

Debaty

o współczesnym świecie



MUZEUM
WARSZAWY

90 LAT
1936—2026



POLSKA DEBATUJE

MOWA NIENAWIŚCI

część 1.

Należy karać za użycie mowy nienawiści w Internecie



WPROWADZENIE

Kiedy dziś myślimy o mediach społecznościowych, trudno uwierzyć, że ich historia zaczęła się całkiem niepozornie. Na początku lat 2000. w Internecie pojawiały się pierwsze serwisy, które pozwalały użytkownikom zakładać profile i łączyć się ze znajomymi - jak Friendster (2002) czy MySpace (2003). To właśnie MySpace był jednym z pierwszych miejsc, gdzie ludzie masowo „zakładali swoje strony”, a około 2004 roku zgromadził już ponad milion aktywnych użytkowników miesięcznie, co wtedy wydawało się absolutnym fenomenem.

Prawdziwy przełom przyniosła platforma Facebook, uruchomiona w 2004 roku jako skromny projekt dla studentów Harvardu. W ciągu kolejnych miesięcy serwis otworzył się najpierw na inne uczelnie, a później na wszystkich internautów, szybko zyskując popularność.

W tym samym czasie rodziła się zupełnie nowa forma internetowej ekspresji: wideo online. Gdy w 2005 r. pojawił się YouTube, ludzie po raz pierwszy mogli nie tylko oglądać zabawne filmiki, ale też tworzyć własne treści i natychmiast dzielić się nimi ze światem. Ten rozwój był jedynie zapowiedzią tego, co miało dopiero nadejść.

Lata 2006–2010 można uznać za moment, w którym media społecznościowe wkroczyły do absolutnego mainstreamu. MySpace wciąż był wówczas niezwykle popularny - w 2006 roku stał się najczęściej odwiedzaną stroną w USA, ale to Facebook i nowe platformy zaczęły coraz mocniej przejmować uwagę użytkowników. Twitter dynamicznie zyskiwał na znaczeniu, oferując zupełnie nowy, szybki sposób komunikacji, a Facebook, już otwarty dla wszystkich, przekroczył 100 milionów aktywnych kont¹ w 2008 roku i tym samym zdefiniował sposób, w jaki zaczęliśmy budować swoje życie towarzyskie w sieci.

Przełomowym momentem okazała się także ekspansja smartfonów i rozwój aplikacji mobilnych. Gdy w 2010 roku zadebiutował Instagram, a dostęp do internetu mobilnego stał się standardem, media społecznościowe przeniosły się z komputerów do naszych kieszeni. Od tego momentu stały się stałym elementem codziennego życia - zawsze obecne, zawsze pod ręką, towarzyszące nam od poranka do późnego wieczora

Dziś skala korzystania z mediów społecznościowych jest trudna do porównania z czymkolwiek innym w cyfrowym świecie. Na początku 2025 roku globalnie działa już ponad 5 miliardów profili użytkowników, co oznacza, że z social mediów korzysta ponad połowa ludzkości, a przeciętny internauta spędza w nich każdego dnia około dwóch i pół godziny.

¹ Schroeder, Stan (August 26, 2008). "Facebook's 100 Million Users: How Much are They Worth?". Mashable.

Wciąż dominują platformy takie jak Facebook, Instagram i YouTube, ale w ostatnich latach wyjątkowo dynamicznie rośnie TikTok, który w 2025 roku przyciąga około 1,6 miliarda aktywnych użytkowników miesięcznie, zmieniając sposób konsumowania krótkich treści wideo na całym świecie

W Polsce obraz wygląda bardzo podobnie. Na początku 2025 roku z mediów społecznościowych korzystało ponad trzy czwarte społeczeństwa, a internet jest powszechnie dostępny w większości gospodarstw domowych. Użytkownicy zwykle funkcjonują na wielu platformach jednocześnie, korzystając z każdej w nieco innym celu: YouTube pełni rolę „televizji internetowej”, Facebook i Messenger pomagają utrzymywać lokalne więzi i codzienny kontakt, Instagram i TikTok odpowiadają za treści wizualne i krótkie formaty, a LinkedIn obsługuje sferę zawodową. Razem tworzy to codzienny, wielowarstwowy ekosystem, w którym media społecznościowe stały się naturalnym tłem życia - miejscem, do którego zaglądamy automatycznie, bez zastanowienia, często wiele razy dziennie.

Właśnie dlatego, że media społecznościowe są dziś tak powszechne, tak łatwe w obsłudze i tak głęboko wpisane w rytm naszego dnia, coraz wyraźniej widzimy również ich drugą stronę. Przestrzeń, która łączy miliardy ludzi i daje niespotykane wcześniej możliwości komunikacji, jest jednocześnie miejscem, w którym z ogromną prędkością mogą rozprzestrzeniać się treści krzywdzące, agresywne czy nienawistne. A to sprawia, że pytanie o to, jak powinna wyglądać odpowiedzialność za słowa w Internecie, przestaje być tylko teoretyczną dyskusją - staje się jednym z najważniejszych wyzwań współczesnej debaty publicznej.

MOWA NIENAWIŚCI DAWNIEJ

Zanim mowa nienawiści przeniosła się do internetu, działała przede wszystkim poprzez prasę, radio, plakaty i szkolne podręczniki. Jej mechanizmy były podobne do dzisiejszych, choć narzędzia wyglądały inaczej. W Niemczech lat trzydziestych regularnie wykorzystywały ją gazety takie jak Der Stürmer, które publikowały sensacyjne historie, powtarzały kłamstwa o rzekomych zbrodniach Żydów i zamieszczały karykatury przedstawiające ich jako istoty zagrażające społeczeństwu. Wszystko to służyło ich dehumanizacji i budowaniu przekonania, że eliminacja tej grupy jest rzekomo konieczna dla bezpieczeństwa państwa. Podobne działania dotyczyły Romów czy Słowian, nazywanych „podludźmi”, co miało uzasadniać ich prześladowanie i przemoc ze strony aparatu państwowego. W Stanach Zjednoczonych przez dekady funkcjonowania systemu Jim Crow nienawiść podsycano poprzez karykatury w prasie i kulturze masowej, przedstawiające

Afroamerykanów w sposób ośmieszający, fałszujący rzeczywistość i utrwalający przekonanie o ich rzekomej niższości. Taki przekaz miał legitymizować segregację i akceptację przemocy. W Rwandzie, tuż przed i w trakcie ludobójstwa w 1994 roku, radio RTL M nazywało Tutsich „karaluchami” i wprost nawoływało do ich atakowania; audycje łączono z muzyką i potocznym językiem, co sprawiało, że propaganda trafiała do bardzo szerokiego grona odbiorców i obniżała opór wobec przemocy. W Europie końca XX wieku podobne mechanizmy obserwowano w czasie wojen jugosłowiańskich, gdzie media państwowe wykorzystywały pamięć historycznych krzywd, aby podsycać wrogość między grupami etnicznymi i usprawiedliwiać agresję. Antyromski język pogardy również ma długą historię - w wielu krajach Europy stereotypy i obelgi powtarzane przez polityków i media wzmacniały społeczne przyzwolenie na wykluczanie tej mniejszości. Nawet w krajach demokratycznych istniały momenty, gdy politycy legitymizowali uprzedzenia. Klasycznym przykładem jest brytyjskie przemówienie „Rivers of Blood” z 1968 roku, które ostrym, alarmistycznym językiem uderzało w imigrantów i wzmocniło napięcie rasowe w Wielkiej Brytanii. Wszystkie te przypadki łączy wspólny fundament: powtarzalne etykiety, ośmieszanie, wzbudzanie strachu i prezentowanie danej grupy jako zagrożenia, co miało przygotować społeczeństwo do tolerowania lub podejmowania działań przemocowych wobec niej.

MOWA NIENAWIŚCI W MEDIACH SPOŁECZNOŚCIOWYCH

Mowa nienawiści to taki rodzaj wypowiedzi, który nie polega na ostrym krytykowaniu czyjejś tezy czy poglądów. Mowa nienawiści uderza w samą tożsamość danej osoby lub grupy. Chodzi o język, który ma upokorzyć, wykluczyć lub przestraszyć, a czasem także zachęcić innych do wrogich działań – ze względu na cechy „przynależne” (etniczność, religia, pochodzenie narodowe, kolor skóry, płeć, orientacja seksualna, niepełnosprawność, tożsamość płciowa, wiek, itp.). W praktyce to nie pojedynczy brzydki epitet, tylko cały sposób mówienia: przypisywanie negatywnych cech „z definicji”, odczłowieczanie (porównania do zwierząt, brudu, choroby), etykietowanie („oni tacy są”), podżeganie („trzeba ich...”, „powinniśmy z nimi...”) i nawoływanie do bojkotu, przemocy czy „wypchnięcia” z przestrzeni publicznej. W wersji internetowej ten język bywa pozornie „żartobliwy” albo zaszyfrowany w memach i skrótach, ale skutek jest podobny: ma realnie ograniczać bezpieczeństwo i możliwość zabierania głosu przez atakowanych.

W sieci ten język ubiera się w lekkie, potoczne formy, żeby łatwiej przechodzić „pod radarem”. Memy, które na pierwszy rzut oka wyglądają jak żart, ale budują skojarzenia z brudem, inwazją czy pasożytem; krótkie filmiki z ironiczną muzyką i napisami, które trywializują bardzo ciężkie treści. Zrzuty ekranu wyrwane z kontekstu, z dopisanymi sugestywnym komentarzami i multiplikowane w kolejnych udostępnieniach. „Mrugnięcia okiem” w postaci kodów, skrótów i emotikonów, które dla wtajemniczonych znaczą dużo więcej niż dla kogoś z zewnątrz. Do tego dochodzą zorganizowane fale nękania, gdy po jednym podaniu dalej na czyjś profil w kilka godzin spada lawina obelg, docinków i prywatnych wiadomości – ludzi nie łączy nic poza tym, że algorytm podsunął im tę samą treść, ale efekt takich działań po drugiej stronie ekranu jest realny: poczucie osaczenia, zagłuszony głos, autocenzura. Ważnym elementem są też „przeróbki” i deepfaki: obrazy lub krótkie filmy, które ośmieszają, seksualizują albo demonizują całe grupy, często bez nawet jednego wulgarnego słowa, za to z potężnym ładunkiem emocjonalnym, które zbierają reakcje szybciej niż jakikolwiek tekst. W tle działa również nieformalna infrastruktura hejtu: publiczne i półprywatne kanały służące umawianiu tzw. „najazdów” na profile, tworzeniu list „wrogów”, przypinaniu gotowych pakietów obraźliwych komentarzy i gotowych odpowiedzi, które potem widać pod setkami postów niczym skopiowane z jednego szablonu. I jeszcze jeden wymiar, o którym łatwo zapomnieć: internet nie jest równoległym światem, więc to, co tam mówimy, nie zostaje tylko w sieci – jeżeli ktoś należy do grupy, o której od miesięcy czyta, że jest „szkodnikiem” albo „zagrożeniem”, prędzej czy później poczuje skutki poza ekranem – w pracy, w szkole, na ulicy. Mechanizm ten zachodzi również wśród osób niedotkniętych bezpośrednio mową nienawiści. Jeżeli ktoś przez dłuższy czas żywi się takimi memami i komentarzami, zaczyna naprawdę wierzyć, że cele tych ataków „nie czują się tak jak my”, „nie zasługują na tyle samo”, „sam się proszą”, i proste złośliwe żarty stają się wygodnym alibi dla braku empatii. Dlatego współczesna mowa nienawiści jest tak skuteczna: łączy lekkość formy z ciężarem skutków, działa szybko, na skróty, przez obrazy i przyjemne dla ucha dźwięki, wchodzi do przeglądanych treści przy okazji śmiesznego wideo, a potem – już bez wykrzykników i wielkich słów – konsekwentnie zwięża komuś pole do zabierania głosu i bycia obecnym w przestrzeni, która dla większości z nas jest dziś podstawową sceną życia społecznego.

Warto podkreślić, że rdzeń współczesnych zjawisk niewiele różni się od mechanizmów sprzed epoki platform: nadal chodzi o ten sam schemat odczłowieczania, straszenia i wskazywania wroga, a dzisiejsze memy i krótkie filmiki pełnią funkcję dawnych środków masowego przekazu, tylko w nowym opakowaniu. Podobna jest też rola „autorytetu”

nadawcy i efekt „pieczęci wiarygodności”². W czasach przed internetowych takie przyzwolenie często nadawały media kontrolowane przez państwo lub głośne wystąpienia polityków. W sieci tę funkcję pełnią wpływowe konta i duże profile: kiedy one nadają ton, przekaz szybciej zyskuje zasięg i społeczne przyzwolenie.

Różnice są jednak kluczowe. Po pierwsze skala i tempo. Z mediów społecznościowych korzysta dziś ponad pięć miliardów osób. Oznacza to, że nawet marginalny komunikat może w kilka godzin stać się globalnym „hitem”, a krótkie wideo – format o najwyższej dynamice udostępnień – potrafi zdominować uwagę milionów. Po drugie architektura uwagi. Algorytmy rekomendacji filtrują i porządkują strumienie treści tak, by maksymalizować zaangażowanie. A ponieważ złość i strach są silnymi wyzwalaczami reakcji, materiały agresywne i polaryzujące mają naturalną przewagę. Po trzecie, próg wejścia i siła nagłośnienia. Kiedyś do nadawania potrzebne były redakcje, anteny i budżety, dziś wystarczy zwykłe konto i kilkanaście sekund, by opublikować coś, co szybko trafi do wielu osób. To obniża barierę nie tylko dla wartościowych głosów, ale też dla nienawiści i nękania. Po czwarte, sam sposób działania sieci. Internet ułatwia zbiorowe ataki na profile i błyskawiczne powielanie tych samych obelg w krótkim czasie, co działa jak mnożnik presji na osobę będącą celem. Po piąte, warstwa obrazu i dźwięku. Krótkie filmy, podkłady muzyczne, filtry i różnego rodzaju przeróbki potrafią zbudować bardzo silny przekaz emocjonalny, który łatwo przyciąga uwagę i rozchodzi się szybciej niż tekst. Wreszcie znaczenie środowiska. Media społecznościowe stały się codzienną infrastrukturą życia - miejscem pracy, nauki i rozmów - więc to, co kiedyś żyło w rubrykach gazet, dziś natychmiast krąży między aplikacjami i wraca do świata offline. W takich warunkach mowa nienawiści rzadko bywa „tylko słowami”: to, co widzimy przeglądając treści, realnie wpływa na relacje w pracy i w szkole, na poczucie bezpieczeństwa osób atakowanych i na to, czy w ogóle decydują się zabierać głos.

² „Pieczęć wiarygodności” to określenie używane na opisaniu sytuacji, w której *autorytet nadawcy* wzmacnia odbiór i skuteczność przekazu.

“Należy karać za użycie mowy nienawiści w Internecie.”

DRUŻYNA PROPOZYCJI (“ZA”)

Poniższa sekcja to materiał do pracy dla całej drużyny propozycji. Jej celem jest uporządkowanie przygotowań i ułatwienie budowy spójnego wystąpienia: od zdefiniowania problemu, przez uzasadnienie celu karania, po wskazanie skutecznych i proporcjonalnych rozwiązań. Lista ta nie wyczerpuje tematu - wskazuje kierunek pracy, podsuwa możliwe konstrukcje argumentów i ma ułatwić własne doprecyzowanie linii argumentacyjnej, przykładów i definicji.

Zaczynając przygotowanie do debaty zastanów się nad trzema podstawowymi aspektami:

1. **Cel** - po co społeczeństwo ma reagować: jakie szkody realnie ograniczamy i jakie dobra wspólne wzmacniamy.
2. **Granica** - jak rozróżniamy ostrą krytykę (dopuszczalną) od mowy nienawiści (niedopuszczalnej) oraz jak uzasadniamy, że sankcja nie jest zamachem na debatę, lecz warunkiem jej podmiotowości.
3. **Skuteczność i proporcja** – jakie środki są adekwatne, mierzalne i skuteczne, aby cele były osiągalne bez nadmiernej ingerencji.

CEL

Pytanie 1. Jakie szkody uzasadniają wprowadzanie sankcji?

W debacie można wyróżnić trzy poziomy szkody: indywidualny (zdrowie psychiczne, poczucie bezpieczeństwa), grupowy (wyciszenie mniejszości w debacie), społeczny (eskalacja konfliktów, polaryzacja, „efekt zarażania” nienawiścią).

Pytanie 2. Co chcemy osiągnąć poprzez karanie?

Prewencja (zmniejszenie liczby ataków), ochrona (szybkie zatrzymanie szkody), norma społeczna (jasny sygnał: to nie jest akceptowalne), sprawiedliwość (konsekwencje dla sprawcy, wsparcie dla poszkodowanych).

Pytanie 3. Co się dzieje, gdy nie reagujemy?

„Spirala normalizacji” - to, co niekarane, bywa powielane; ofiary wycofują się z debaty; rośnie przyzwolenie na ostrzejsze formy przemocy..

GRANICE WOLNOŚCI SŁOWA

Pytanie 4. Dlaczego karanie mowy nienawiści nie niszczy wolności słowa?

Jak można uzasadnić, że kara za mowę nienawiści nie narusza wolności słowa, lecz ją chroni? Podkreśl różnicę między krytyką poglądów (chronioną) a atakiem na tożsamość (raniącym i wykluczającym).

Pytanie 5. Jak odróżnić ostrą krytykę od nienawiści?

- Cel wypowiedzi (pogląd vs cecha wrodzona)
- Odczłowieczanie/ wzywanie do wykluczenia
- Kontekst i skutek (np. fala gróźb po „żarcie”).

SKUTECZNOŚĆ ROZWIĄZAŃ

Pytanie 6. Jakie sankcje są proporcjonalne i działają?

Pokaż różnicowanie kary w zależności od czynu np.: ostrzeżenia i obowiązkowa treść edukacyjna przy lżejszych naruszeniach, kary finansowe lub czasowe ograniczenia przy recydywie, działania naprawcze (np. przeprosiny, wsparcie dla poszkodowanych) przy poważnych sprawach.

Pytanie 7. Skąd pewność, że to zadziała?

Odwołaj się na przykład do intuicyjnej logiki: gdy zachowanie przestaje być „bez kosztów”, spada skłonność do jego powielania; dodatkowo pojawia się „efekt chłodzący” na otoczeniu sprawcy (mniej lajków na hejt, mniej chętnych do dołączania do „chóru”).

Pytanie 8. Jak mierzyć skuteczność?

Podpowiedź: wskaźniki „przed-po”: liczba zgłoszeń i recydywy, spadek powtarzalności tych samych obelg w danym wątku, krótszy czas reakcji, utrzymanie aktywności poszkodowanych (np. nie kasują kont, wracają do dyskusji).

ODPOWIEDZIALNOŚĆ

Pytanie 9. Kto ma ponosić odpowiedzialność?

W debacie skup się na odpowiedzialności jednostkowej osób publikujących treści - to autor wypowiedzi decyduje o języku i formie komunikatu, dlatego to on w pierwszej kolejności odpowiada za skutki swoich słów.

BEZPIECZEŃSTWO I PSYCHOLOGIA

Pytanie 10. Jak mowa nienawiści wpływa na jednostkę?

Niska „kosztowność” hejtu dla sprawcy kontra wysoki koszt dla ofiary (stres, lęk, bezsenność, unikanie sieci).

Pytanie 11. Czy kara ma wymiar sygnału społecznego?

Jak normy prawne oddziałują na formowanie norm społecznych? Sankcjonowanie mowy nienawiści w Internecie wzmacnia zakaz dyskryminacji i stabilizuje standard życia publicznego, oparty na poszanowaniu godności oraz odpowiedzialności za słowo.

Pytanie 12. Dlaczego dehumanizacja jest tak groźna?

Dehumanizacja „odrywa” nas od empatii - odbiorcom łatwiej zaakceptować ostrzejsze formy krzywdzenia.

DRUŻYNA OPOZYCJI (“PRZECIW”)

Poniższa sekcja to materiał do pracy dla całej drużyny opozycji. Jej celem jest uporządkowanie przygotowań i ułatwienie budowy spójnego wystąpienia: od zdefiniowania problemu, przez uzasadnienie celu karania, po wskazanie skutecznych i proporcjonalnych rozwiązań. Lista ta nie wyczerpuje tematu - wskazuje kierunek pracy, podsuwa możliwe konstrukcje argumentów i ma ułatwić własne doprecyzowanie linii argumentacyjnej, przykładów i definicji.

RYZIKO NADREGULACJI

Pytanie 1. Jakie są największe zagrożenia związane z karaniem za treści słowne?

Możesz wskazać ryzyko nadmiernej ingerencji w debatę, efekt mrozący wobec ostrych, lecz merytorycznych wypowiedzi oraz możliwość „przekwalifikowywania” kontrowersji w wykroczenia.

Pytanie 2. Czy mechanizmy karne mogą być nadużywane przez władzę lub grupy wpływu?

Możesz uwypuklić problem asymetrii siły: ci, którzy dysponują zasobami, częściej kształtują interpretację pojęć i inicjują postępowania, co grozi selektywną ochroną.

Pytanie 3. Czy kategoria „mowa nienawiści” może zostać jednoznacznie zdefiniowana?

Możesz zwrócić uwagę na płynność granicy między surową krytyką a nienawiścią, różnice kulturowe i kontekstowe, a także na trudność w obiektywnym ustalaniu intencji.

WOLNOŚĆ SŁOWA

Pytanie 4. Czy wprowadzenie kar może doprowadzić do autocenzury i tłumienia debaty?

Możesz podkreślić, że strach przed błędną kwalifikacją wypowiedzi skłania do przemilczeń, zubażając pluralizm i ograniczając spór do bezpiecznych banałów. Zastanów się, dlaczego to ważne.

SKUTECZNOŚĆ KAR

Pytanie 5. Czy karanie faktycznie redukuje mowę nienawiści czy raczej spycha ją do zamkniętych przestrzeni?

Możesz pokazać, że karanie za mowę nienawiści może doprowadzić do przeniesienia się użytkowników do “głębszych” części internetu. Sprawiając, że zarówno treści te będą trudniejsze do namierzenia jak i ich autorzy pozostaną anonimowi.

Pytanie 6. Jakie alternatywy wobec kar mogą być skuteczniejsze?

Możesz wskazać: wyraźne standardy społeczności, szybkie, proporcjonalne moderowanie treści, narzędzia wyciszania, kontrnarracje oraz programy edukacyjne wzmacniające kompetencje odbiorców, wsparcie dla grup dotkniętych mową nienawiści.

PRAKTYCZNE TRUDNOŚCI

Pytanie 7. Kto miałby decydować, co jest nienawiścią, a co nie?

Możesz podnieść kwestię standardów oceny, kompetencji decydentów i przejrzystości procedur. Możesz akcentować potrzebę kontekstu, wielopoziomowej weryfikacji oraz zachowania domniemania dobrej wiary tam, gdzie brak jednoznacznych przesłanek.

Pytanie 8. Czy kara ma sens przy trudności ustalania tożsamości użytkowników?

Możesz zauważyć, że anonimowość i łatwość zakładania nowych kont osłabiają prewencję, a spór przenosi się z jakości treści na technalia identyfikacji.

SPOŁECZEŃSTWO I KULTURA

Pytanie 9. czy podejście represyjne rozwiązuje problem, czy raczej go maskuje?

Możesz twierdzić, że bez pracy nad przyczynami (uprzedzenia, niewiedza, polaryzacja) same sankcje działają krótkotrwale.

PYTANIA DO OBU STRON (pomagające przewidzieć kontrargumenty)

1. Jakie minimum wolności słowa jest niezbędne w demokratycznej debacie?
2. Co jest większym problemem: nadmierna tolerancja wobec nienawiści czy nadmierna cenzura?
3. Czy wszystkie grupy społeczne są jednakowo narażone na skutki mowy nienawiści?
4. Jaką rolę w ograniczaniu problemu mogą odgrywać edukacja, kampanie społeczne i moderatorzy platform?
5. Czy formalne kary zmieniają postawy, czy tylko zachowania?
6. Jak rozróżnić realne zagrożenie od prowokacji, ironii czy mowy emocjonalnej?

Organizator:



Partner:



Współfinansowanie:



Partner Edukacji
Muzeum Warszawy:



Patron
medialny:



TVP | info_