Op:Code
Open Code for Hate-free Communication

Współfinansowane z Programu Praw, Równości
i Obywatelstwa Unii Europejskiej (2014-2020)

Podsumowanie prac badawczych przeprowadzonych przez Stowarzyszenie „NIGDY WIĘCEJ" w ramach projektu OpCode w okresie wrzesień 2019 – wrzesień 2020

Autor: Jacek Dzięgielewski

STOWARZYSZENIE
„NIGDY WIĘCEJ"
www.nigdywiecej.org

A) Zwięzła prezentacja roli waszej organizacji w ramach projektu OpCode (Open Code for Hate-Free Communication) oraz badań dotyczących fenomenu mowy nienawiści

Stowarzyszenie „NIGDY WIĘCEJ", założone w 1996 roku, jest organizacją non-profit, skupioną na monitorowaniu mowy nienawiści i przestępstw motywowanych nienawiścią, jak również na prowadzeniu działań edukacyjnych. Projekt OpCode ma na celu przeciwstawianie się mowie nienawiści w internecie. Opiera się on na wielowymiarowym podejściu obejmującym monitorowanie, analizowanie, podejmowanie działań, opracowywanie bezpłatnych i otwartych rozwiązań programowych do moderowania treści generowanych przez użytkowników oraz angażowanie się w rzecznictwo i konsolidowanie międzynarodowej sieci współpracy (International Network Against Cyber Hate). Projekt ma również na celu wzmocnienie wysiłków organizacji społeczeństwa obywatelskiego, aby uczestniczyły w europejskich badaniach i stale monitorowały sieć, dostarczając tym samym danych, które mogłyby stanowić podstawę przyszłych regulacji. Ponadto naszym zamiarem jest przedstawienie krajowym i europejskim decydentom oraz firmom informatycznym rekomendacji dotyczących sposobu przeciwdziałania problemowi mowy nienawiści w sieci. Rezultaty tego projektu zostaną uwzględnione w rekomendacjach Komisji Europejskiej odnośnie do sposobów skutecznego przeciwstawienia się nielegalnym treściom w sieci, co świadczy o konieczności współdziałania i wzajemnej wymiany pomiędzy firmami technologicznymi a organizacjami społeczeństwa obywatelskiego na rzecz dobra publicznego. W ramach projektu OpCode w okresie od września 2019 roku do września 2020 Stowarzyszenie „NIGDY WIĘCEJ" wzięło udział w dwóch ćwiczeniach monitoringowych wraz z innymi europejskimi partnerami, jak również współtworzyło specjalny raport na temat nienawiści w sieci, związanej z pandemią koronawirusa.

B) Zwięzła prezentacja krajowego kontekstu w ostatnim roku (2019-2020) w związku z fenomenem mowy nienawiści

Ostatnie lata przyniosły wybuch mowy nienawiści i przestępstw motywowanych nienawiścią. Często nienawiść jest wzniecana przez radykalne ruchy polityczne (chociaż mainstreamowe partie polityczne również odegrały ważną rolę w nasyceniu debaty publicznej nienawiścią i treściami nawołującymi do przemocy). W 2019 i w 2020 roku w Polsce odbyły się wybory do Sejmu i na urząd prezydenta. Podżeganie do nienawiści miało miejsce podczas obydwu kampanii wyborczych. Celem, w który wymierzona była mowa nienawiści polityków i osób publicznych, byli członkowie społeczności LBGT+. Politycy, dziennikarze i celebryci wykorzystywali nienawiść wobec tej społeczności, aby umocnić radykalny elektorat. Wiele przykładów takiej nienawiści ujęto w aktualnym wydaniu „Brunatnej Księgi", czyli monitoringu przestępstw motywowanych nienawiścią i mowy nienawiści prowadzonego przez Stowarzyszenie „NIGDY WIĘCEJ". Rok 2020 przyniósł pandemię COVID-19 i nienawiść skierowaną przeciwko ludziom pochodzenia azjatyckiego, która później rozwinęła się i objęła wszystkich nie-Polaków (osoby indywidualne oraz całe grupy, uznawane za „inne"). Skutkiem pandemii był też wysyp teorii spiskowych, często zawierających elementy antysemickie,

homofobiczne i rasistowskie. Zarówno nienawiść skierowana przeciw LGBT+, jak i nienawiść związana z koronawirusem trafiły na ulice, powodując akty przemocy i dyskryminacji. Zauważalną prawidłowością jest to, że mowa nienawiści ze strony osób publicznych prowadzi do przemocy w realnym życiu.

C) Główne trendy i istotne wyniki pierwszych dwóch ćwiczeń monitorujących Kodeks postępowania

Platformy mediów społecznościowych w Polsce nie reagują dostatecznie na zgłoszenia. Duża część tych treści, które zgłosiliśmy do usunięcia podczas Ćwiczeń Monitoringu, nie została nawet rozpatrzona (albo my nie dostaliśmy żadnej informacji, że została rozpatrzona) i ogromna część tych treści pozostała dostępna w sieci. Facebook wydaje się reagować najczęściej (choć zwykle odpowiedź jest negatywna), podczas gdy Twitter i YouTube generalnie nie odpowiadają na nasze zgłoszenia. Większość przypadków, które zgłosiliśmy do firm IT, zawierała treści antysemickie, rasistowskie, homofobiczne i podżegające do przemocy.

Podczas niedawnego Ćwiczenia Monitoringu Stowarzyszenie „NIGDY WIĘCEJ" zgłosiło 58 nienawistnych komentarzy lub innych materiałów. Większość z nich zgłosiliśmy do Facebooka i Twittera. Bardzo zaskakującym zjawiskiem był brak jakiegokolwiek odzewu ze strony Twittera. Nie otrzymaliśmy żadnej informacji na temat tego, jak oceniono zgłoszone przypadki. Te tweety zawierały ostre sformułowania homofobiczne, antysemickie, rasistowskie – także w odniesieniu do ruchu Black Lives Matter oraz kryzysu związanego z koronawirusem. Podczas poprzednich Ćwiczeń Monitoringu był pewien odzew ze strony Twittera. W niektórych przypadkach nienawistne treści zostały usunięte. A w innych przypadkach otrzymaliśmy informacje, że Twitter uznaje, iż niektóre treści naruszają ich reguły i powinny zostać usunięte – ale tak się nie stało, treści te nie zostały usunięte i nadal były dostępne w serwisie. Podajemy przykłady mowy nienawiści, które zostały zignorowane przez Twittera: „Hitler był LGBT", „Żydzi mają jeden cel – zabijać i zniewolić nie-Żydów, mają prawo mordować, kłamać i oszukiwać", „musimy zniszczyć LGBT", „Żydzi zabili więcej Polaków, niż Hitler", „Wy żydzi [sic] zginiecie, wy zabiliście Jezusa Chrystusa […], powinniście sobie wyciąć penisy i macice". Facebook usunął niektóre ze zgłoszonych treści, na przykład komentarze: „Pier…lić Żydów", „Największym złem całego świata są żydzi [sic] – mordercy Syna Bożego", „Precz z Żydami", „do wora z nim" (komentarz nawołujący do zabicia migranta). Facebook nie usunął następujących komentarzy: „Może ktoś wyrzuci tę plagę z Polski" (o Żydach), „Modlę się codziennie, aby Bóg pomógł oczyścić Polskę z tych zdrajców" (także o Żydach). Facebook nie przedstawił jakiejkolwiek uzasadnienia i nie usunął zgłoszonego komentarza: „Brakuje Hitlera. On by powystrzelał i zagazował te wszystkie małpy" (o czarnoskórych i migrantach). Facebook także nie usunął sprzedaży t-shirtów z hasłami wyrażającymi poparcie dla rasistowskiego mordercy Janusza Walusia, który zastrzelił walczącego z apartheidem działacza i polityka południowoafrykańskiego Chrisa Haniego (nie usunął też koszulek ze sloganem „white lives matter" – tłum. „życie białych ma znaczenie"). Stowarzyszenie „NIGDY WIĘCEJ" zastosowało metodologię opartą na porównaniu grupy

badanej i grupy kontrolnej. Grupa badana składała się z badacza, którzy brał udział w poprzednich Ćwiczeniach Monitoringu, a więc jego adres IP i informacje osobowe znane były firmie IT. Grupa kontrolna składała się z różnych innych nicków tegoż badacza i używanych VPN (wirtualna sieć prywatna), tak że osoba zgłaszająca treści z tego konta była „nieznana" serwisowi społecznościowemu. Wyniki potwierdziły tezę, że media społecznościowe (szczególnie Facebook) reagują chętniej na zgłoszenia pochodzące od osoby, która już wcześniej wysyłała zgłoszenia, niż od zupełnie „nowego" użytkownika.

D) Główne wyzwania i ograniczenia podczas pierwszych dwóch ćwiczeń monitorujących Kodeks postępowania

Największym wyzwaniem pozostaje przekonanie platform mediów społecznościowych, aby zaczęły reagować na nasze zgłoszenia. Bez odpowiedniego odzewu trudno jest zwalczać mowę nienawiści skutecznie. Ponadto technologia i procedury używane do oceny zgłoszeń są mgliste i niewyraźnie zdefiniowane; nie wiemy, dlaczego jedno zgłoszenie zostało rozpatrzone, a inne nie; nie wiemy, według jakich kryteriów jedne treści zostają usunięte, a inne nie. Firmy IT powinny zaprojektować jasny system zgłaszania nienawistnych oraz/lub nielegalnych treści i ustalić przejrzysty i sprawny sposób komunikowania się z użytkownikami sieci, którzy wysyłają zgłoszenia. To niedopuszczalne, że treści pełne nienawiści, bez względu na to, czy mieszczące się w normach prawnych, czy nie, nie są usuwane przez platformy mediów społecznościowych. Zauważyliśmy, że firmy IT reagowały głównie na „najcięższe" treści, takie jak nawoływanie do zabójstwa, ale w większości ignorowały „zwykłą" werbalną homofobię, rasizm i antysemityzm. W naszej opinii także stwierdzenia pochwalające antysemityzm lub rasizm stwarzają wielkie niebezpieczeństwo (przemoc zwykle zaczyna się od słów). Tak więc ludzie decydujący o tym, który komentarz usunąć, a który zostawić bez zmian, powinni reagować na wszystkie zgłoszenia, a nie tylko na „ciężkie przypadki". To dotyczy głównie Facebooka, ponieważ Twitter w ogóle nie reaguje na większość zgłoszeń, nie usuwając ani nawet nie oceniając tweetów zawierających ewidentne nawoływanie do przemocy (na przykład tweet z linkiem do artykułu o Unabomberze – Tedzie Kaczyńskim, terroryście, który, w opinii autora tegoż artykułu, robił świetną robotę, zwalczając „lewaków").

E) Główne trendy i kluczowe zagadnienia zidentyfikowane w ramach badania nienawiści związanej z koronawirusem

Badanie nienawiści związanej z koronawirusem wykazało, że fala nienawiści była pierwotnie wymierzona w ludzi pochodzenia azjatyckiego. Później, najprawdopodobniej z powodu przymusowej izolacji, nienawiść ta rozszerzyła się na wszystkich nie-Polaków. Zauważyliśmy także wysyp teorii spiskowych, często rozpowszechnianych przez osoby publiczne, jak politycy, celebryci, artyści, osoby duchowne itd. Teorie te często zawierały elementy antysemickie, rasistowskie i homofobiczne. Rozpowszechnianie teorii spiskowych jest także wysoce niebezpieczne, ponieważ wystawia zdrowie publiczne na duże ryzyko, gdyż wiele

takich teorii wiąże się z negowaniem istnienia pandemii lub sprzeciwem wobec szczepień. Schemat obserwowany w nienawiści związanej z pandemią jest podobny do poprzednich wybuchów zachowań ksenofobicznych, rasistowskich, antysemickich i homofobicznych. Niektórzy politycy i partie polityczne wykorzystują istniejące w społeczeństwie lęki lub stwarzają nowe źródła strachu w celu zdobywania głosów. W 2015 r. można to było zaobserwować w nienawistnej narracji skierowanej przeciwko uchodźcom (identyfikującej wszystkich uchodźców/imigrantów jako terrorystów itp.), w rezultacie czego dochodziło do aktów dyskryminacji i przemocy. W latach 2019 i 2020 wielu polityków wykorzystywało homofobię jako paliwo polityczne. To nastawienie wytworzyło poczucie strachu w bardziej podatnej na takie emocje części elektoratu, która przeraziła się „ideologią LGBT" lub „ideologią gender". Tego rodzaju przekaz zawierał też fake newsy, np. powiązanie orientacji seksualnej z pedofilią lub manipulowanie dyrektywą Światowej Organizacji Zdrowia po to, aby przedstawiać edukację seksualną na przykład jako „uczenie małych dzieci, jak się masturbować". Politycy używający nienawistnego języka homofobicznego osiągnęli sukces w wyborach. Podobnie, osoby publiczne takie jak politycy, dziennikarze, duchowni (np. księża katoliccy) promowali teorie spiskowe i nienaukowe podejście do tematu pandemii. Stwierdzenia takie, powtarzane często, stworzyły bezpieczną przestrzeń dla tych, którzy zaczęli używać przemocy lub dyskryminowali ludzi, którym bezpodstawnie zarzucano roznoszenie wirusa. To wszystko pokazuje, jak wielka jest odpowiedzialność osób publicznych.

F) Główne wyzwania i ograniczenia w dokumentowaniu nienawiści związanej z koronawirusem w ramach prowadzonych badań

Przede wszystkim dużym problemem jest ogólna niechęć platform mediów społecznościowych do odpowiadania na zgłoszenia i do usuwania nienawistnych treści. Inny problemem jest nastawienie mediów – zarówno tradycyjnych, jak i społecznościowych – ponieważ albo dają przestrzeń do wypowiedzi ludziom rozsiewającym fake newsy i teorie spiskowe, albo nie zwracają należytej uwagi, gdy trzeba usunąć takie treści, które zostały już opublikowane w sieci.

Innym istotnym ograniczeniem, jak wykazano powyżej, jest również to, że platformy mediów społecznościowych nie zapewniają jasnych zasad komunikacji z użytkownikami, którzy zgłaszają nienawistne treści. Nie możemy działać skutecznie, jeśli algorytmy są nieprzejrzyste i nie jest zrozumiałe, na jakiej podstawie zapadają decyzje, które treści mają być usunięte, a które nie. Jakkolwiek kilka platform mediów społecznościowych opracowało strategie do zwalczania fake newsów i dezinformacji (głównie Facebook i YouTube) jeszcze wiele materiałów video i wpisów na tych stronach zawiera teorie spiskowe oraz informacje niesprawdzone i potencjalnie szkodliwe.