

Dr Robert Dziembowski
Katedra Postępowania Karnego i Prawa Karnego Wykonawczego
Wydział Prawa i Administracji
Uniwersytet Warmińsko-Mazurski w Olsztynie
ORCID: 0000-0002-1697-637X
e-mail: robert.dziembowski@uwm.edu.pl

Dr Malwina Treder
Katedra Postępowania Karnego i Prawa Karnego Wykonawczego
Wydział Prawa i Administracji
Uniwersytet Warmińsko-Mazurski w Olsztynie
ORCID: 0000-0002-9489-0096
e-mail: malwina.treder@uwm.edu.pl

MOWA NIENAWIŚCI W SERWISIE SPOŁECZNOŚCIOWYM FACEBOOK – ZAGADNIENIA WYBRANE

Streszczenie

Celem niniejszego artykułu jest omówienie wybranych zagadnień związanych z mową nienawiści na portalu społecznościowym Facebook. W pierwszej kolejności została omówiona krótka geneza tego zjawiska, następnie przytoczone zostały przykładowe definicje mowy nienawiści – zawarta w Rekomendacji Nr R 97/20 Komitetu Ministrów Rady Europy, funkcjonująca w orzecznictwie sądów polskich oraz zamieszczona w Standardach Społeczności Facebooka. Następnie omówione zostały cechy charakterystyczne mowy nienawiści funkcjonującej w Internecie i serwisach społecznościowych. Przeanalizowane zostały także dane opublikowane przez Meta - dotyczące częstotliwości występowania mowy nienawiści na portalu społecznościowym Facebook, ilości działań podejmowanych w stosunku do treści zawierających mowę nienawiści oraz danych procentowych prezentujących działania podjęte w stosunku do przypadków mowy nienawiści przez zgłoszeniem szkodliwych treści przez użytkowników portalu w okresie od III kwartału 2020 r. do I kwartału 2024 r. Następnie omówiona została procedura uzyskiwania danych na temat sprawców czynów zabronionych przez organy ścigania od Meta. W ramach tej części artykułu przeanalizowano także dane dotyczące ilości

wniosków złożonych przez odpowiednie polskie organy oraz pozytywnie rozpatrzonych przez Meta w latach 2017-2023. We wnioskach autorka podsumowała omówione wyżej treści jak również zaproponowała zmiany w polskim systemie prawnym, które mogłyby pozytywnie wpłynąć na ściganie sprawców przestępstw z zakresu mowy nienawiści, a tym samym spełnić funkcję prewencyjną, co potencjalnie mogłoby ograniczyć występowanie tego zjawiska nie tylko na portalach społecznościowych typu Facebook, ale również poza Internetem.

Słowa kluczowe: mowa nienawiści, Facebook, Internet, serwisy społecznościowe, polskie prawo karne.

WSTĘP

Internet jest medium, które jest coraz popularniejsze w naszym codziennym życiu, zaś w szczególności po wybuchu pandemii SARS-CoV-2, kiedy wiele aktywności przeniosło do sieci trudno wyobrazić sobie funkcjonowanie bez niego.

Zgodnie z raportem „Digital 2025: Poland” na początku 2025 roku w Polsce odnotowano 35,75 mln internautów, z mediów społecznościowych korzystało 27,90 mln użytkowników, co stanowiło 68,8% całej populacji. W tym okresie użytkownikami Facebooka było 17,10 mln osób, co oznacza, że 47,83% polskich internautów korzystało z tego serwisu¹. Zgodnie z danymi statystycznymi zawartymi na stronie Similarweb na stan z 1 czerwca 2025 r. najpopularniejszym portalem społecznościowym użytkowanym w Polsce był Facebook, zaś na kolejnych miejscach znalazły się: Instagram oraz X (wcześniej funkcjonujący pod nazwą „Twitter”)².

Facebook posiada wiele możliwości aktywności, w tym zaproszenia do znajomych, „tagowanie” innych, publikowanie komentarzy, zdjęć i tworzenie

¹ Digital 2025: Poland [online:] <https://datareportal.com/reports/digital-2024-poland> [dostęp: 02.02.2025].

² Similarweb: Top Websites Ranking [online:] <https://www.similarweb.com/top-websites/poland/computers-electronics-and-technology/social-networks-and-online-communities>, [dostęp: 01.06.2025].

aktualizacji statusu, przy czym większość funkcji ułatwia interakcję między użytkownikiem, a jego gronem wirtualnych znajomych. Mimo, że uważany jest za prototypowy portal społecznościowy, początków takich witryn należy upatrywać pod koniec lat 70, bowiem w 1979 roku Tom Truscott i Jim Ellis z Duke University stworzyli „Usenet” - ogólnoświatowy system dyskusyjny, który umożliwiał użytkownikom Internetu publikowanie wiadomości publicznych. Era mediów społecznościowych, jaką znamy dzisiaj, rozpoczęła się prawdopodobnie w 1998, kiedy Bruce i Susan Abelson założyli „Open Diary” portal, który zrzeszał autorów dzienników internetowych w jedną społeczność. Coraz powszechniejszy globalnie dostęp do Internetu dodatkowo zwiększył popularność koncepcji tworzenia internetowych społeczności, prowadząc do powstania serwisów społecznościowych, takich jak MySpace (w 2003 r.) i Facebook (w 2004 r.)³.

W związku z tym, że społeczności internetowe odzwierciedlają w sieci zachowania ludzkie z realnego życia nie może dziwić, iż do wirtualnej rzeczywistości przeniknęły również te negatywne obecne w naszym życiu codziennym⁴. Jak wskazywał Rzecznik Praw Obywatelskich w 2019 roku – zarówno dorośli jak i młodzież w Polsce z uwagi na częsty kontakt z mową nienawiści w mediach i sytuacjach życia codziennego sami skłonni są do jej stosowania⁵.

Celem niniejszego artykułu jest omówienie wybranych zagadnień związanych z występowaniem, zwalczaniem oraz przeciwdziałaniem mowy nienawiści w portalu społecznościowym Facebook, z uwzględnieniem współpracy z podmiotem zarządzającym przedmiotową platformą. Nie ulega wątpliwości, iż jest to zjawisko powszechne, którego doświadcza na co dzień przeciętny użytkownik Internetu. Z uwagi na fakt, iż obecnie jednym z najpopularniejszych portalem społecznościowym aktualnie zarówno w Polsce

³ A. M. Kaplan, M. Haenlein, Users of the world, unite! The challenges and opportunities of Social Media, „Business Horizons”, nr 53, 2010, s. 60.

⁴ M. Treder [w:] Social media. Analiza prawnokarna i kryminologiczna. Zagadnienia wybrane, A. Opalska, M. Treder, A. Tymieniecka, Szczecin 2016, s. 36.

⁵ A. Jędrzejczyk, 30 przykładów spraw "mowy nienawiści", w których działania prokuratury budzą wątpliwości, [online:] <https://bip.brpo.gov.pl/pl/content/30-przykladow-mowy-nienawisci-w-ktorych-dzialania-prokuratury-budza-watpliwosci-RPO> [dostęp: 02.02.2025].

jak i na świecie jest Facebook, tematyka przedmiotowej pracy zawężona została to ww. platformy. Jako metodę badawczą zastosowano głównie analizę danych statystycznych opublikowanych przez Meta, a dotyczących częstotliwości występowania, mowy nienawiści na portalu społecznościowym Facebook, działań podejmowanych w celu zwalczania zjawiska (w tym również przez samych użytkowników portalu), jak również współpracy z polskimi organami ścigania w tym zakresie. Dodatkowo przeanalizowano definicje mowy nienawiści oraz cechy charakterystyczne w odniesieniu do jej występowania w mediach społecznościowych.

DEFINICJA MOWY NIENAWIŚCI

Mowa nienawiści istniała jeszcze przed upowszechnieniem się Internetu i mediów społecznościowych, jednakże w dobie komunikacji internetowej osiągnęła nową, często „viralową jakość”. W nie ma jednej definicji tego zjawiska, istnieje wiele odmiennych, zaś tworzone są przez przedstawicieli wielu różnych gałęzi nauk – prawnych, politycznych czy językowych⁶.

Literatura przedmiotu wskazuje, iż termin „mowa nienawiści” (ang. hate speech) został stworzony i spopularyzowany pod koniec lat 80. w Stanach Zjednoczonych przez amerykańskich uczonych z zakresu prawa w opracowaniach ukazujących w jaki sposób różne systemy prawne radziły sobie z niektórymi rodzajami szkodliwej mowy rasistowskiej. Pierwszy raz terminem tym posłużyła się Mari Matsuda w swoim artykule z 1989 roku⁷ opisując w jaki sposób system prawny w Stanach Zjednoczonych zawiódł ofiary szkodliwych wypowiedzi rasistowskich, nie zapewniając im odpowiednich środków dochodzenia zadośćuczynienia, zarówno na gruncie postępowania cywilnego, jak i karnego⁸.

Jedną z najczęściej cytowanych i używanych definicji jest ta zawarta w Rekomendacji Nr R (97) 20 Komitetu Ministrów Rady Europy, zgodnie z którą

⁶ J. Vučković, S. Lučić, Hate speech and social media, „Teme”, Vol. XLVII, nr 1, 2023, s. 193.

⁷ M. J. Matsuda, Public Response to Racist Speech: Considering the Victim's Story, „Michigan Law Review”, Vol. 87, No. 8, 1989.

⁸ A. Brown, What is hate speech? Part 1: the myth of hate, „Law and Philosophy”, nr 36, 2017, s. 424.

mowę nienawiści należy rozumieć jako obejmującą wszelkie formy wypowiedzi, które szerzą, podżegają, promują lub usprawiedliwiają nienawiść rasową, ksenofobię, antysemityzm lub inne formy nienawiści oparte na nietolerancji, w tym: nietolerancję wyrażającą się w agresywnym nacjonalizmie i etnocentryzmie, dyskryminację i wrogość wobec mniejszości, migrantów i osób pochodzenia imigranckiego pochodzenie⁹.

Przenosząc się na grunt polskiego orzecznictwa warto przytoczyć pogląd Sądu Najwyższego, zgodnie z którym pod tym pojęciem powszechnie rozumie się wypowiedzi i wizerunki łączące, wyszydzające i poniżające grupy i jednostki z powodów całkowicie lub po części od nich niezależnych, a także przynależności do innych naturalnych grup społecznych, przy czym grupy naturalne to takie których się nie wybiera, udział w jednych determinowany jest biologicznie (płeć, kolor skóry), w innych społecznie, jak przynależność etniczna, religijna, język¹⁰.

Facebook jako portal społecznościowy posługuje się własną definicją mowy nienawiści na potrzeby funkcjonowania tego serwisu wskazując w swoich Standardach Społeczności, iż jest to przemoc lub odczłowieczająca mowa, twierdzenia o niższości, wezwania do wykluczenia lub segregacji w oparciu o cechy chronione lub obelgi. Cechy te obejmują rasę, pochodzenie etniczne, pochodzenie narodowe, przynależność religijną, orientację seksualną, kastę, płeć, tożsamość płciową oraz poważną niepełnosprawność lub chorobę¹¹.

MOWA NIENAWIŚCI W SERWISIE SPOŁECZNOŚCIOWYM FACEBOOK

Jak już wskazano wyżej Facebook jest jednym z najpopularniejszych serwisów społecznościowych, nie tylko w Polsce, ale również na świecie – w kwietniu

⁹ Recommendation No. R (97) 20 of The Committee Of Ministers to Member States On "Hate Speech", Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies) [online:] <https://rm.coe.int/1680505d5b> [dostęp: 02.02.2025].

¹⁰ Wyrok Sądu Najwyższego z dnia 8 lutego 2019 r., sygn. akt IV KK 38/18, LEX nr 2621830.

¹¹ Meta, Transparency Center, Hate Speech [online:] <https://transparency.fb.com/data/community-standards-enforcement/hate-speech/facebook> [dostęp: 02.02.2025]

2025 roku korzystało z niego ponad 3 miliardy aktywnych użytkowników¹².

Warto zauważyć, iż mowa nienawiści, którą można zaobserwować w Internecie różni się od tej spotykanej poza siecią, zaś jej cechami charakterystycznymi według A. Browna¹³ są:

1. Anonimowość – internauci mogą ukrywać swoją prawdziwą tożsamość, jeśli nie chcą zostać zidentyfikowani, to z kolei może być czynnikiem, który ośmiela ich do publikowania wypowiedzi bardziej bulwersujących, odpychających bądź nienawistnych, aniżeli miałyby to miejsce w konwersacji „twarzą w twarz”.
2. Niewidzialność – między nadawcą a odbiorcami konkretnych komunikatów istnieje najczęściej fizyczna odległość, w związku z tym sprawcy nie odbierają klasycznych społeczno-psychologicznych sygnałów np. potępienia danego zachowania, co w interakcji na żywo skutkowałoby kontrolowaniem szkodliwych lub aspołecznych zachowań.
3. Społeczność – człowiek posiada wrodzone pragnienie, by nawiązywać kontakt z innymi osobami posiadającymi takie same poglądy, zaś Internet daje możliwość wirtualnego łączenia się w grupy i tworzenia społeczności pomimo odległych miejsc zamieszkania. Ponadto nowi lub przyszli członkowie takich grup mogą angażować się w rozpowszechnianie mowy nienawiści w Internecie, jeśli czują się odizolowani lub pragną dobrej opinii innych osób danej społeczności realizując swoją potrzebę akceptacji innych osób.
4. Natychmiastowość - w porównaniu ze sposobami komunikacji offline, Internet sprzyja wypowiedziom będących mową nienawiści, które są spontaniczne, natychmiastowe, instynktowne, nierzadko są

¹²Datereportal, Global Social Media Statistics [online:] <https://datereportal.com/social-media-users> [dostęp: 30.05.2025].

¹³Alexander Brown – profesor nadzwyczajny na University of East Anglia, autor monografii: “Hate Speech Law: A Philosophical Examination” (Routledge, 2015); “The Politics of Hate Speech Laws” (we współautorstwie z Adrianą Sinclair) (Routledge, 2020); “Hate Speech Frontiers: Exploring the Limits of the Ordinary and Legal Concepts” (we współautorstwie z Adrianą Sinclair) (Cambridge University Press, 2023).

nieprzemyślanymi osądami i pierwszymi przemyśleniami na dany temat, bowiem sama konstrukcja funkcjonowania mediów społecznościowych zachęca do szybkich interakcji.

5. Szkoda – mowa nienawiści w mediach społecznościowych może mieć dalej idące skutki z uwagi, z uwagi na fakt, iż zarówno po stronie sprawców jak i ofiar potencjalnie można zanotować większą liczbę osób, więc działania takie będą mieć dużo większą skalę aniżeli czyny dokonywane poza Internetem, a więc co za tym idzie – wywoływać również większą szkodę¹⁴.

Sprawcy mowy nienawiści na portalach społecznościowych działają samodzielnie bądź w ramach zorganizowanej grupy. Podobnie jest z ofiarami takich działań – czyny skierowane są albo przeciwko konkretnym osobom albo grupom dyskryminowanych ze względu na określone cechy, takie jak rasa czy płeć. W drugim przypadku w proceder jest zaangażowana duża grupa osób, zaś nienawiść może skutkować się przemocą fizyczną lub agresywnymi działaniami poza Internetem¹⁵.

Meta¹⁶ publikuje na swojej stronie internetowej dane dotyczące częstotliwości występowania mowy nienawiści w serwisie jako jedną z „treści naruszającą zasady”. Uwzględniają one liczbę wyświetleń danych elementów, zaś wartości procentowe to szacunkowy odsetek tych wyświetleń, które zawierały mowę nienawiści. Podmiot ten szacuje więc jak często dane materiały są wyświetlane, nie zaś ilość publikowanych postów, zdjęć, wideo czy komentarzy, by ustalić określić, jak bardzo wpływają one na użytkowników Facebooka. Tego rodzaju pomiary wierniej odzwierciedlają wpływ na społeczność, bowiem jeden materiał zawierający mowę nienawiści może być wyświetlony przez większą liczbę użytkowników¹⁷. Obecnie na stronie

¹⁴ A. Brown, What is so special about online (as compared to offline) hate speech?, „Ethnicities”, Vol. 18(3), 2018, s. 298-308.

¹⁵ F. Del Vigna., A. Cimino, F. Dell’Orletta, M. Petrocchi, M. Tesconi, Hate me, hate me not: Hate speech detection on Facebook [w:] “Proceedings of the First Italian Conference on Cybersecurity (ITASEC17)” Venice 2017, s. 86.

¹⁶ Meta Platforms, Inc. - amerykański konglomerat, w skład którego wchodzi m.in. Facebook, Instagram i WhatsApp.

¹⁷ Meta, Transparency Center, Prevalence, <https://transparency.fb.com/policies/improving/prevalence-metric> [dostęp: 30.12.2024].

internetowej Meta znajdują się dane dotyczące częstotliwości występowania mowy nienawiści na Facebooku za okres od III kwartału 2020 roku do I kwartału 2024 roku, które zawarte zostały w Tabeli 1 przedstawione graficznie na Wykresie 1.

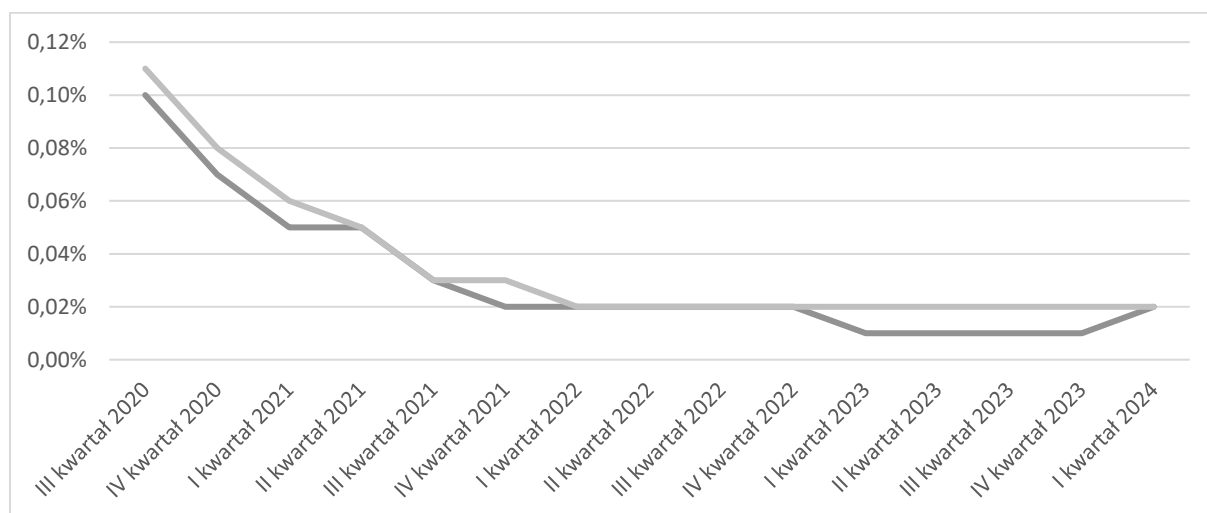
Tabela 1. Częstotliwość występowania mowy nienawiści na portalu społecznościowym Facebook w okresie od III kwartału 2020 do I kwartału 2024.

Kwartał	Wartości maksymalne	Wartości minimalne
III kwartał 2020	0,10%	0,11%
IV kwartał 2020	0,07%	0,08%
I kwartał 2021	0,05%	0,06%
II kwartał 2021	0,05%	0,05%
III kwartał 2021	0,03%	0,03%
IV kwartał 2021	0,02%	0,03%
I kwartał 2022	0,02%	0,02%
II kwartał 2022	0,02%	0,02%
III kwartał 2022	0,02%	0,02%
IV kwartał 2022	0,02%	0,02%
I kwartał 2023	0,01%	0,02%

II kwartał 2023	0,01%	0,02%
III kwartał 2023	0,01%	0,02%
IV kwartał 2023	0,01%	0,02%
I kwartał 2024	0,02%	0,02%

Źródło: Dane opublikowane przez Meta (<https://transparency.fb.com/reports/community-standards-enforcement/hate-speech/facebook/#content-actioned>)

Wykres 1. Częstotliwość występowania mowy nienawiści na portalu społecznościowym Facebook w okresie od III kwartału 2020 r. do I kwartału 2024 r.



Źródło: Dane opublikowane przez Meta (<https://transparency.fb.com/reports/community-standards-enforcement/hate-speech/facebook/#content-actioned>).

Powyższe dane wskazują na to, iż we wskazanym okresie częstotliwość wyświetlania treści naruszających standardy społeczności w zakresie mowy nienawiści na portalu społecznościowym Facebook oscylowały między 0,11% do 0,01%, przy czym wartości te z kwartału na kwartał systematycznie spadają, jedynie w I kwartale 2024 roku można było zanotować niewielki wzrost.

Meta udostępnia także dane dotyczące ilości działań podjętych przez pracowników serwisu społecznościowego, zgodnych z regulaminem, zarówno

w stosunku do samych treści jak i kont, jeszcze przed otrzymaniem zgłoszeń od użytkowników. Podjęcie działań może obejmować usunięcie treści z serwisu jak również wyłączenie czy ograniczenie funkcjonalności kont publikujących szkodliwe materiały¹⁸. W Tabeli 2 i Wykresie 2 przedstawiono dane dotyczące tego rodzaju działań za okres od III kwartału 2020 roku do I kwartału 2023 roku.

Tabela 2. Działania podejmowane w stosunku do treści zawierających mowę nienawiści na portalu społecznościowym Facebook w okresie od III kwartału 2020 r. do I kwartału 2023 r.

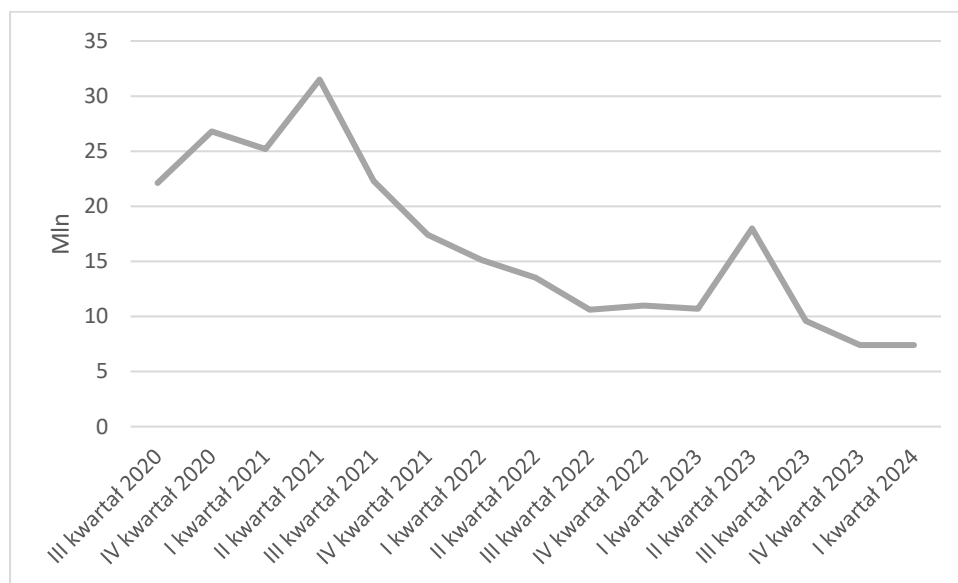
Kwartał	Podjęte działania (mln)
III kwartał 2020	22,1
IV kwartał 2020	26,8
I kwartał 2021	25,2
II kwartał 2021	31,5
III kwartał 2021	22,3
IV kwartał 2021	17,4
I kwartał 2022	15,1
II kwartał 2022	13,5
III kwartał 2022	10,6

¹⁸ Meta, Transparency Center, Content actioned [online:] <https://transparency.fb.com/policies/improving/content-actioned-metric> [dostęp: 30.12.2024].

IV kwartał 2022	11
I kwartał 2023	10,7

Źródło: Dane opublikowane przez Meta (<https://transparency.fb.com/reports/community-standards-enforcement/hate-speech/facebook/#content-actioned>).

Wykres 2. Działania podejmowane w stosunku do treści zawierających mowę nienawiści na portalu społecznościowym Facebook w okresie od III kwartału 2020 r. do I kwartału 2024 r.



Źródło: Dane opublikowane przez Meta (<https://transparency.fb.com/reports/community-standards-enforcement/hate-speech/facebook/#content-actioned>)

Z powyższych danych wynika, że działania Meta związane z przeciwdziałaniem rozprzestrzeniania mowy nienawiści w serwisie społecznościowym Facebook w ww. okresie wynosiły między 31,5 mln (w II kwartale 2021 r.), a 7,4 mln (w IV kwartale 2023 i I kwartale 2024). Zauważyć można tendencję spadkową w liczbie podejmowania takich działań od II kwartału 2021 r. do III kwartału 2022 r. Jednakże co istotne - Meta wskazuje, iż liczba interwencji może rosnąć lub spadać z powodu wielu czynników zewnętrznych. Jeśli na przykład zorganizowana grupa zaleje Facebook tym samym postem zwielowrotnionym 10 milionów razy, liczba interwencji w danym kwartale wzrośnie o 10 milionów. Nie są więc to dane, które obrazowałyby skuteczność zwalczania mowy nienawiści w tym serwisie

społecznościowym, a wskazujące jedynie na ilość szkodliwych treści, które opublikowane zostały na Facebooku¹⁹.

Do publicznej wiadomości opublikowano również informacje, które wskazują na odsetek wszystkich treści lub kont, w odniesieniu do których Facebook samodzielnie podjął działania zanim użytkownicy dokonali stosownego zgłoszenia. Meta dysponuje narzędziami, które automatycznie identyfikują treści potencjalnie naruszające Standardy Społeczności. Mimo, że sama technologia jest obiecująca upływie jeszcze wiele lat, zanim stanowić będzie ona skuteczne narzędzie w przypadku wszelkiego rodzaju naruszeń, istnieją bowiem ograniczenia w możliwości zrozumienia kontekstu i niuansów, szczególnie w przypadku treści tekstowych²⁰. W Tabeli 3 i Wykresie 3 przedstawiono dane dotyczące działań podejmowanych przez Facebooka oraz zgłoszeń użytkowników za okres od III kwartału 2020 roku do I kwartału 2023 roku.

Tabela 3. Mowa nienawiści – działania samodzielnie podejmowane przez Facebooka oraz zgłoszenia użytkowników w okresie od III kwartału 2020 r. do I kwartału 2024 r.

Kwartał	Działania samodzielne Facebooka	Zgłoszenia użytkowników
III kwartał 2020	94,80%	5,20%
IV kwartał 2020	97,10%	2,90%
I kwartał 2021	96,70%	3,30%
II kwartał 2021	97,60%	2,40%
III kwartał 2021	96,50%	3,50%

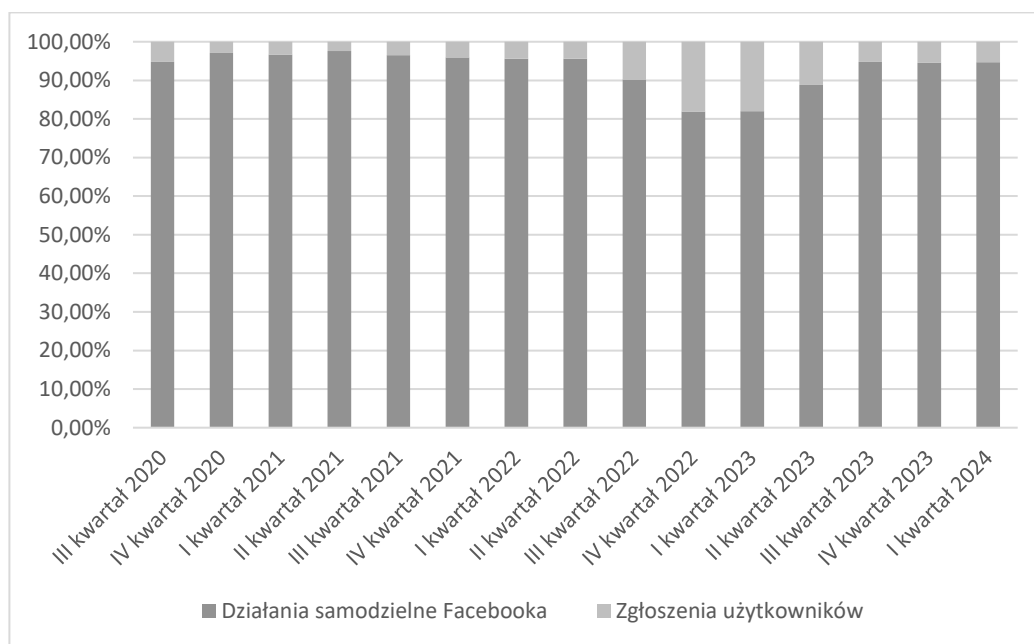
¹⁹ Ibidem.

²⁰ Meta, Transparency Center, Proactive rate [online:] <https://transparency.fb.com/policies/improving/proactive-rate-metric> [dostęp: 30.12.2024]

IV kwartał 2021	95,90%	4,10%
I kwartał 2022	95,60%	4,40%
II kwartał 2022	95,60%	4,40%
III kwartał 2022	90,20%	9,80%
IV kwartał 2022	81,90%	18,10%
I kwartał 2023	82%	18%
II kwartał 2023	88,80%	11,20%
III kwartał 2023	94,80%	5,20%
IV kwartał 2023	94,50%	5,50%
I kwartał 2024	94,70%	5,30%

Źródło: dane opublikowane przez Meta (<https://transparency.fb.com/reports/community-standards-enforcement/hate-speech/facebook/#content-actioned>).

Wykres 3. Mowa nienawiści – działania samodzielnie podejmowane przez Facebooka oraz zgłoszenia użytkowników w okresie od III kwartału 2020 r. do I kwartału 2023 r.



Źródło: dane opublikowane przez Meta (<https://transparency.fb.com/reports/community-standards-enforcement/hate-speech/facebook/#content-actioned>)

Z powyższych danych wynika, iż skuteczność narzędzi do wykrywania mowy nienawiści na Facebooku jest dosyć duża i waha się między 82% (I kwartał 2023), a 97,60% (w II kwartale 2021 roku) i w większości przypadków odpowiednie działania podejmowane są jeszcze przed zgłoszeniami dokonanych przez samych użytkowników. Jednocześnie można zauważyć, iż od II kwartału 2021 roku do IV kwartału 2022 roku coraz większą rolę zaczęły odgrywać zgłoszenia użytkowników – procentowo ich ilość wahała się między 4,40%, a 18,10%.

W kontekście walki z mową nienawiści warto wspomnieć, iż Meta angażuje się również w regionalne inicjatywy międzynarodowe. Komisja Europejska w maju 2016 roku wraz z Facebookiem, Microsoftem, Twitterem i YouTube'em stworzyła „Kodeks postępowania w sprawie zwalczania nielegalnej mowy nienawiści w Internecie”. Do porozumienia tego w 2018 r. dołączył Instagram, Snapchat i Dailymotion, w styczniu 2019 r. Jeuxvideo.com, w 2020 r. TikTok, w LinkedIn w 2021 r., zaś odpowiednio w maju i czerwcu 2022 r. Rakuten Viber i Twitch²¹. W ramach tej inicjatywy Komisja Europejska

²¹ European Commission, The EU Code of Conduct [online:] <https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental->

okresowo publikuje informacje z oceny stosowania przedmiotowego Kodeksu w ramach działalności ww. serwisów społecznościowych. Ostatnie dane – z listopada 2022 roku wskazują, iż zanotowano spadek wyników przedsiębiorstw w zakresie powiadamiania i podejmowania działań tj. liczba zgłoszeń rozpatrzonych przez serwisy społecznościowe w ciągu 24 godzin spadła w porównaniu z dwoma ostatnimi monitoringami z 90,4% w 2020 r. do 81% w 2021 r., i 64,4% w 2022 r. Z kolei wskaźnik usuwania wyniósł 63,6% i był podobny jak w 2021 r. (62,5%), ale nadal niższy niż w 2020 r. (71%), przy czym aktywność Facebooka w tym zakresie pozostała na niemalże takim samym poziomie (69,1% w 2022 r. i 70,2% w 2021 r.)²².

UZYSKIWANIE DANYCH SPRAWCÓW PRZESTĘPSTW OD META

Najbardziej rozpoznawalnymi przestępstwami publicznoskargowymi z zakresu mowy nienawiści są czyny penalizowane przez art. 256 i 257 Kodeksu karnego, czyli przestępstwa publicznego propagowania faszyzmu i nawoływania do nienawiści oraz publicznego znieważenia lub naruszenia nietykalności z powodów narodowościowych, etnicznych, rasowych i wyznaniowych lub bezwyznaniowości²³. Są to przestępstwa publicznoskargowe, co oznacza, iż państwowy organ powołany do ścigania prowadzi postępowanie przygotowawcze i oskarża przed sądem²⁴. W związku z tym obowiązek zebrania odpowiedniego materiału dowodowego spoczywa na organach ścigania. Przydatne mogą więc być dane gromadzone przez serwis społecznościowy, chociażby z tego względu, iż Facebook przechowuje m.in. informacje związane

rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_pl [dostęp: 02.02.2025]

²² European Commission, EU Code of Conduct against online hate speech: latest evaluation shows slowdown in progress [online:] https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7109 [dostęp: 02.02.2025]

²³ M. Urbańczyk, Definiowanie mowy nienawiści w orzecznictwie sądów polskich w sprawach karnych, „Studia nad Autorytaryzmem i Totalitaryzmem”, nr 1, 2020, s. 80.

²⁴ L. Gardocki, Pojęcie przestępstwa i podziały przestępstw w polskim prawie karnym, „Annales Universitatis Mariae Curie-Skłodowska. Sectio G. Ius”, nr 2, 2013, s. 38.

z aktywnością w aplikacji i sieci, lokalizację użytkownika, czy też informacje o urządzeniach z których nastąpiło logowanie w tym o adresie IP²⁵.

Meta co 6 miesięcy publikuje nowe dane statystyczne w ramach „Transparency Report”, gdzie wskazuje m.in. liczbę wniosków wystosowanych przez odpowiednie organy ścigania o ujawnienie danych użytkowników z każdego kraju („Government Requests for User Data”). Zgodnie z udostępnionymi informacjami – podmiot ten odpowiada na żądania udostępnienia danych zgodnie z obowiązującym prawem oraz swoimi warunkami świadczenia usług. Podmiot ten deklaruje, iż otrzymane żądanie jest dokładnie sprawdzane pod kątem zgodności z prawem oraz zastrzega, że wniosek taki może spotkać się z odmową bądź też z prośbą o skonkretyzowanie, jeśli pytania są zbyt ogólne bądź niejasne²⁶.

Podmiot zarządzający Facebookiem rozróżnia dwa rodzaje wniosków składanych przez odpowiednie organy państwowe, definiowane zgodnie z samodzielnie przejętymi kryteriami, niekonieczne znajdujących odzwierciedlenie ze specyfiką postępowań w danych państwach tj.:

- „Legal process requests” - wnioski związane z toczącymi się postępowaniami sądowymi²⁷. Meta ujawnia dane konta wyłącznie zgodnie z Terms of Service i obowiązującym prawem²⁸.

- „Emergency disclosure requests” - w nagłych przypadkach organy ścigania mogą składać wnioski przed rozpoczęciem postępowań. W zależności od okoliczności sprawy Meta deklaruje dobrowolnie ujawnić informacje organom ścigania, jeśli działając „w dobrej wierze” mają powód, by sądzić, że

²⁵ A. Kotowski, Jak sprawdzić, jakie dane zebrał na twój temat Facebook i Google? „Prawda może niektórych przerazić” [online:] <https://www.komputerswiat.pl/artykuly/redakcyjne/jak-sprawdzic-jakie-dane-zebral-na-twoj-temat-facebook-i-google-prawda-moze-yhmd7vq> [dostęp: 02.02.2025]

²⁶ Meta, Transparency Center, Government Requests for User Data [online:] <https://transparency.fb.com/reports/government-data-requests> [dostęp: 30.12.2024]

²⁷ Z uwagi na fakt, iż Facebook na swoich stronach publikuje informacje dotyczące informacji dla przedstawicieli organów ścigania, gdzie mowa jest o wnioskach „w związku z toczącym się śledztwem” ten rodzaj wniosków dotyczy również dowodów pozyskiwanych na etapie postępowania przygotowawczego (zob. Informacje dla przedstawicieli organów ścigania [online:] <https://www.facebook.com/help/494561080557017> [dostęp: 30.12.2024])

²⁸ Meta, Transparency Center, Data types [online:] <https://transparency.fb.com/data/government-data-requests/data-types> [dostęp: 30.12.2024]

sprawa wiąże się z bezpośrednim ryzykiem poważnych obrażeń fizycznych lub śmierci²⁹.

Zdecydowana większość wniosków dotyczy postępowań karnych. W wielu z tych przypadków wnioski obejmują podstawowe informacje o użytkownikach takich jak imię i nazwisko, data rejestracji i okres korzystania z danej usługi. Inne żądania mogą również dotyczyć dzienników adresów IP lub zawartości konta³⁰.

Do publicznej wiadomości zostały także udostępnione informacje dotyczące wniosków złożonych o udzielenie informacji przez odpowiednie organy w poszczególnych państwach oraz ile z nich procentowo w stosunku do wszystkich zostało zrealizowanych. Szczegółowe dane dotyczące wniosków składanych przez polskie organy do Meta w latach 2017-2023 przedstawione zostały w Tabeli 4 i Wykresie 4.

Tabela 4. Wnioski o udostępnienie informacji złożone do Meta przez polskie organy w latach 2017-2023.

Półrocze		Złożone wnioski	Procent zrealizowanych wniosków
2017	I półrocze	1476	53%
2017	II półrocze	1476	53%
2018	I półrocze	1842	54%
2018	II półrocze	1912	52%
2019	I półrocze	3397	56%

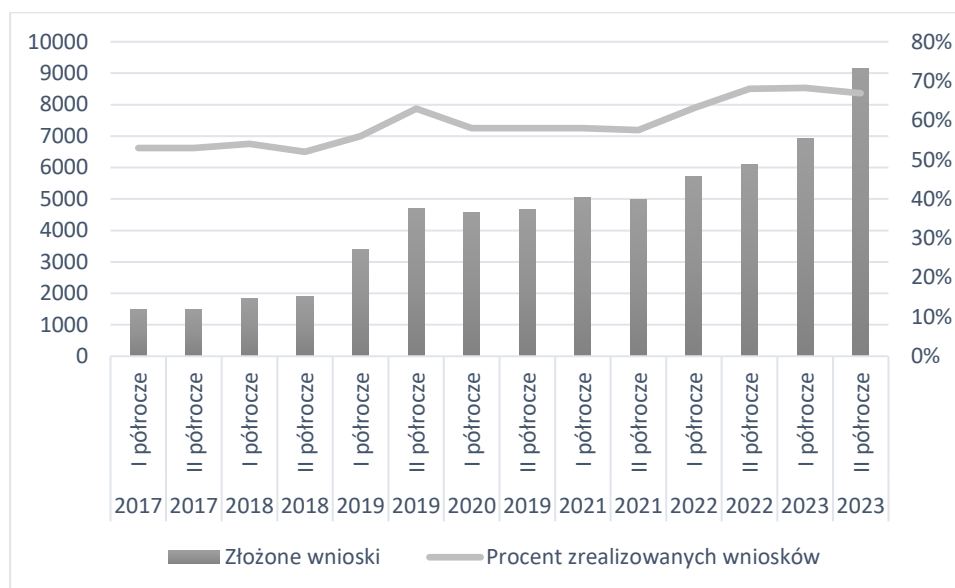
²⁹ Ibidem.

³⁰ Meta, Transparency Center, Further asked questions [online:] <https://transparency.fb.com/data/government-data-requests/further-asked-questions> [dostęp: 30.12.2024].

2019	II półrocze	4688	63%
2020	I półrocze	4572	58%
2019	II półrocze	4673	58%
2021	I półrocze	5033	58%
2021	II półrocze	4982	57,53%
2022	I półrocze	5720	63,11%
2022	II półrocze	6083	68,07%
2023	I półrocze	6910	68,26%
2023	II półrocze	9135	66,91%

Źródło: Dane udostępniane przez Meta (<https://transparency.fb.com/reports/government-data-requests/country/PL/>)

Wykres 4. Wnioski o udostępnienie informacji złożone do Meta przez polskie organy w latach 2017-2023.



Źródło: dane udostępniane przez Meta (<https://transparency.fb.com/reports/government-data-requests/country/PL/>)

Z powyższych danych wynika, iż liczba wniosków składanych do Meta przez polskie organy wzrasta, bowiem w I i II półroczu 2017 roku było ich 1476, zaś w II półroczu 2023 liczba ta wzrosła do 9135, a więc wzrosła sześciokrotnie. W przypadku procentowego oznaczenia pozytywnie rozpoznanych wniosków, co do zasady można zauważyć tendencję wzrostową, przy czym informacje przedstawione w „Transparency Report” udostępniają zbiorcze dane bez podziału na wnioski zrealizowane w całości i w części³¹. Zauważyć można, iż z roku na rok procent wniosków uwzględnionych wzrasta w odniesieniu do polskich organów, bowiem I półroczu 2017 roku było to 53%, zaś w I półroczu 2023 roku było to aż 68,26%, co daje wzrost o 15,26%.

WNIOSKI

Niniejsze opracowanie potwierdza, iż mowa o nienawiści, z którą można spotkać się w jednym z najpopularniejszych serwisie społecznościowym jakim jest Facebook stanowi realny problem. Pozytywnie należy ocenić działania podejmowane przez Meta w celu zwalczania tego zjawiska. Przede wszystkim

³¹ Meta, Transparency Center, Government Requests for User Data, Poland, [online:] <https://transparency.fb.com/reports/government-data-requests/country/PL/> [dostęp: 30.12.2024]

w Standardach Społeczności zdefiniowana została mowa nienawiści, co pozwala na dosyć precyzyjne określenie jakie działania są zakazane i będą zwalczane przez portal społecznościowy. Treści stanowiące mowę nienawiści są usuwane, zaś konto użytkownika, który je zamieścił może zostać ograniczone lub wyłączone³². Facebook dysponuje także narzędziami pozwalającymi na wykrycie i usunięcie szkodliwych materiałów jeszcze przed zgłoszeniem ich przez zaniepokojonych użytkowników, jednakże można zauważyć, iż mimo ich stosunkowo wysokiej skuteczności, od połowy 2021 roku coraz większą rolę zaczynają odgrywać zgłoszenia użytkowników. Mechanizmy te powodują, iż problem mowy nienawiści wprawdzie nie został wyeliminowany, jednakże systematycznie spada częstotliwość wyświetlania tego rodzaju treści.

Zdaje się, iż wobec tak rozbudowanych działań Meta, kolejnymi krokami podjętymi w celu walki z tymi zjawiskami powinny być odpowiednie narzędzia prawne funkcjonujące nie tyle co na szczeblu międzynarodowym co krajowym w poszczególnych państwach. Odpowiednio surowe kary oraz większa skuteczność organów ścigania z całą pewnością byłyby działaniami prewencyjnymi, co z kolei wpłynęłoby na skuteczną walkę z mową nienawiści na portalu społecznościowym Facebook.

Podkreślić należy, iż na chwilę obecną Meta udostępnia dane użytkowników na wniosek odpowiednich organów z poszczególnych państw, jak również współpracuje chociażby z Komisją Europejską, czego efektem było stworzenie „Kodeksu postępowania w sprawie zwalczania nielegalnej mowy nienawiści w Internecie”.

Przechodząc na grunt polskiego prawodawstwa, zaproponować można dwie główne zmiany, które wpłynęłyby na większą skuteczność walki z mową nienawiści w serwisach społecznościowych typu Facebook. Po pierwsze jest to wprowadzenie ustawowej definicji mowy nienawiści, zaś po drugie – ściganie z urzędu czynów motywowanymi uprzedzeniami innymi niż na tle rasowym,

³²Facebook, Centrum pomocy, [online:] https://www.facebook.com/help/adsmanagerbuiltin/260743102021762?helpref=related&locale=pl_PL [dostęp: 30.12.2024].

pochodzenia narodowego lub etnicznego bądź wyznania. Są to zresztą dwie z dwudziestu rekomendacji Rzecznika Praw Obywatelskich dotyczące walki z mową nienawiści³³. W zakresie definicji na chwilę obecną opieramy się bądź na Rekomendacji R (97) 20 Komitetu Ministrów Rady Europy bądź na dorobku orzecznictwa sądów powszechnych, co wydaje się niewystarczające. Z kolei rozszerzenie katalogu przestępstw ściganych z urzędu zmniejszyłoby poczucie bezkarności osób posługujących się mową nienawiści na portalach społecznościowych. Z całą pewnością dobrym krokiem byłoby stworzenie typów kwalifikowanych przestępstw w zakresie mowy nienawiści, w przypadkach jeśli sprawca działałby za pomocą środków masowego komunikowania, a więc zastosowanie konstrukcji analogicznej jak w art. 212§2 czy 216§2 Kodeksu Karnego.

Bibliografia

Akty normatywne

Recommendation No. R (97) 20 of The Committee Of Ministers to Member States On "Hate Speech", Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies).

Orzecznictwo

Wyrok Sądu Najwyższego z dnia 8 lutego 2019 r., sygn. akt IV KK 38/18, LEX nr 2621830.

Literatura

Brown A., What is hate speech? Part 1: the myth of hate, "Law and Philosophy", nr 36, 2017.

Brown A., What is so special about online (as compared to offline) hate speech?, "Ethnicities", Vol. 18(3), 2018.

Del Vigna F., Cimino A., Dell'Orletta F., Petrocchi M., Tesconi M., Hate me, hate me not: Hate speech detection on Facebook [w:] "Proceedings of the First Italian Conference on Cybersecurity (ITASEC17)" Venice 2017.

³³ Ł. Starzewski, Jak walczyć z mową nienawiści. 20 rekomendacji RPO dla premiera [online:] <https://bip.brpo.gov.pl/pl/content/jak-walczy-cz-mowa-nienawisci-20-rekomendacji-rpo-dla-premiera> [dostęp: 01.06.2025]

Gardocki L., Pojęcie przestępstwa i podziały przestępstw w polskim prawie karnym, „Annales Universitatis Mariae Curie-Skłodowska. Sectio G. Ius”, nr 2, 2013.

Kaplan A. M., Haenlein M., Users of the world, unite! The challenges and opportunities of Social Media, „Business Horizons”, nr 53, 2010.

Matsuda M. J., Public Response to Racist Speech: Considering the Victim's Story, „Michigan Law Review”, Vol. 87, No. 8.

Treder M. [w:] Social media. Analiza prawnokarna i kryminologiczna. Zagadnienia wybrane, A. Opalska, M. Treder, A. Tymieniecka, Szczecin 2016.

Urbańczyk M., Definiowanie mowy nienawiści w orzecznictwie sądów polskich w sprawach karnych, „Studia nad Autorytaryzmem i Totalitaryzmem”, nr 1, 2020.

Vučković J., Lučić S., Hate speech and social media, „Teme”, Vol. XLVII, nr 1, 2023.

Źródła internetowe

Jędrzejczyk A., 30 przykładów spraw "mowy nienawiści", w których działania prokuratury budzą wątpliwości, [online:] <https://bip.brpo.gov.pl/pl/content/30-przykladow-mowy-nienawisci-w-ktorych-dzialania-prokuratury-budza-watpliwosci-RPO>.

Kotowski A., Jak sprawdzić, jakie dane zebrał na twój temat Facebook i Google? "Prawda może niektórych przerazić" [online:] <https://www.komputerswiat.pl/artykuly/redakcyjne/jak-sprawdzic-jakie-dane-zebral-na-twoj-temat-facebook-i-google-prawda-moze/yhmd7vq>.

Meta, Transparency Center, Content actioned [online:] <https://transparency.fb.com/policies/improving/content-actioned-metric>.

Meta, Transparency Center, Data types [online:] <https://transparency.fb.com/data/government-data-requests/data-types>.

Meta, Transparency Center, Further asked questions [online:] <https://transparency.fb.com/data/government-data-requests/further-asked-questions>.

Meta, Transparency Center, Government Requests for User Data [online:] <https://transparency.fb.com/reports/government-data-requests>.

Meta, Transparency Center, Government Requests for User Data, Poland [online:] <https://transparency.fb.com/reports/government-data-requests/country/PL>.

Meta, Transparency Center, Hate Speech [online:] <https://transparency.fb.com/data/community-standards-enforcement/hate-speech/facebook>.

Meta, Transparency Center, Prevalence [online:]
<https://transparency.fb.com/policies/improving/prevalence-metric>.

Meta, Transparency Center, Proactive rate [online:]
<https://transparency.fb.com/policies/improving/proactive-rate-metric>.

Similarweb: Top Websites Ranking [online:] <https://www.similarweb.com/top-websites/poland/computers-electronics-and-technology/social-networks-and-online-communities>.

Starzewski Ł., Jak walczyć z mową nienawiści. 20 rekomendacji RPO dla premiera [online:] <https://bip.brpo.gov.pl/pl/content/jak-walczyc-z-mowa-nienawisci-20-rekomendacji-rpo-dla-premiera>.

HATE SPEECH ON THE SOCIAL NETWORKING SITE FACEBOOK – SELECTED ISSUES

Abstract

The aim of this article is to discuss selected issues related to hate speech on the social networking site Facebook. First, a brief genesis of this phenomenon was discussed, then exemplary definitions of hate speech were provided - those included in the Recommendation No. R (97) 20 of the Committee of Ministers of the Council of Europe, used in the jurisprudence of Polish courts and included in the Facebook Community Standards. Then, the characteristic features of hate speech operating on the Internet and social networking sites were discussed. Data published by Meta was also analyzed - regarding the frequency of occurrence of hate speech on the social networking site Facebook, the number of actions taken in relation to content containing hate speech and percentage data presenting actions taken in relation to cases of hate speech before reporting harmful content by users of the portal in the period from third quarter of 2020 to the first quarter of 2024. Then, the procedure for obtaining data on perpetrators of prohibited acts by law enforcement agencies from Meta was discussed. In this part of the article - data on the number of applications submitted by the relevant Polish authorities and positively considered by Meta in the years 2017-2023 also was analyzed. In the conclusions, the author

summarized the content discussed above and proposed changes in the Polish legal system that could have a positive impact on the prosecution of perpetrators of hate speech crimes, and thus fulfill a preventive function, which could potentially reduce the occurrence of this phenomenon not only on social networking sites such as Facebook, but also offline.

Keywords: hate speech, Facebook, Internet, social networking sites, Polish criminal law.