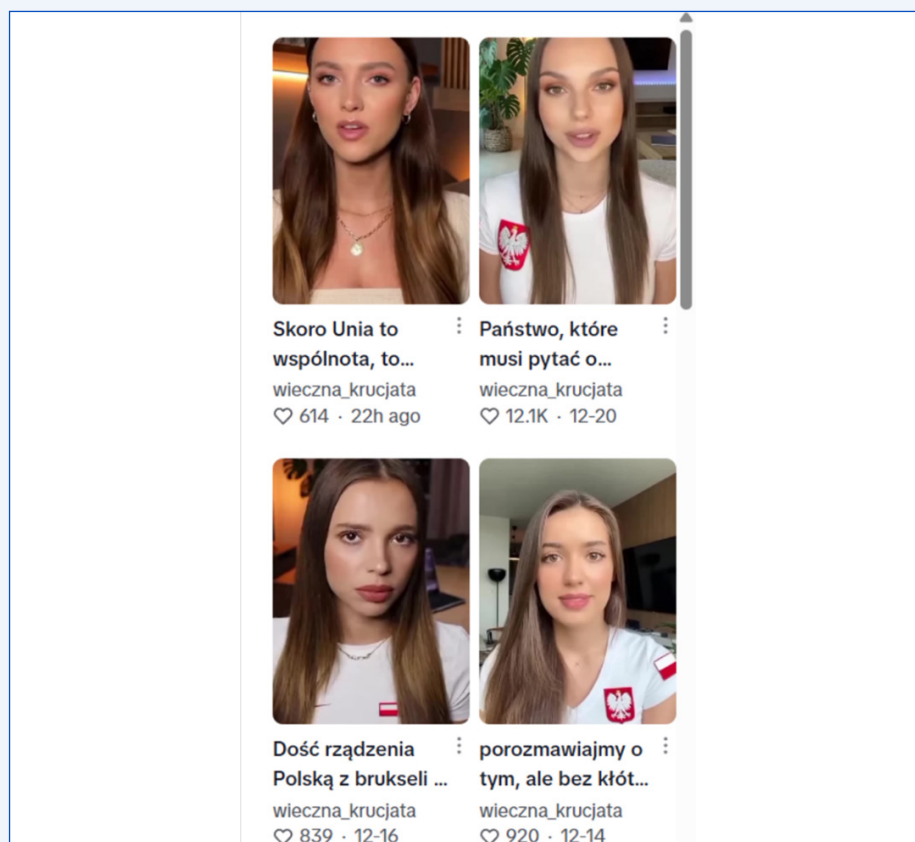
W grudniu 2025 roku na platformie TikTok zaobserwowano zorganizowaną kampanię dezinformacyjną wykorzystującą materiały generowane przez sztuczną inteligencję, mającą na celu promowanie postaw antyunijnych i narracji sprzyjającej opuszczeniu przez Polskę Unii Europejskiej. Kampania była skierowana głównie do użytkowników w wieku 15–30 lat, a konta wykorzystywały wizerunek nieistniejących młodych kobiet odwołujących się do wartości patriotycznych i tradycji.



OBRAZ 122.

Profil „Prawilne_Polki” na platformie TikTok (konto zablokowane)

ŹRÓDŁO: TIKTOK.COM



OBRAZ 123.

Filmy zamieszczone przez konto "Prawilne_Polki" (filmy zablokowane)

ŹRÓDŁO: TIKTOK.COM

OAD NASK zidentyfikował 11 podobnych kont rozpowszechniających szkodliwe treści, z czego 6 funkcjonowało na TikToku, 4 na X i 1 na Facebooku. Administracja platformy TikTok zablokowała wszystkie zgłoszone profile, natomiast konta na X i Facebooku w badanym okresie pozostały aktywne.

Podszywanie pod organizację rządową

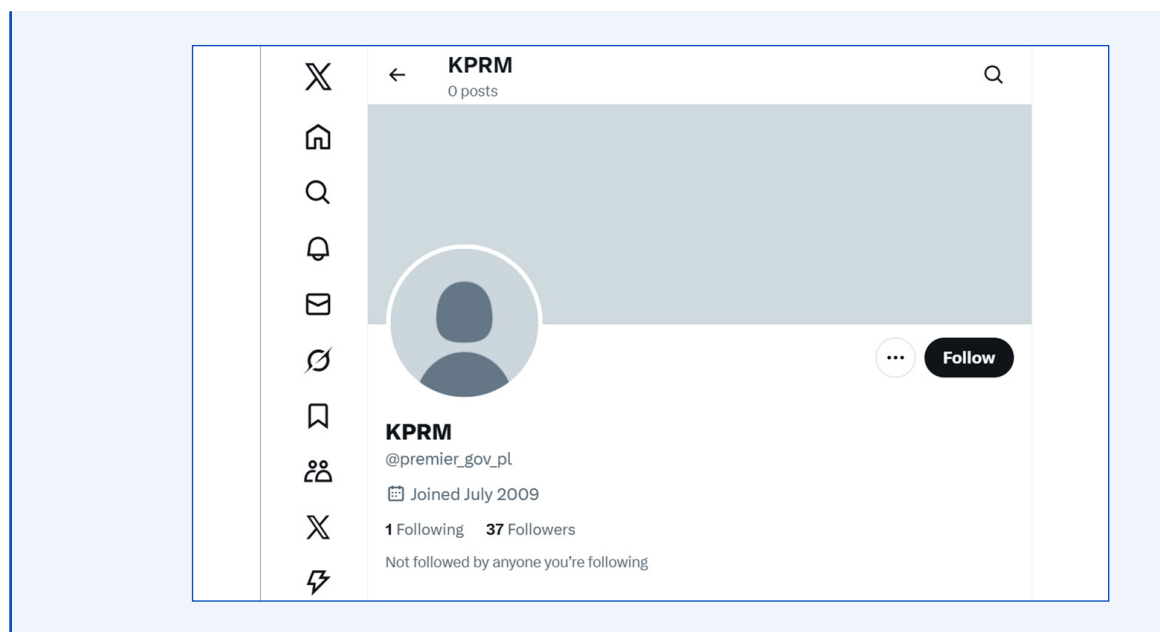
W IV kwartale 2025 roku zidentyfikowano konta na platformie X podszywające się pod instytucje rządowe. Nazwy tych profili często zawierały elementy takie jak „ministerstwo”, „gov” czy odniesienia do straży granicznej.



OBRZ 124. Profil podszywający się pod MKiDN (konto zablokowane)
ŹRÓDŁO: X.COM



OBRZ 125. Profil podszywający się pod MSiT (konto zablokowane)
ŹRÓDŁO: X.COM



OBRAZ 126. Profil podszywający się pod KPRM (konto zablokowane)

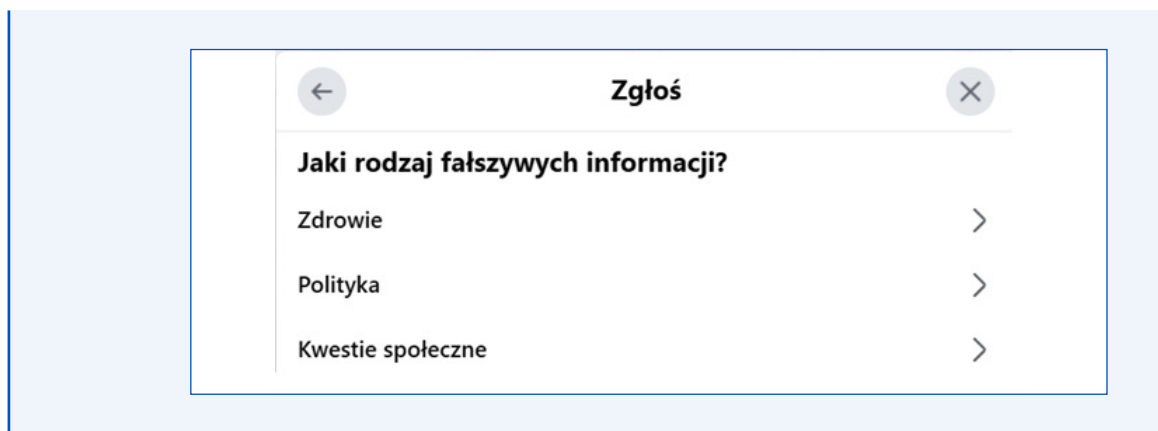
ŹRÓDŁO: [X.COM](https://x.com)

Zgłoszone konta zostały w większości poddane moderacji, a część z nich usunięta przez administrację serwisu X, co ograniczyło ich dalszą widoczność i wpływ na odbiorców.

Podsumowanie: współpraca z platformami – trudności i wyzwania

Platformy Meta – Facebook, Instagram i Threads

- A. Choć platforma Meta umożliwia zgłaszanie treści zawierających fałszywe informacje, kategoria ta odnosi się jedynie do trzech obszarów: kwestii zdrowotnych, politycznych i społecznych. Brakuje natomiast kategorii służącej do zgłaszania treści zmanipulowanych, dezinformacyjnych. W regulaminie platformy nie istnieje bowiem takie naruszenie jak „dezinformacja”.

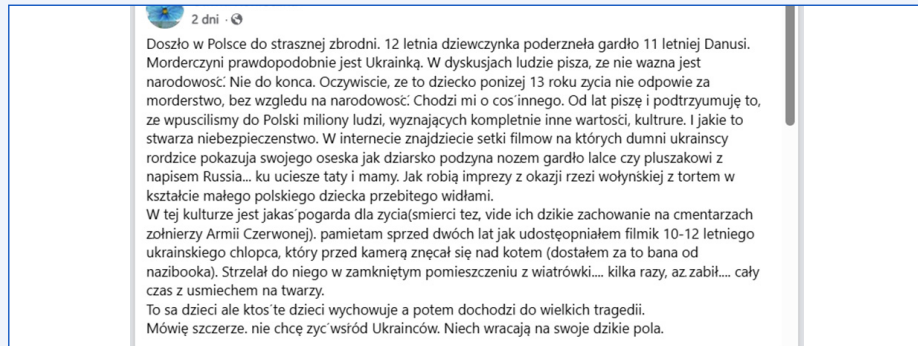


OBRAZ 127.

Rodzaje fałszywych informacji możliwe do zgłoszenia na platformie Facebook

ŹRÓDŁO: FACEBOOK.COM

- B. Brak spójności w moderowaniu treści. Zdarza się, że moderatorzy platformy Meta, oceniając podobne materiały zawierające dezinformację, usuwają jedynie część z nich, pozostawiając pozostałe bez reakcji. W badanym okresie było to widoczne na przykładzie rozpowszechniania fałszywej informacji o Ukraince, która miała być rzekomo sprawczynią zabójstwa popełnionego na 11-letniej Polce. Część wpisów zawierających tę tezę zostało zablokowanych, inne wciąż są dostępne dla użytkowników platformy Facebook.



OBRAZ 128.

Materiał zawierający fałszywą informację o zabójstwie 11-latk (wpis zablokowany)

ŹRÓDŁO: FACEBOOK.COM



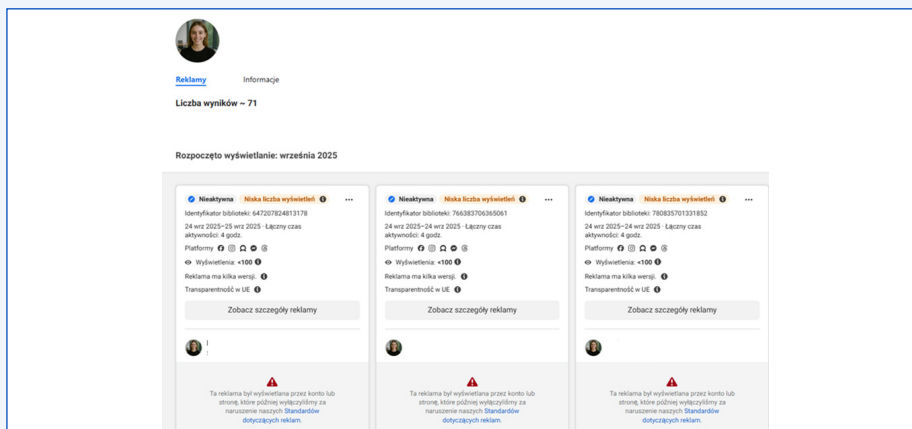
OBRAZ 129.

Materiał zawierający fałszywą informację o zabójstwie 11-latk (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

- C. Analitycy OAD nie zawsze otrzymują informację zwrotną na temat statusu zgłoszonej publikacji. Z tego powodu czasami niemożliwe jest zamknięcie prowadzonej sprawy i tym samym poinformowanie zgłaszającego podmiotu o akcji podjętej przez platformę społecznościową. Z tego względu zgłoszenia, które nie doczekają się żadnej reakcji przez okres dwóch tygodni, zostają uznane za odrzucone.
- D. Przy odrzuceniu zgłoszenia platformy nie podają uzasadnienia swojej decyzji. Z poziomu narzędzia do obsługi zgłoszeń nie ma również dostępu do pierwotnej treści zgłoszenia, a więc ewentualne złożenie odwołania jest utrudnione.

- E.** W ramach zapowiedzi transparentnych działań przed wyborami przedstawiciele platform deklarowali szczególny nacisk na monitorowanie tego obszaru. Fałszywe informacje dotyczące samego procesu wyborczego czy zasad głosowania wskazano jako kategorię treści, które będą poddawane moderacji. Nie pojawiło się jednak jasne stanowisko co do innych rodzajów dezinformacji wyborczej, jak na przykład popularnej narracji sugerującej ingerencję Unii Europejskiej w wybory w Polsce, czyli tak zwanego „wariantu rumuńskiego”. Z praktyki OAD NASK wynika, że treści zawierające takie twierdzenia nie były usuwane przez platformę mimo ich negatywnego wpływu na dyskurs wyborczy w Polsce.
- F.** Platforma Meta nie dysponuje skutecznymi, proaktywnymi mechanizmami wykrywania i blokowania treści typu deepfake w krajach Unii Europejskiej. Skutkuje to między innymi pojawianiem się reklam i wpisów wykorzystujących wizerunek osób publicznie znanych w celu zachęcania do udziału w fałszywych inwestycjach czy nakłaniania do kupna podejrzanych produktów.
- G.** Istnieją techniczne trudności w identyfikowaniu szkodliwych reklam publikowanych na platformach należących do firmy Meta. Informacje zawarte w udostępnionej przez firmę bibliotece reklam są aktualizowane z opóźnieniem. Zdarza się, że użytkownicy serwisu Facebook natrafiają na szkodliwe treści, które nie są jeszcze widoczne w rejestrze, co znacząco utrudnia zgłaszanie ich do zespołu moderującego platformę.
- H.** Problemem jest praktyka serwisu, która pozwala na dalsze funkcjonowanie reklamodawców mimo wcześniejszego naruszenia przez nich wewnętrznego regulaminu platformy oraz stosowania przestępczych praktyk związanych z wyludzaniem danych i środków finansowych. Takie podejście umożliwia użytkownikom kontynuowanie szkodliwych działań w serwisie i sprawia, że przeciwdziałanie manipulacyjnym praktykom staje się nieefektywne.

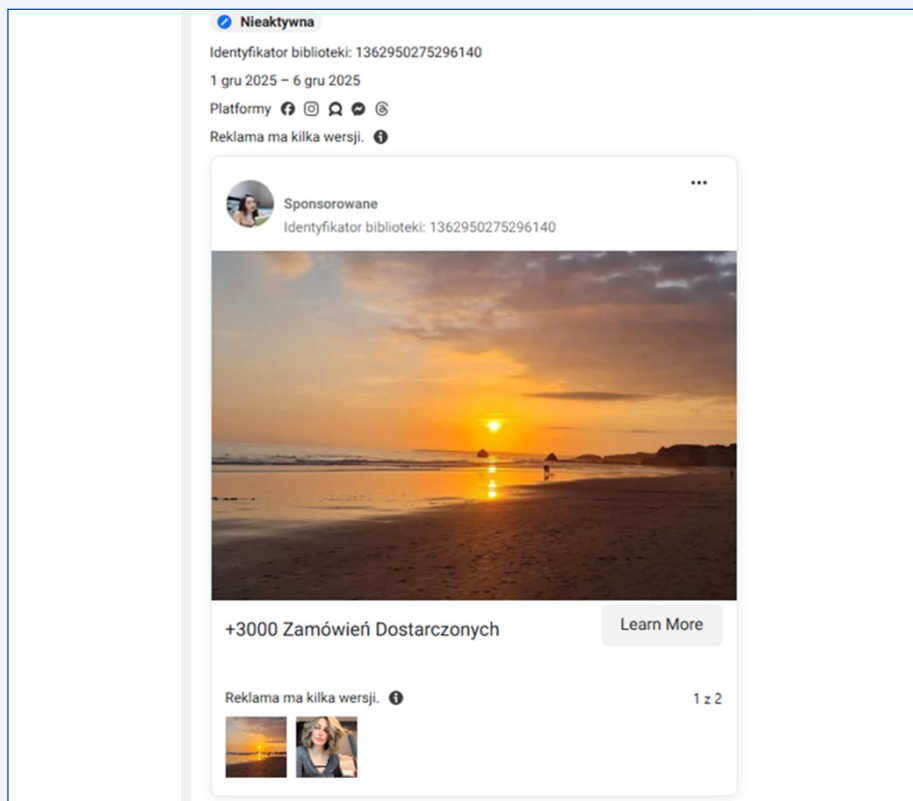


OBRAZ 130.

Przykład reklamodawcy wielokrotnie naruszającego regulamin platformy (zgłoszenie odrzucone)

ŹRÓDŁO: FACEBOOK.COM

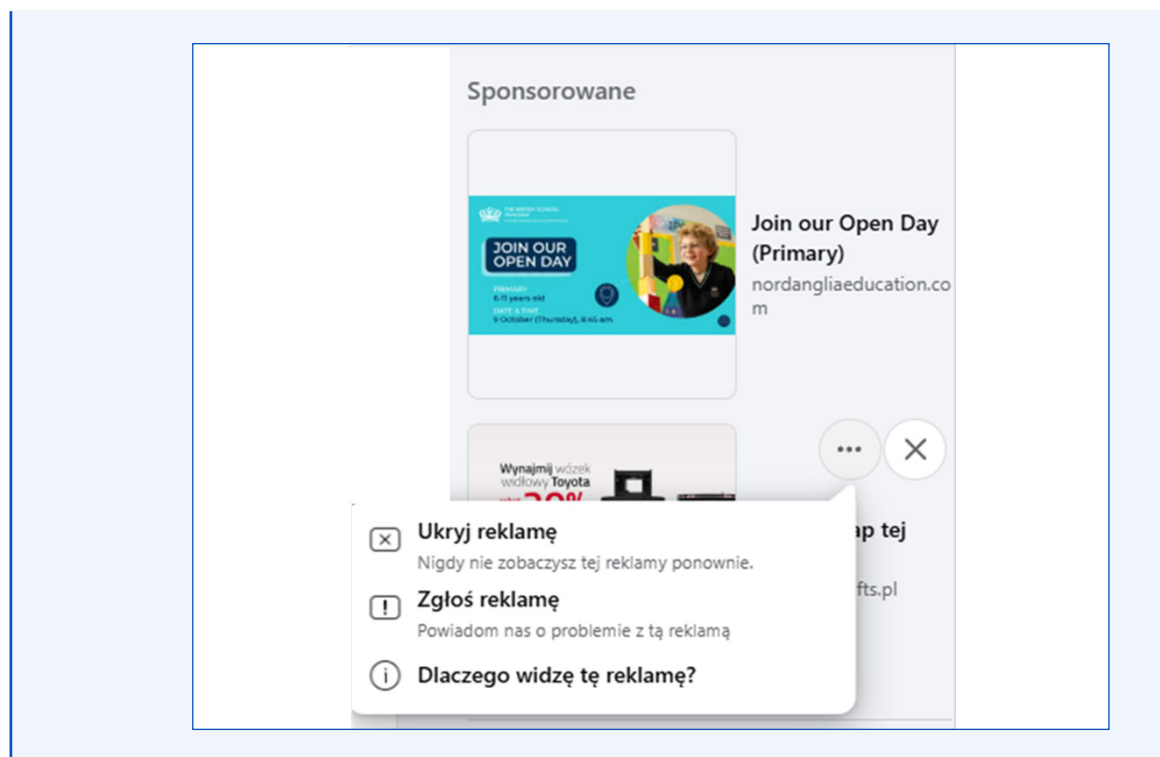
- I. Często praktyką stosowaną przez aktorów dezinformacji w celu utrudnienia identyfikacji szkodliwych treści jest publikowanie tej samej reklamy w wielu wersjach. Główny wariant odsyła do przekazu niezwiązanego z właściwą treścią reklamy, często przybierając formę neutralnego obrazu, na przykład zachodu słońca lub elementu garderoby. Zdarza się, że w wersjach dostępnych w bibliotece reklam nie występuje oryginalny przekaz, który wcześniej był widoczny na tablicy użytkownika. Zabieg ten nie tylko utrudnia identyfikację szkodliwego materiału, ale również ogranicza możliwość przygotowania dokładnego zgłoszenia do administracji serwisu Meta. W przeszłości proceder ten był zgłaszany do punktu kontaktowego, jednak nie doprowadziło to do ograniczenia takich praktyk na platformach firmy Meta.



OBRAZ 131. Reklama opublikowana w sekcji „Sponsorowane”

ŹRÓDŁO: FACEBOOK.COM

- J. Publikacja reklamy w sekcji „Sponsorowane” na platformie Facebook sprawia, że niemożliwe jest pozyskanie linku do jej treści, a tym samym skuteczne zgłoszenie. W tej sytuacji możliwe jest jedynie przekazanie opisu incydentu za pomocą wiadomości mailowej skierowanej do punktu kontaktowego. Ścieżka ta nie pozwala jednak na śledzenie statusu zgłoszenia.



OBRAZ 132. Reklama opublikowana w sekcji „Sponsorowane”

ŹRÓDŁO: FACEBOOK.COM

- K.** Problem treści deepfake na platformie dotyczy również dezinformacji wyborczej. W analizowanym okresie pojawiły się generowane za pomocą AI filmy z kandydatami, które miały ich ośmieszyć i zdyskredytować. Tworzono także nieprawdziwe obrazy rzekomych wyborców danego kandydata, mające sugerować popieranie go przez określoną grupę społeczną. Tego rodzaju materiały nie były usuwane ani moderowane przez platformę, pomimo potencjalnie dużego wpływu na postrzeganie kandydatów przez wyborców (zarówno w fałszywie pozytywnym, jak i negatywnym kontekście).
- L.** Brak szybkiej i adekwatnej reakcji w przypadku kryzysowych sytuacji. Mimo że w takich przypadkach jak incydent z dronami materiały przekazywane są bezpośrednio do punktu kontaktowego z prośbą o ich priorytetową obsługę, to często pozostają one bez reakcji ze strony administracji serwisu. Z kolei w przypadku przekazywania zgłoszeń poprzez dedykowaną platformę administracja ma 48 godzin na rozpatrzenie zgłoszenia. Sprawia to, że platforma podejmuje działania w sytuacjach kryzysowych ze znacznym opóźnieniem, co powoduje, że materiały dezinformacyjne są swobodnie rozpowszechniane w infosferze. Pomimo zaangażowania punktu kontaktowego w czasie sytuacji kryzysowej i deklaracji o priorytetowym podejściu, zgłoszenia były rozpatrywane, ale w większości przypadków zostały odrzucone. Administracja deklarowała jednak, że ograniczała widoczność wskazanych materiałów. Dla OAD NASK nie jest to w żaden sposób mierzalne.

- M.** Administracja podjęła również ograniczone działania w przypadku zidentyfikowanej przez zespół OAD NASK siatki botów rozpowszechniającej dezinformację na temat dronów. Facebook jedynie w kilku przypadkach oznaczył posty jako „fałszywe informacje”, jednak konta, które je opublikowały, nie zostały zablokowane i nadal funkcjonują w serwisie.
- N.** Administracja nie dopatrzyła się również naruszenia zasad społeczności w treściach publikowanych przez Ambasadę Izraela w Polsce na temat sytuacji w Strefie Gazy. Uznała jedynie, że kwalifikują się one do przekazania do zewnętrznej weryfikacji przez zespół fact-checkingowy. Mimo takiego stanowiska platformy treści te są cały czas dostępne i nie towarzyszy im community note czy komunikat uzupełniający kontekst o artykuł fact-checkingowy.

X.com

Wśród problemów i wyzwań, jakie analitycy OAD NASK zauważają w kontekście współpracy z serwisem X, należy wymienić:

- A.** Punkt kontaktowy nie zajmuje się bezpośrednią moderacją treści, a jedynie przekazuje zgłoszone materiały do moderacji.
- B.** Bezpośrednio do moderacji trafiają wyłącznie wnioski przesłane za pośrednictwem formularza „Legal request submissions”. Narzędzie to nie umożliwia jednak podglądu złożonych zgłoszeń ani sprawdzenia ich bieżącego statusu. Zespół OAD NASK otrzymuje jedynie informację zwrotną o tym, czy zgłoszona treść została poddana moderacji, czy nie. W treści wiadomości przesłanej na skrzynkę mailową osoby zgłaszającej zazwyczaj brakuje uzasadnienia decyzji dotyczącej pozostawienia szkodliwego materiału na platformie, co znacząco utrudnia proces odwoławczy. Taki system zgłaszania treści jest mało efektywny i znacząco wydłuża czas obsługi zgłoszeń.
- C.** Punkt kontaktowy zalecił, by treści fałszywe, wprowadzające w błąd czy też stanowiące pomówienia oznaczać poprzez wypełnianie community notes na platformie X. Ta opcja dostępna jest jednak wyłącznie dla użytkowników, którzy zweryfikowali swój numer telefonu. Ze względu na specyfikę pracy zespołu (szybkie reagowanie), ta opcja nie sprawdza się w codziennym funkcjonowaniu. Ponadto przekierowuje odpowiedzialność za moderowanie treści na zewnętrzne instytucje lub użytkowników, choć to platforma powinna podejmować proaktywne działania.
- D.** Platforma X na przestrzeni III kwartału 2025 roku podjęła decyzję o informowaniu użytkowników (również kont znanych z wcześniejszego szerzenia dezinformacji),

których dotyczyły zgłoszenia OAD NASK w kwestii usunięcia lub moderacji treści, że w systemie pojawił się wniosek dotyczący ich wpisu. Dotyczyło to zgłoszeń realizowanych przez formularz „Legal request submissions”.

- E.** W formularzu zgłoszeniowym na platformie X brakuje opcji wyboru kategorii „Dezinformacja” czy podobnej. Stanowi to szczególnie istotny problem, gdyż w praktyce dezinformacja nie jest rozpatrywana jako naruszenie regulaminu X.
- F.** Brak reakcji zespołu moderującego na wpisy dezinformacyjne, które mogły negatywnie wpłynąć na przebieg demokratycznych wyborów.
- G.** Przykładem szczególnie szkodliwej treści, która nie spotkała się z reakcją platformy X, jest wpis opublikowany w czasie pierwszej tury wyborów prezydenckich. Rozpowszechniano wtedy informację, że skopiowanie zaświadczenia o prawie do głosowania, pozwoli na oddanie głosu w wielu lokalach.



OBRAZ 133. Wpis zachęcający do oddania głosu w wielu lokalach wyborczych

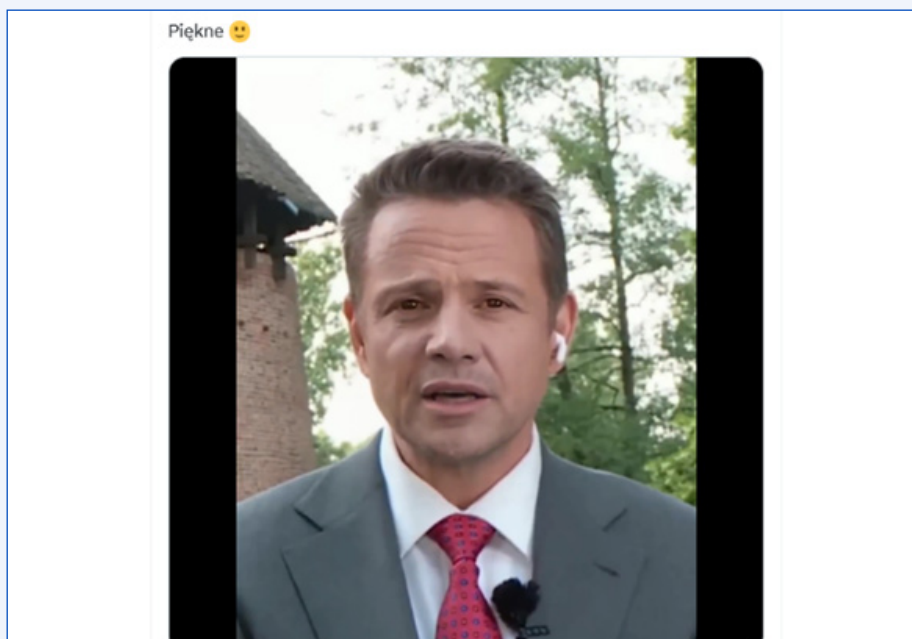
ŹRÓDŁO: X.COM

- H.** Słaba efektywność platformy X w usuwaniu dezinformacyjnych treści może wynikać z faktu, że w zespole moderacji serwisu znajduje się tylko jedna osoba biegle posługująca się językiem polskim⁴³.
- I.** Podejście do moderacji treści przez platformę X jest niekonsekwentne. Bardzo często zdarza się, że moderatorzy platformy X, mając do czynienia z podobnymi treściami dezinformacyjnymi, decydują się na usunięcie jedynie części z nich.
- J.** Kolejnym wyzwaniem jest zgłaszanie treści wygenerowanych przez sztuczną inteligencję. Jak informuje platforma: „Nie wolno udostępniać w serwisie X treści nieprawdziwych, które mogą wprowadzać ludzi w błąd lub prowadzić do szkód. Dotyczy to między innymi: Media syntetyczne i zmanipulowane: Nie wolno udostępniać mediów nieautentycznych, w tym zmanipulowanych lub wyrwanych z kontekstu, które mogą powodować powszechne zamieszanie w kwestiach publicznych, wpływać na bezpieczeństwo publiczne lub powodować

⁴³ <https://transparency.X.com/dsa-transparency-report-2025-october.html> (dostęp: 17.02.2026, godz. 14:38)

poważne szkody („media wprowadzające w błąd”). Możemy korzystać z własnej technologii lub otrzymywać raporty za pośrednictwem partnerstw z podmiotami zewnętrznymi w celu ustalenia czy media zostały zmanipulowane lub wyrwane z kontekstu. W sytuacjach, gdy nie jesteśmy w stanie wiarygodnie określić, czy są to media wprowadzające w błąd, możemy nie podejmować żadnych działań”⁴⁴.

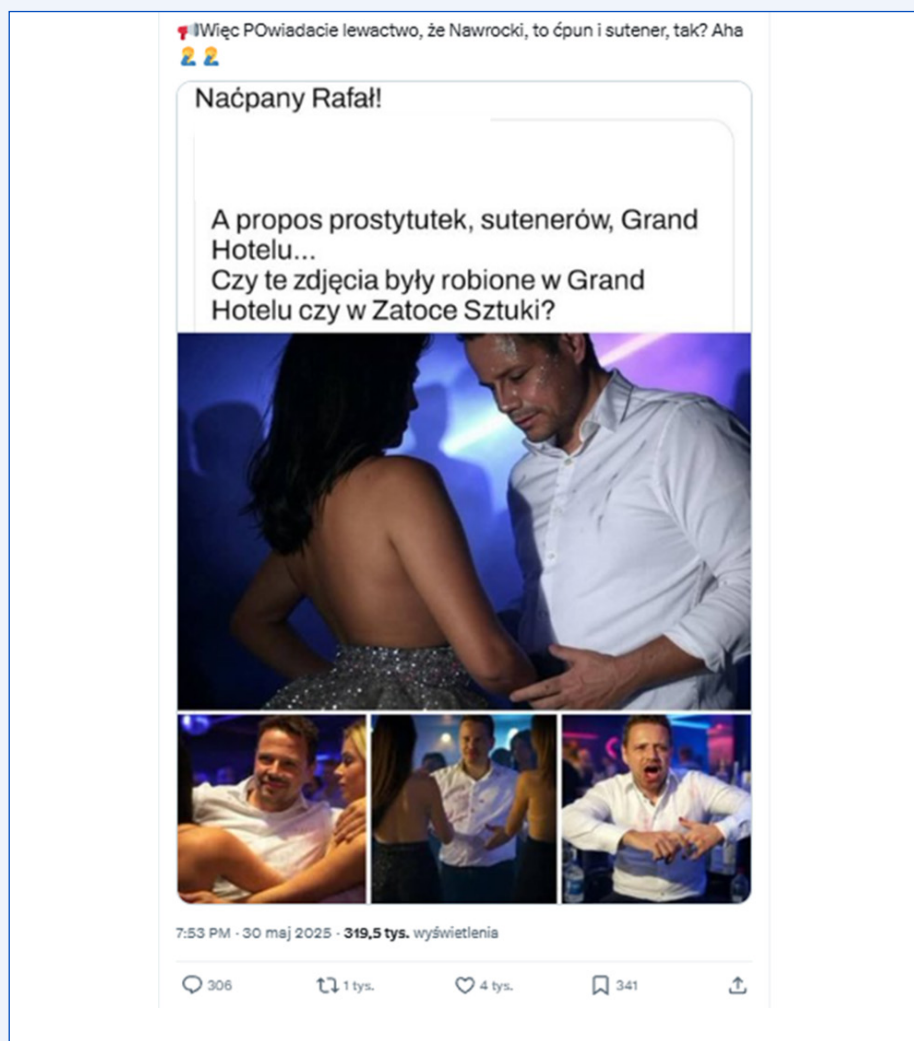
- K.** Mimo że przesyłane do zespołu moderującego treści spełniają wymienione kryteria, to jest: są materiałami nieoznaczonymi jako wygenerowane przy zastosowaniu sztucznej inteligencji, w przypadku wielu z nich nie zostają podjęte działania moderacyjne.



OBRAZ 134. Materiał typu deepfake z wykorzystaniem wizerunku Rafała Trzaskowskiego (zgłoszenie odrzucone)

ŹRÓDŁO: X.COM

⁴⁴ <https://help.X.com/en/rules-and-policies/authenticity> (dostęp: 29.09.2025 godz. 12.40)



OBRAZ 135. Materiał typu deepfake z wykorzystaniem wizerunku Rafała Trzaskowskiego (zgłoszenie odrzucone)

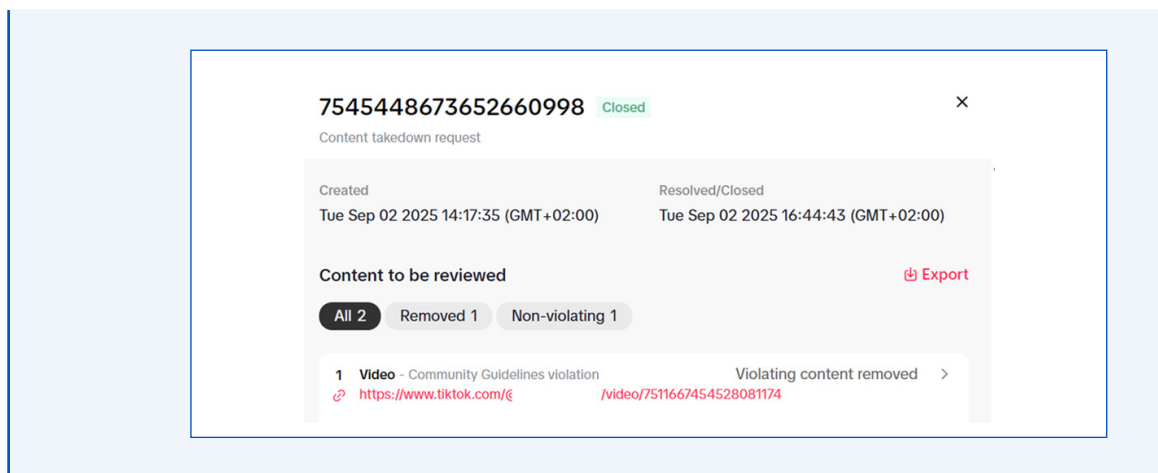
ŹRÓDŁO: X.COM

- L. Zespół OAD NASK zaobserwował, że platforma X sprawnie reagowała na zgłoszenia związane z aktywnością siatek botów. Wynika to najprawdopodobniej z łatwiejszej identyfikacji tego typu kont niż treści dezinformacyjnych, które często wymagają znajomości języka i kontekstu.

TikTok

Wśród problemów i wyzwań, jakie analitycy OAD NASK zauważają w kontekście współpracy z serwisem TikTok, należy wymienić:

- A. Zespół OAD NASK zaobserwował ograniczoną reakcję platformy na treści szerzące szkodliwą dezinformację medyczną, mogące stanowić zagrożenie dla życia i zdrowia obywateli i obywaterek. Szerzej temat ten został opisany w rozdziale Case studies.
- B. OAD NASK odnotowuje przypadki świadczące o niekonsekwentnym podejściu administracji serwisu TikTok do zgłaszanych incydentów. Zebrane dane wskazują na to, że problem ten jest szczególnie widoczny w sytuacjach dotyczących podszywania się pod osoby publiczne. Część kont wykorzystujących wizerunek takich osób została zablokowana przez zespół moderujący treści na platformie. Zauważono jednak, że w serwisie nadal funkcjonują profile podszywające się między innymi pod Premiera Donalda Tuska, pomimo ich wcześniejszego zgłoszenia do administracji platformy.
- C. System raportowania udostępniony przez platformę TikTok nie umożliwia zaawansowanego filtrowania ani sortowania zgłoszeń. Analitycy nie mogą wyszukiwać zgłoszonych treści według autora lub linku do publikacji. W konsekwencji identyfikacja konkretnych materiałów i sprawdzanie ich statusów jest czasochłonne.
- D. Zdarza się, że komunikat wyświetlany w narzędziu zgłoszeniowym wprowadza osobę zgłaszającą incydent w błąd. W takich sytuacjach brak możliwości odwołania się od decyzji zespołu moderującego uniemożliwia szybkie wyjaśnienie sprawy. Poniżej przedstawiono zrzut ekranu z systemu służącego do raportowania naruszeń. Mimo że TikTok poinformował o usunięciu



OBRAZ 136.

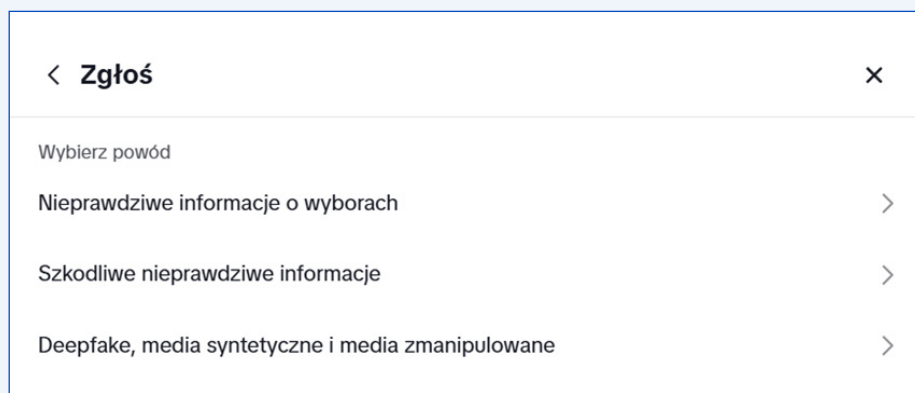
Wprowadzający w błąd komunikat wyświetlany w narzędziu zgłoszeniowym

ŹRÓDŁO: PLATFORMA ZGŁOSZENIOWA TIKTOK

- E.** Na spotkaniu związanym z dezinformacją wyborczą przedstawiciele platformy TikTok poinformowali pracowników OAD NASK, że przy niektórych zgłoszeniach treści dezinformacyjnych administracja nie podejmuje decyzji o usunięciu danego materiału od razu, ponieważ odbiera jej to możliwość śledzenia zachowania kont. Innymi słowy, administracja koncentruje się na wykrywaniu siatek nieautentycznych kont (CIB). Jeśli chodzi o konta oznaczone jako „parody account”, są one blokowane na podstawie analizy zachowania użytkownika (na przykład po ocenie częstotliwości publikowania), a nie bezpośrednio publikowanej przez niego treści. Takie podejście sprawia, że szkodliwe materiały mogą dłużej oddziaływać na użytkowników i szeroko się rozprzestrzeniać.
- F.** Platforma TikTok nie posiada skutecznych mechanizmów wykrywania i blokowania treści typu deepfake, co prowadzi między innymi do pojawiania się zmanipulowanych materiałów wykorzystujących wizerunek polityków oraz innych znanych osób.
- G.** Z poziomu użytkownika nie jest możliwe zgłaszanie wszystkich rodzajów incydentów dezinformacji. Kategoria „Błędne informacje” zawiera poniższy podział:
- **Nieprawdziwe informacje o wyborach, to jest:**
 - A. Nieprawdziwe informacje dotyczące sposobu głosowania lub kandydowania na urząd;
 - B. Nieprawdziwe informacje na temat ostatecznych wyników lub wyników wyborów.
 - **Szkodliwe nieprawdziwe informacje, to jest:**
 - A. Nieprawdziwe informacje, które stanowią zagrożenie dla bezpieczeństwa publicznego lub mogą wywołać panikę, takie jak wykorzystywanie nagrań z przeszłego wydarzenia i fałszywe przedstawianie ich jako aktualnych lub rozpowszechnianie nieprawdziwych twierdzeń, że podstawowe artykuły, takie jak żywność lub woda, nie są już dostępne;
 - B. Nieprawdziwe informacje medyczne, które stanowią zagrożenie dla zdrowia publicznego, takie jak wprowadzające w błąd informacje o szczepionkach i niedokładne porady medyczne, które zniechęcają ludzi do uzyskania odpowiedniej opieki medycznej;
 - C. Nieprawdziwe informacje na temat zmian klimatu, które są sprzeczne z dobrze ugruntowanym konsensusem naukowym, takie jak zaprzeczanie istnieniu zmian klimatu;
 - D. Niebezpieczne teorie spiskowe, które promują przemoc, nienawiść lub wymierzone są w osoby, na przykład te, które powodują uprzedzenia wobec określonej grupy i wyrządzają krzywdę.

- Deepfake, media syntetyczne i media zmanipulowane, to jest:
 - A. Materiały syntetyczne lub zmanipulowane przedstawiające realistyczne sceny, które nie są wyraźnie ujawnione ani oznaczone w filmie
 - B. Materiały syntetyczne, które zawierają podobiznę (wizualną lub dźwiękową) prawdziwej osoby, gdy są używane do rekomendacji politycznych lub komercyjnych lub jeśli naruszają nasze zasady społeczności;
 - C. Materiał zredagowany w sposób, który może wprowadzić w błąd co do rzeczywistych wydarzeń;
 - D. Dozwolone są media syntetyczne przedstawiające osobę publiczną w kontekście artystycznym i edukacyjnym, na przykład celebrytę wykonującego popularny taniec lub postać historyczną występującą na lekcji historii.

Obecna klasyfikacja wyklucza więc możliwość zgłaszania fałszywych informacji dotyczących kwestii międzynarodowych, które mogą wpływać na dyskurs polityczny w Polsce, czy zmanipulowanych treści wykorzystywanych do promowania stereotypów dotyczących mniejszości narodowych czy określonych grup społecznych.



OBRAZ 137.

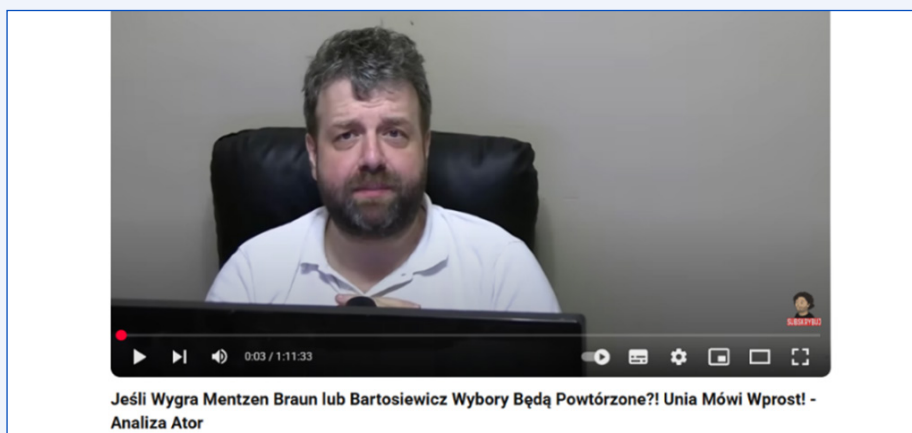
Podkategorie dostępne w ramach kategorii „Błędne informacje” na platformie TikTok.com

ŹRÓDŁO: TIKTOK.COM

Google (YouTube i Google Ads)

Wśród problemów i wyzwań, jakie analitycy OAD NASK zauważają w kontekście współpracy z serwisami firmy Google, należy wymienić:

- A. Mimo poczynionych z administracją ustaleń, że głównym sposobem zgłaszania treści przez OAD NASK będzie bezpośrednie wysyłanie ich do punktu kontaktowego, działania w zakresie usuwania treści są ograniczone, szczególnie w przypadku materiałów z serwisu YouTube. Dotyczy to również kanałów szerzących dezinformację na temat wyborów prezydenckich.

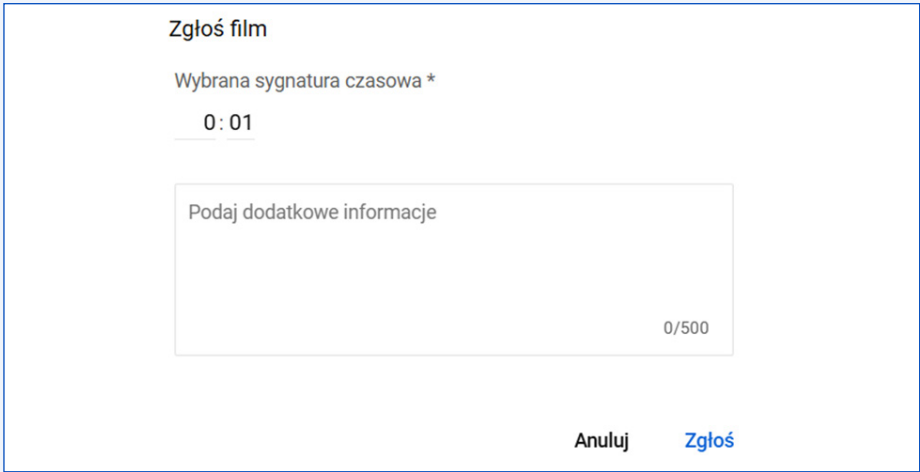


OBRAZ 138. Przykład filmu zawierającego dezinformację wyborczą (zgłoszenie odrzucone)
ŹRÓDŁO: YOUTUBE.COM



OBRAZ 139. Przykład filmu zawierającego dezinformację wyborczą (zgłoszenie odrzucone)
ŹRÓDŁO: YOUTUBE.COM

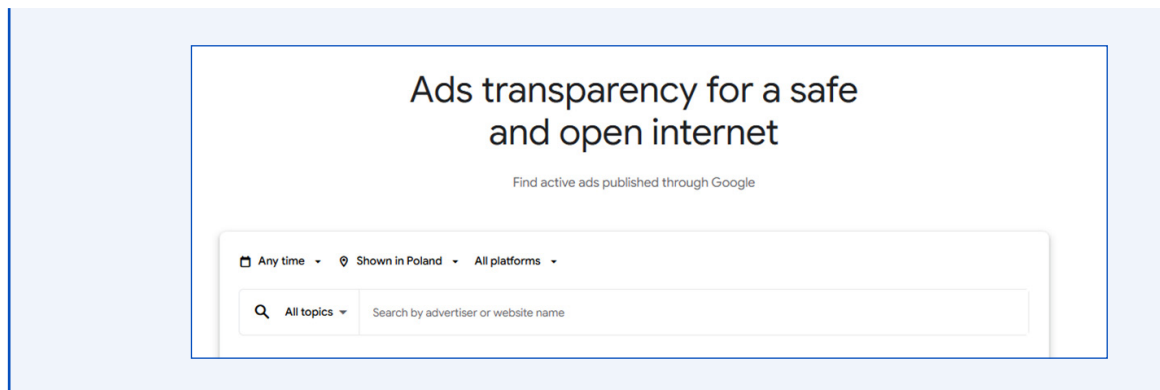
- B.** W związku z ograniczeniem możliwości zgłaszania szkodliwych treści za pomocą formularza kontaktowego jedynie do serwisu Google Play, utrudnione jest monitorowanie aktualnego statusu zgłoszeń. Osoba wyznaczona przez Google do kontaktu pełni wyłącznie rolę pośrednika, przekazując informacje o zgłoszeniach do odpowiednich zespołów moderujących. W rezultacie analitycy OAD NASK rzadko otrzymują informacje zwrotne dotyczące podjętych przez serwisy działań.
- C.** W przypadku zgłaszania treści z poziomu użytkownika na platformie YouTube konieczne jest wskazanie konkretnej sygnatury czasowej, w której dochodzi do naruszenia, co uniemożliwia zgłaszanie kilku incydentów występujących w ramach jednego nagrania.

The image shows a screenshot of a web form titled "Zgłoś film" (Report video). Below the title, there is a label "Wybrana sygnatura czasowa *" (Selected timestamp *). Under this label, the text "0 : 01" is displayed. Below the timestamp, there is a text input field with the placeholder text "Podaj dodatkowe informacje" (Provide additional information). To the right of the input field, the character count "0/500" is shown. At the bottom right of the form, there are two buttons: "Anuluj" (Cancel) and "Zgłoś" (Report).

OBRAZ 140. Pole wyboru sygnatury czasowej na platformie YouTube

ŹRÓDŁO: [YOUTUBE.COM](https://www.youtube.com)

- D.** Ograniczona możliwość przeglądania reklam publikowanych w serwisach należących do platformy. W bibliotece reklam udostępnionej przez serwis nie ma możliwości wyszukiwania treści po słowach kluczowych. Aby uzyskać informacje o publikowanych materiałach, konieczne jest wskazanie nazwy konkretnego reklamodawcy.



OBRAZ 141. Panel wyszukiwania dostępny w bibliotece reklam Google

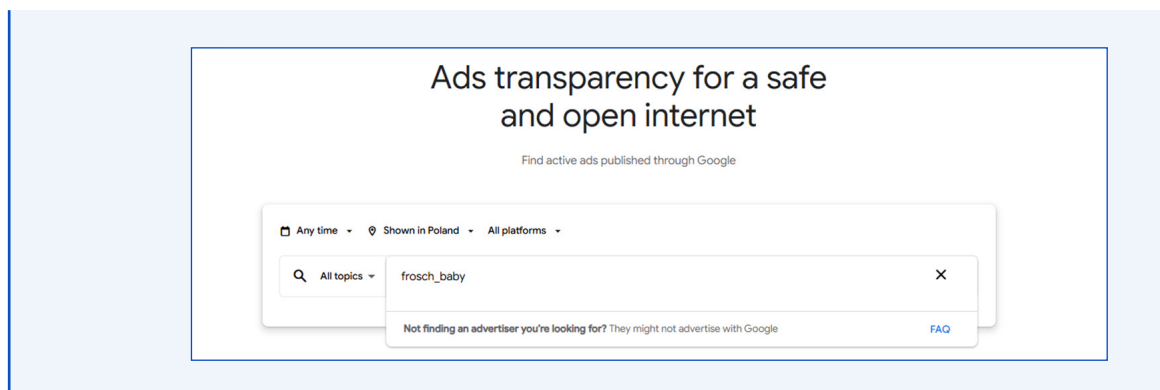
ŹRÓDŁO: [ADSTRANSparency.GOOGLE.COM](https://adstransparency.google.com)

- E.** Mimo że analitykom znane są dane reklamodawcy, zdarza się, że publikowane przez konto materiały sponsorowane nie pojawiają się w bibliotece reklam. Fakt ten uniemożliwia sprawne przekazywanie zgłoszeń do serwisu Google.



OBRAZ 142. Materiał ilustrujący mechanizm działania biblioteki reklam Google

ŹRÓDŁO: [YOUTUBE.COM](https://www.youtube.com)



OBRAZ 143. Wynik wygenerowany po wyszukaniu reklamodawcy frosch_baby w bibliotece reklam

ŹRÓDŁO: [ADSTRANSparency.GOOGLE.COM](https://adstransparency.google.com)

- F. Stosunkowo wysoki poziom reakcji ze strony platformy Google w zakresie moderacji treści zgłoszonych przez zespół OAD NASK odnotowano w przypadku reklam podszywających się pod instytucje państwowe oraz polityków.
- G. Brak systemu zautomatyzowanego i konsekwentnego moderowania treści generatywnych. Zespół OAD NASK zaobserwował brak spójności podejścia platformy Google do usuwania i moderowania treści generowanych przez AI. Tego rodzaju sytuacja miała miejsce między innymi w przypadku zgłoszonych kont w serwisie YouTube publikujących wyłącznie materiały stworzone przez sztuczną inteligencję. W większości prezentowały one sensacyjne i nieprawdziwe informacje dotyczące wydarzeń z polskiej oraz międzynarodowej sceny politycznej.
- H. Zespół OAD NASK obserwuje coraz częstsze pojawianie się na platformie YouTube tego typu kanałów. Mogą być one wykorzystywane nie tylko do rozpowszechniania dezinformacji, ale również do różnego rodzaju oszustw. Niewykluczone, że wiele z nich operuje w ramach zorganizowanej sieci kont, o czym może świadczyć wysoka częstotliwość oraz regularność pojawiania się nowych materiałów. Odpowiedzią na tego typu działania mógłby być zautomatyzowany system oznaczania oraz blokowania materiałów AI, który zapobiegłby masowemu rozprzestrzenianiu się zmanipulowanych treści.
- I. Google umożliwia reklamodawcom oznaczanie reklam jako należących do konkretnej kategorii⁴⁵. W przypadku zgłoszeń dotyczących reklam publikowanych przez Izrael platforma przeniosła je do kategorii Law & Government (Prawo i administracja). Należy jednak podkreślić, że kategoria ta znajduje się w obrębie kategorii ogólnych i nie jest przez Google kwalifikowana jako Political ads (Reklamy polityczne). Google jako reklamy polityczne kwalifikuje wyłącznie treści związane z wyborami⁴⁶.
- J. Umieszczenie spotów w kategorii ogólnej oznacza, że tych nie obejmują ich rygorystyczne wymagania przejrzystości stosowane w przypadku reklam politycznych, takie jak między innymi konieczność ujawniania informacji na temat sztucznie wygenerowanych lub zmodyfikowanych cyfrowo treści, ani ograniczenia związane z kierowaniem reklam do określonych grup odbiorców. Dodatkowym problemem, przy obsłudze tego typu zgłoszeń są opóźnienia w publikowaniu reklam w bibliotece. Google stosuje okres opóźnienia⁴⁷ liczący 48–72 godzin, co jest problematyczne w szczególności przy reklamach wyświetlanych przez okres krótszy niż 72 godziny. Praktyka ta sprawia, że raportowanie takich treści jest niemożliwe w trakcie, kiedy aktywnie wyświetlają się one użytkownikom, a jedyna reakcja platformy możliwa jest już po fakcie.

⁴⁵ <https://support.google.com/displayvideo/answer/2697826> (dostęp: 27.01.2026, godz. 9:30)

⁴⁶ <https://support.google.com/displayvideo/answer/6014595> (dostęp: 27.01.2026, godz. 9:30)

⁴⁷ <https://adstransparency.google.com/faq?region=PL> (dostęp: 27.01.2026, godz. 9:30)

LinkedIn

Wśród problemów i wyzwań, jakie analitycy OAD NASK zauważają w kontekście współpracy z serwisem LinkedIn, należy wymienić:

- A. Monitoring platformy LinkedIn prowadzony przez OAD NASK nie wykazał wielu incydentów dezinformacji na platformie. Sieć ta może jednak być wykorzystywana do tego typu działań na szerszą skalę w przyszłości. Choć możliwe jest zgłaszanie materiałów z zaznaczeniem kategorii „dezinformacja”, to z pewnością skuteczniejszym sposobem byłby kontakt bezpośrednio z dedykowanym opiekunem, szczególnie przy wystąpieniu groźniejszych incydentów.
- B. Mimo że platforma umożliwia zgłaszanie incydentów dezinformacyjnych, nie udostępniono pola pozwalającego na dodanie opisu do zgłoszenia. W konsekwencji nie ma możliwości przekazania zespołowi moderującemu informacji o kontekście ani specyfice danego materiału.

OBRAZ 144.

Formularz służący do zgłaszania incydentów dezinformacji na platformie LinkedIn

ŹRÓDŁO: LINKEDIN.COM

Zgłoś publikację ×

Wybierz zasady, które mają tu zastosowanie

Molestowanie Spam lub scam Spam Dezinformacja Mowa nienawiści

Groźby lub przemoc Samookaleczanie Treści drastyczne

Niebezpieczne lub ekstremistyczne organizacje Treści seksualne Fałszywe konto

Wykorzystywanie dzieci Nielegalne produkty i usługi Naruszenie

Zgłaszasz nielegalne treści?

Zgłoś te treści, jeśli są niezgodne z prawem Wielkiej Brytanii, prawem Unii Europejskiej lub prawem państw członkowskich Unii Europejskiej. [Zgłoś nielegalne treści](#)

Szukasz czegoś innego?

Czasem nasi członkowie wyłączają widoczność niektórych treści, zamiast je zgłaszać.

Powiedz nam, co Ci się nie podoba. →

Dalej

OBRAZ 145.

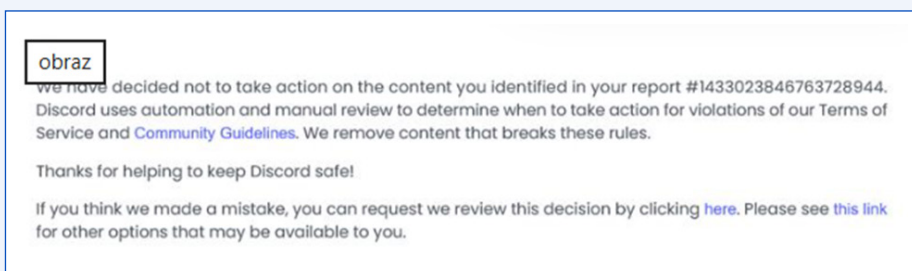
Kategorie naruszeń dostępne w formularzu na platformie LinkedIn

ŹRÓDŁO: LINKEDIN.COM DISCROD

Discrod

Wśród problemów i wyzwań, jakie analitycy OAD NASK zauważają w kontekście współpracy z serwisem Discrod, należy wymienić:

- A. Brak punktu kontaktowego z platformą Discrod stanowi znaczne utrudnienie dla procesu zgłaszania treści dezinformacyjnych. Było to szczególnie widoczne w momencie pojawienia się na platformie w styczniu 2025 roku serwerów, na których prowadzono sprzedaż podrobionej aplikacji mObywatel, co zostało omówione szerzej w rozdziale Case studies.
- B. W momencie wystąpienia incydentu niemożliwe było zgłoszenie platformie całego serwera, a jedynie pojedynczych wiadomości, które się na nim pojawiały, co znacznie wydłużało proces zgłaszania. Nie istniała kategoria związana z dezinformacją czy też inne, pasujące do charakteru tego procederu. Żaden ze zgłaszanych serwerów nie został usunięty, choć incydent ten stanowił poważne zagrożenie, szczególnie w kontekście osób nieletnich.
- C. Obecnie platforma rozszerzyła kategorie zgłoszeń. Pojawiła się między innymi kategoria „dezinformacja” oraz opcja zgłoszeń związanych z naruszeniami Aktu o usługach cyfrowych (DSA).
- D. W systemie dostępnym z poziomu użytkownika po wysłaniu zgłoszenia przez formularz DSA zgłaszający otrzymuje mail z potwierdzeniem oraz numerem porządkowym. Następnie po rozpatrzeniu sprawy platforma wysyła wiadomość ze swoją decyzją. W przypadku decyzji odmownej funkcja odwołania zdaje się nie działać.



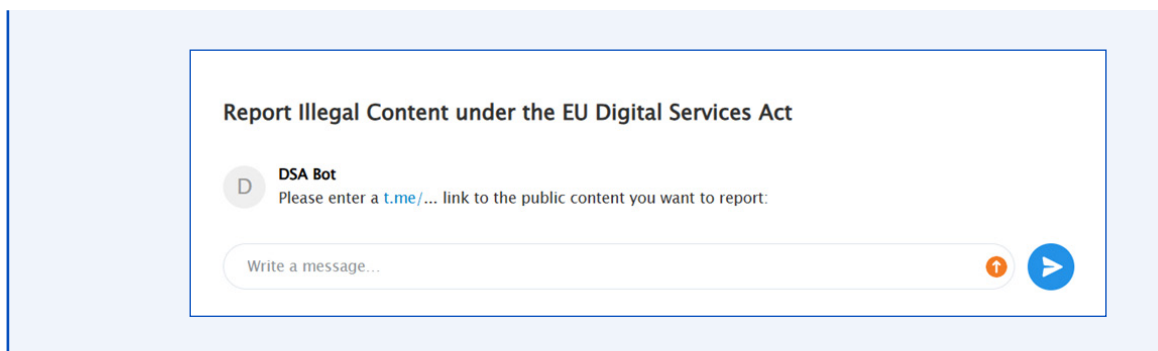
OBRAZ 146. Wiadomość od administracji Discrod w sprawie zgłoszenia

ŹRÓDŁO: POTWIERDZENIE E-MAIL Z DISCROD

Telegram

Wśród problemów i wyzwań, jakie analitycy OAD NASK zauważają w kontekście współpracy z serwisem Telegram, należy wymienić:

- A. W przypadku platformy Telegram można przekazywać informacje o incydentach jedynie za pomocą ogólnodostępnego formularza. Wydaje się, że jego udostępnienie stanowi próbę stworzenia pozoru wprowadzenia procedur wymaganych przez Akt o usługach cyfrowych (DSA). Ze względu na swoją dostępność w krajach Unii Europejskiej Platforma musi realizować obowiązki wynikające ze wspomnianej regulacji, między innymi udostępnić stronę do przyjmowania zgłoszeń, jednak ich rozpatrywanie zawsze kończyło się brakiem reakcji.
- B. Kategorie, w których można dokonać zgłoszenia, są bardzo ograniczone. W przypadku kategorii: „Przemoc”, „Przemoc wobec dzieci”, „Terroryzm”, „Prawa autorskie” oraz „Inne” użytkownik przechodzi do pola, w którym może wpisać szczegółowe informacje dotyczące zgłoszenia. Natomiast w kategoriach „Nie podoba mi się” oraz „To nie jest nielegalne, ale musi zostać usunięte” nie ma możliwości dodania dodatkowych informacji – po wybraniu opcji zgłoszenie zostaje automatycznie przesłane do moderatorów.
- C. Platforma Telegram udostępniła możliwość zgłaszania nielegalnych treści również za pośrednictwem chatbota. Podczas interakcji z chatbotem można jednak zgłosić tylko pojedyncze materiały – brak jest możliwości zgłoszenia całego konta. Do tej pory analitykom OAD NASK nie udało się skutecznie przesłać zgłoszenia za pomocą tej funkcji. Wszystkie dotychczasowe próby wykorzystania chatbota do przekazania informacji o szkodliwym materiale kończyły się wyświetleniem komunikatu o błędzie.



OBRAZ 147.

Usługa chatbota umożliwiająca zgłaszanie treści na platformie Telegram

ŹRÓDŁO: TELEGRAM.COM BLUESKY

Bluesky

Wśród problemów i wyzwań, jakie analitycy OAD NASK zauważają w kontekście współpracy z serwisem Bluesky, należy wymienić:

- A.** Analitycy OAD NASK nie posiadają dostępu do dedykowanego narzędzia, które pozwalałoby na zgłoszenie treści dezinformacyjnych, pojawiających się na platformie Bluesky. Wszystkie zgłoszenia dotyczące tego serwisu są przekazywane z poziomu interfejsu użytkownika. Z tego powodu utrudnione jest śledzenie statusu zgłoszonych incydentów.
- B.** Jak wynika z dotychczasowych doświadczeń analityków, użytkownik zgłaszający treść nie otrzymuje potwierdzenia o przyjęciu treści do oceny.

