

sieć obywatelska  
**WATCHDOG**<sup>^</sup>

**PRAVDA**

**DEMAGOG**

**CyberDefence 24**

**SPIDER'SWEB** +

**Nauka**  
to lubię

**CRAZYNAUKA**

Instytut  
**ZAMENHOFA**

*Stemcel*

**NASK**

**#FakeHunter**



**pap**

POLSKA AGENCJA PRASOWA

# Kodeks dobrych praktyk

## Wspólnie przeciw dezinformacji

kwiecień 2022 r.

## SPIS TREŚCI

<b>Podsumowanie - czyli o czym przeczytasz w dokumencie.....</b>	<b>3</b>
<b>Wstęp.....</b>	<b>5</b>
<b>Czym jest dezinformacja.....</b>	<b>6</b>
<b>Jakimi kanałami rozpowszechniane są treści dezinformacyjne.....</b>	<b>7</b>
<b>a. Trolle.....</b>	<b>8</b>
<b>b. Boty.....</b>	<b>8</b>
<b>c. Liderzy opinii.....</b>	<b>9</b>
<b>d. Politycy.....</b>	<b>9</b>
<b>e. Zwyczajni użytkownicy.....</b>	<b>10</b>
<b>f. Media.....</b>	<b>10</b>
<b>Jakie są najpopularniejsze metody działań dezinformacyjnych w polskiej infosferze.....</b>	<b>12</b>
<b>Wybrane metody działań dezinformacyjnych.....</b>	<b>12</b>
<b>Jak monitorować treści dezinformacyjne.....</b>	<b>15</b>
<b>Informacje i opinie w kontekście dezinformacji - jak je odróżnić.....</b>	<b>16</b>
<b>Jak informować o pojawieniu się treści dezinformacyjnych.....</b>	<b>17</b>
<b>Jak rozpoznawać zjawisko dezinformacji.....</b>	<b>18</b>
<b>Jak przeciwdziałać zjawisku dezinformacji.....</b>	<b>19</b>
- Fact-checking.....	19
- Edukacja użytkowników.....	19
- Wysokie standardy działań medialnych.....	20
- Zgłaszanie / blokowanie treści / Trust flagging kont / treści dezinformacyjnych.....	20
- Rola platform społecznościowych.....	21

## Podsumowanie - czyli o czym przeczytasz w dokumencie?

- Dezinformacja - bez względu na to czy świadoma, czy też nie - wywiera wpływ na postawy w realnym świecie, co może skutkować poważnymi konsekwencjami dla zdrowia i życia ludzi;
- Dezinformacja może dotyczyć nie tylko faktów, ale również opinii. Szkodliwe opinie, które wprowadzają w błąd co do stanu faktycznego i wywołują agresywne postawy lub zachowania powinny być piętnowane na równi z nieprawdziwymi informacjami;
- Wykorzystywanie opinii, spekulacji i plotek w celu wprowadzania w błąd, co do faktycznego stanu rzeczy powinno być wskazywane i zwalczane. Treści skutkujące okazywaniem agresji i nienawiści w relacjach społecznych powinny być sygnalizowane i w miarę możliwości, zgodnie z prawem i poszanowaniem konstytucyjnej wolności słowa, zwalczane;
- Boty i trolle internetowe to tylko część kanałów jakimi udostępniane są treści dezinformacyjne. Działania dezinformacyjne często łączą pozornie niepowiązane ze sobą czynniki i kanały komunikacji;
- Źródłem informacji dezinformacyjnych może być każdy - zwykły użytkownik, lider opinii, polityk czy dziennikarz;
- Szczególna odpowiedzialność za publikowane treści spoczywa na najbardziej zasięgowych autorach - liderach opinii, politykach i przedstawicielach mediów;
- Kampanie dezinformacyjne mogą wykorzystywać różne środki, jednak mają jeden wspólny mianownik - ich celem jest zmiana postawy lub zachowania grupy osób, do której jest kierowana;
- Kampanie dezinformacyjne mogą być prowadzone w różnym okresie czasu - najskuteczniejsze trwają w charakterze ciągłym/wieloletnim;

- Monitorowanie treści dezinformacyjnych nie może się ograniczać jedynie do analizy ilościowej i automatycznego określania sentymentu publikowanych wzmianek;
- Skuteczny monitoring treści szkodliwych musi zawierać analizę jakościową - uwzględniającą wiele czynników pomagających scharakteryzować skalę zjawiska dezinformacji;
- W kontekście działań dezinformacyjnych plotki, opinie i spekulacje często mieszają się z faktami. Daje to pole do skutecznego wykorzystywania wielu technik dezinformacji. Odpowiednie rozpoznawanie tych treści i ich klasyfikowanie jest bardzo ważną częścią przeciwdziałania zjawisku dezinformacji;
- Wyrażanie opinii, zadawanie pytań z tezą czy cytowanie treści może również być działaniem o charakterze dezinformacyjnym. Wyznacznikiem jest tutaj wprowadzanie w błąd co do stanu faktycznego i poziom szkodliwości treści - nie jej forma;
- Nieodpowiednia, nieprofesjonalna komunikacja dotycząca wykrytych treści dezinformacyjnych może przynieść więcej szkód niż pożytku. Kluczową kwestią jest forma prezentowanego komunikatu;
- Platformy społecznościowe powinny dostosować swoje działania i narzędzia do realiów, w jakich się znajdujemy. Ich bierność może spowodować nieodwracalne szkody dla zdrowia i życia ludzi, a także relacji społecznych;
- Kampanie edukacyjne, odpowiedzialne działanie mediów, oznaczanie, zgłaszanie treści dezinformacyjnych są kluczowymi działaniami pozwalającymi na skuteczną walkę z dezinformacją.

## Wstęp

W obecnych czasach społeczeństwo ma styczność z ogromną ilością danych w infosferze, które coraz trudniej jest zweryfikować. Szerzące się zmanipulowane lub fałszywe informacje są zjawiskiem zagrażającym debacie publicznej. Jednym z celów dezinformacji jest tworzenie podziałów wśród społeczeństwa poprzez narzucanie określonej narracji. W konsekwencji wzbudza ona poczucie zagrożenia, strachu, a także odrzucenia powszechnie znanych faktów, w tym również popartych badaniami naukowymi, a w skrajnych przypadkach prowadzi także do propagowania nienawiści wobec wybranych grup społecznych, etnicznych, narodowościowych, wyznaniowych, grup związanych z orientacją seksualną i tożsamością płciową oraz osób niepełnosprawnych. Techniki tworzenia i rozpowszechniania treści dezinformacyjnych są wysoce zaawansowane i wymagają często wiele trudu, aby je zdemaskować.

Ofiarami zafałszowanych komunikatów mogą padać nie tylko przeciętni użytkownicy Internetu, ale również media czy osoby powszechnie uznawane za autorytety. W duchu odpowiedzialności za jakość debaty publicznej oraz woli przeciwdziałania zjawisku dezinformacji został opracowany Kodeks Dobrych Praktyk. Eksperti *Crazy Nauki*, *CyberDefence24*, *FakeHunter*, *Instytutu Zamenhofa*, *Fundacji Rozwoju Przez Całe Życie*, *Fundacji Nauka. To Lubię*, *Spider's Web+*, *Stowarzyszenia Demagog*, *Stowarzyszenia Pravda*, *Stowarzyszenia Sieć Obywatelska Watchdog Polska*, *Naukowej i Akademickiej Sieci Komputerowej – PIB* przygotowali opracowanie, które tłumaczy, jak rozumieć zachodzące w polskiej infosferze procesy dezinformacyjne, wskazuje ich charakterystykę oraz wybrane sposoby zapobiegania rozprzestrzeniania szkodliwych treści.

Wypracowany Kodeks Dobrych Praktyk jest jednym z kamieni milowych w uspołnieniu standardów walki z dezinformacją. Eksperti zawarli kluczowe zagadnienia w tym obszarze, a także w zakresie rekomendowanych działań i zmian, które powinny pojawić się zarówno wśród oficjalnych podmiotów odpowiedzialnych za przekazywanie treści w infosferze, jak i wśród zwykłych użytkowników.

## Czym jest dezinformacja?

Dezinformacja może być dowolnym przekazem tekstowym lub audiowizualnym, który jest rozpowszechniany w sposób świadomy lub nieświadomy i oddziałuje negatywnie na odbiorców. Działania szkodliwe mogą wpływać na zmianę opinii, decyzji czy ugruntowanie określonego światopoglądu. Dezinformacja dotyczy także sfery wiedzy, postaw osób lub całych grup społecznych. Treści dezinformacyjne mogą doprowadzić do podjęcia określonych działań lub ich zaniechania.

**Dezinformacja** – celowe działanie, którego celem jest sfabrykowanie lub zaburzenie przekazu informacyjnego, by osiągnąć własne korzyści polityczne, społeczne, finansowe, militarne, etc. Efektem takiej narracji jest wprowadzenie w błąd drugiej strony (osoby/grupy/populacji), będącej odbiorcą przekazu dezinformacyjnego.

**Misinformation** (dezinformacja niezamierzona) – to w efekcie braku wiedzy, znajomości kontekstu itp. powielanie nieprawdziwych i niezweryfikowanych informacji przez użytkownika, który bezkrytycznie przekazuje zmanipulowane treści. Tego typu materiały są udostępniane bez intencji wyrządzenia krzywdy.

**Malinformation** (złośliwe lub szkodliwe wiadomości) – informacje prawdziwe, udostępniane z intencją wyrządzenia krzywdy. Przykładem takiej sytuacji było zhakowanie i upublicznienie przez Rosjan treści maili sztabu wyborczego Hillary Clinton w celu zaszkodzenia wizerunkowi kandydatki Partii Demokratycznej.

Dezinformacja **dotyczy faktów, ale także może wprowadzać w błąd za pomocą zmanipulowanych przekazów mieszanych zawierających zarówno opinie, jak fałszywe przedstawianie faktów**. Mogą to być wyrwane z kontekstu cytaty, pytania, opinie z wbudowaną tezą wprowadzającą w błąd co do stanu faktycznego, materiały wywołujące silną reakcję emocjonalną przez manipulowanie stanem faktycznym i wprowadzanie w błąd odbiorcy, co wpływa na zmiany zachowania lub poglądów u odbiorców.



## Jakimi kanałami rozpowszechniane są treści dezinformacyjne?

**Dezinformacja bywa błędnie utożsamiana jedynie z kanałami automatycznymi i mechanicznymi (boty/trolle). Kanałem rozpowszechniania treści dezinformacyjnych mogą być wszystkie nośniki informacji. Od ulotki czy billboardu po wirtualne spotkanie konferencyjne. Warto przy tym zaznaczyć, że działania dezinformacyjne często łączą pozornie niepowiązane ze sobą czynniki i kanały komunikacji.**

Cechą charakterystyczną treści dezinformacyjnych jest fakt, że są w stanie dostosować się do specyficznych elementów danej platformy oraz do charakterystycznego stylu danej grupy odbiorczej (przykładowo mogą wykorzystywać slang i popularne hashtagi). Szkodliwe treści rozpowszechniane są zarówno przez tzw. tradycyjne media (radio, prasa, telewizja), internet (szczególnie media społecznościowe), jak i kanały bezpośrednie (komunikatory, czaty czy SMSy). Działające w sieci podmioty mogą mieć postać zorganizowaną, ale także funkcjonować oddolnie, z własnej inicjatywy czy też braku określonej wiedzy.

Warto zaznaczyć, że działania tzw. botnetów (sieci botów) lub trolli są często jedynie działaniami wspierającymi dezinformację. Ich znaczenie bywa często wyolbrzymiane. W rzeczywistości dużo bardziej szkodliwe są treści wysokozasięgowe generowane przez konta uwiarygodnione. Działania botów mają przede wszystkim na celu sztuczne zwiększanie zasięgu określonych treści tak, aby wywołać iluzję popularności danej treści oraz dotarcie do potencjalnie nowych odbiorców.

**Różnorodność kanałów dystrybucji szkodliwych treści oraz ich pozorowane podobieństwo do działań zwykłych użytkowników sprawia, że kluczowe w kontekście działań dezinformacyjnych są propagowane treści, a nie kanały ich dystrybucji.**

## a. Trolle

„Troll” i „trollowanie” to jedne z najstarszych używanych i popularnych do dziś terminów żargonu sieciowego. W kontekście dezinformacji mówimy o kontach zamieszczających kontrowersyjne lub silnie zabarwione emocjonalnie treści w celu zwrócenia na siebie uwagi albo sprowokowania, ośmieszenia lub obrażenia innych użytkowników.

Troll to prawdziwy użytkownik działający w sieci, czasem na zlecenie innych podmiotów. Publikuje emocjonalne treści, powodujące podziały i wywołujące kontrowersje. Niekiedy działa w ramach większej struktury nazywanej farmą trolli. Sam trolling może być nośnikiem bardzo szkodliwych treści, nawet jeżeli publikacja dotyczy wpisów ironicznych lub prześmiewczych, a intencją nie jest zmiana czyjejs postawy lub opinii.

## b. Boty

**Bot to konto zautomatyzowane lub półzautomatyzowane (tzw. cyborgi lub minotaury), charakteryzujące się powtarzającymi się zachowaniami dotyczącymi interakcji z publikowanymi przez inne konta treściami. Boty często starają się naśladować zachowania zwykłych użytkowników danej platformy i mogą być wykorzystywane we wspieraniu akcji dezinformacyjnych.**

Boty to zautomatyzowane konta w mediach społecznościowych, których celem jest szybkie reagowanie na określone treści, tworzenie nowych, budowanie obrazu sztucznego zainteresowania danym tematem w celu poparcia lub sprzeciwu wobec określonej opinii czy stanowiska. Mogą istnieć zarówno boty osobowe i bezosobowe w ramach sieci, jak i w sposób całkowicie od siebie niezależny.

Często stosowaną techniką budowy tego typu kont jest udostępnienie treści angażujących (np. treść skrajnie emocjonalna, ankiety czy pytania otwarte wymagające komentarza), które prowokują użytkowników do działań zwiększających zasięgi publikacji.



## c. Liderzy opinii

**O tym, czy dany profil można nazwać liderem opinii, decyduje w internecie zasięg, czyli możliwości szerokiego rozpowszechniania i angażowania ludzi poprzez publikowane treści. Liderzy opinii to osoby o ugruntowanej oraz opiniotwórczej pozycji, które wpływają na poglądy użytkowników, realnie propagując określone tezy. Tego typu konta są skutecznym nośnikiem świadomej lub nieświadomej dezinformacji.**

Lider opinii to osoba/strona/grupa (itd.), która odpowiada za dużą część aktywności wokół danego tematu. Wśród liderów opinii możemy również spotkać osoby budujące swój autorytet na nieprawdziwej tożsamości. Aby określić lidera opinii, musimy jasno zdefiniować temat, w którym chcemy takiego lidera znaleźć. Może on być odpowiedzialny za rozpowszechnianie zarówno informacji pomocnych, jak i tych dezinformujących. Przykładowo, w temacie szczepień liderem opinii jest tak samo dr n. med. Paweł Grzesiowski, jak i pani Justyna Socha, chociaż reprezentują inne grupy. Identyfikacja liderów opinii pozwala na szybkie ustalenie głównych argumentów danej grupy, którzy skutecznie kształtują jej poglądy.

Liderzy opinii mogą bardzo szybko rozpowszechniać wysokozasięgowe treści - dlatego wokół tego typu kont można zaobserwować działania tzw. botów (konta automatyczne lub mechaniczne) oraz internautów, którzy powielają, reagują i wspierają treści lidera. Wierząc w jego autorytet, użytkownicy nie mają świadomości szkodliwości takiego działania.

## d. Politycy

**Politycy są specyficzną grupą liderów opinii. Treści pochodzące od polityków, szczególnie w wysoce spolaryzowanym społeczeństwie (takim jak w Polsce), są często wykorzystywane w akcjach dezinformacyjnych.**

Występują tutaj podobne mechanizmy, jak w przypadku zwykłych liderów opinii, jednak wywoływane przez takie wpisy reakcje charakteryzują się kontekstem bezpośrednio politycznym, co w konsekwencji ma dużo wyższy ładunek emocjonalny. Warto zaznaczyć, że pogłębianie polaryzacji sprzyja działaniom dezinformacyjnym. Użytkownicy w naturalny sposób podlegają radykalizacji lub dołączają do tzw. milczącej większości. Treści dezinformacyjne mają przez to zasadniczo prostszą

ścieżkę dotarcia do odbiorców. Odpowiedzialna postawa polityków jest niezbędna w zabezpieczeniu społeczeństwa przed działaniami dezinformacyjnymi.

## e. **Zwykli użytkownicy**

**Zwykli użytkownicy są bardzo często nośnikami treści dezinformacyjnych. Często komentują lub udostępniają treści bez ich uprzedniej weryfikacji. Padając ofiarą działań dezinformacyjnych, stają się również ich narzędziami, często nieświadomie rozprzestrzeniając nieprawdziwą narrację lub szkodliwą opinię.**

Podobnie jak inne treści dezinformacyjne mogą one być nastawione na zbudowanie w odbiorcach mylnego wrażenia opartego na dowodzie społecznej słuszności (liczba komentarzy o zbieżnej treści ma potwierdzać słuszność prezentowanych tez/opinii). Tym kanałem są też często wykorzystywane sparafrazowane i przygotowane wcześniej przekazy kolportowane mechanicznie.

## f. **Media**

**Media i dziennikarze są specjalnym zbiorem funkcjonującym w ramach grupy liderów opinii. W epoce wszechobecnego dostępu do źródeł informacji spoczywa na nich duża odpowiedzialność.**

Wiarygodność dziennikarzy - nawet jeżeli ograniczona do określonej grupy odbiorców - powoduje, że każda opublikowana przez nich treść jest przyjmowana przez odbiorców bezkrytycznie. Nieuważnie udostępniona treść, np. podsunęta przez bota, może być źródłem dla szybko rozprzestrzeniającej się szkodliwej narracji. W Internecie obowiązuje rozszerzona definicja mediów - to również konta dziennikarzy, blogi, mikroblogi, portale prowadzone w mediach społecznościowych.

Media powiązane z reżimami totalitarnymi lub autorytarnymi propagują dezinformację poprzez pielęgnowanie własnego przekazu. Media w państwach demokratycznych również sporadycznie przyczyniają się do dezinformacji, nie weryfikując w sposób wystarczający podawanych wiadomości, bądź też narzucając określoną interpretację danego zjawiska.

Istnieje także specyficzna grupa dziennikarzy/publicystów, których można zaklasyfikować na pograniczu trzech kategorii: liderów opinii, polityków i mediów, ponieważ promują oraz nagłaśniają opinie zgodne przede wszystkim z przekazem konkretnych partii politycznych, w tym treści o znamionach dezinformacyjnych wspierających narracje tych partii.

# Jakie są najpopularniejsze metody działań dezinformacyjnych w polskiej infosferze?

**Kampanie dezinformacyjne mają na celu m.in. polaryzację polskiego społeczeństwa poprzez narzucanie narracji wzbudzających poczucie zagrożenia, strachu, wzmacniające tezy antynaukowe czy propagowanie nienawiści wobec określonych grup społecznych, etnicznych, narodowościowych, wyznaniowych, grup związanych z orientacją seksualną i tożsamością płciową oraz osób niepełnosprawnych. Przykładem tego typu działań są agresywne komunikaty wobec ukraińskich uchodźców, rozbudzanie strachu przed zagrożeniem radiacyjnym czy świadome dyskredytowanie szczepionek przeciw COVID-19.**

Bardzo niebezpieczne są długotrwałe działania dezinformacyjne funkcjonujące w nawiązaniu do sytuacji, które tworzą wątpliwości mimo znanych faktów. Przykładem takiego oddziaływania może być awersja społeczeństwa wobec szczepionek w ogóle, więc i spadku odporności na drobnoustroje, które do tej pory były zneutralizowane przez masowe szczepienia. Warto podkreślić, że tego typu działania są często sprofilowane i dostosowane do docelowej grupy odbiorców. Z czasem zmieniają ich sentyment wobec określonych wydarzeń.

Dzisiejsza kultura informacyjna w połączeniu z propagandą doprowadziła do tego, że duża część odbiorców nie rozróżnia informacji od opinii. Podważają też jakiegokolwiek klasyczne kanały pozyskiwania informacji, co czyni ich podatnymi na metody dezinformacji.

## **Wybrane metody działań dezinformacyjnych:**

- Wykorzystywanie botów do podbijania zasięgów określonych narracji;
- Wykorzystywanie trolli do uwiarygodnienia treści i tworzenia nowych narracji;
- Manipulacja kontekstem (przede wszystkim za pomocą zdjęć i filmów);

- Nagłaśnianie wyrwanych z kontekstu cytatów pochodzących od ogólnie uznanych autorytetów, co ma uwiarygodnić nieprawdziwą tezę;
- Wyciąganie fałszywych wniosków na podstawie poprawnych danych;
- Polaryzacja społeczeństwa dzięki poruszaniu i antagonizowaniu tematów społecznie wrażliwych (osoby LGBT, uchodźcy itp.);
- Przedstawianie skrajnych i niepopularnych poglądów jako norma obowiązująca w społeczeństwie;
- Generalizowanie i uogólnianie;
- CheapFake (proste manipulowanie obrazem za pomocą tanich i powszechnie dostępnych narzędzi);
- DeepFake (zaawansowane metody obróbki obrazu wykorzystujące sztuczną inteligencję, np. poprzez nakładanie i animowanie czyjejs twarzy na inny materiał wizualny);
- Przedstawianie fałszywych informacji jako efektów niezależnego myślenia, nieulegania trendom itp.;
- Wykorzystanie/kreowanie fałszywych ekspertów/autorytetów;
- Osłabianie i podburzanie zaufania do instytucji państwowych i międzynarodowych organizacji zrzeszających ekspertów;
- Cherry picking - wykorzystanie wybiórczych danych, które potwierdzają pozycję lub tezę, ignorując pozostałe materiały badawcze, które temu przeczą;
- Postępowanie się dowodem anegdotycznym - publikowanie zanonimizowanych historii o podobnym schemacie, gdzie komuś wiarygodnemu (przyjacielowi, członkowi rodziny, osobie pełniącej jakąś funkcję) przypisuje się doświadczenie odnoszące się bezpośrednio do bieżącej sytuacji;

- Stawianie chochoła - atakowanie, wyśmiewanie lub podważanie poglądów; przeciwnika, których tak naprawdę nigdy nie wyraził;
- Wyśmiewanie/umniejszanie rangi problemu;
- Whataboutism ("ale co z?", "ale za to ty") - odwracanie uwagi od tematu dyskusji poprzez zarzucanie przeciwnikowi hipokryzji (ty zbierasz fundusze dla uchodźców z Ukrainy, ale co ze środkami na wsparcie ubogich rodzin w Polsce?);
- Promocja niskiej jakości blogów czy stron internetowych, często kreujących się na media lokalne, informacyjne lub branżowe;
- Propagowanie emocjonalnych opinii, pytań lub cytatów;
- Galop Gisha - publikowanie długich spisów rzekomych dowodów (m.in. badań zgromadzonych wg klucza cherry picking) na potwierdzenie danej tezy, czego celem jest przytłoczenie i zalanie przeciwnika argumentami, bez względu na ich rzetelność (metoda stosowana głównie w dyskursie nt. szczepień);
- Sea-lioning - ciągłe domaganie się od rozmówcy kolejnych wyjaśnień do wcześniejszych wypowiedzi, co ma na celu zmęczenie przeciwnika i rozmycie sedna dyskusji (metoda stosowana głównie w dyskursie nt. szczepień);
- Podawanie nieprawdziwych, dawno obalonych twierdzeń, które jednak nadal wywołują duże emocje.



## Jak monitorować treści dezinformacyjne?

Istnieje wiele metod monitorowania treści dezinformacyjnych opartych na zautomatyzowanych badaniach ilościowych, które stanowią podstawę do wyciągania wniosków o charakterze jakościowym.

Aktywność kont o charakterze mechanicznym czy automatycznym stanowi zestaw działań wspierających główne wątki narracyjne. Dlatego też wyciąganie zbyt daleko idących wniosków, jedynie na bazie tej aktywności, może prowadzić do błędnego określenia skali i efektów takiego działania.

### Przykład:

*Monitoring wykazał obecność wielu tysięcy wpisów krytykujących dietę wegetariańską, których zasięg wygenerował 2 mln odbiorców w mediach społecznościowych. Algorytm może wykazać, że 90% z tych treści cechuje się negatywnym wobec wegetarianizmu sentymentem.*

**Czy możemy już ogłosić, że przeprowadzono skuteczną kampanię dezinformacyjną wobec osób unikających spożywania mięsa? Odpowiedź brzmi: nie.**

Warto pogłębić analizę o szereg pytań, które pozwalają opracować pełniejszy obraz sytuacji:

- *Jakie zasięgi osiągnęły poszczególne treści we wskazanym zbiorze?*
- *Ile odnotowały interakcji?*
- *Jakie to były interakcje (komentarze, udostępnienia, like)?*
- *Jak zareagowali na te treści komentujący? W jakim okresie przeprowadzono akcję dezinformacyjną?*
- *Jakie konta brały w tym udział?*
- *Jakich słów używały?*
- *Jakiej metody działań użyto?*
- *Jak szacowany jest rozkład treści organicznych względem treści mechanicznych?*
- *Jaką metodę do tego przyjęto? Czy krytyka była uzasadniona?*

# Informacje i opinie w kontekście dezinformacji - jak je odróżnić?

**W kontekście działań dezinformacyjnych plotki, opinie i spekulacje często mieszają się z faktami. Powoduje to, że odpowiednie rozpoznawanie tych treści i ich klasyfikowanie jest bardzo ważną częścią przeciwdziałania zjawisku dezinformacji.**

Informacje bazują na faktach i są weryfikowane w sposób obiektywny. Opinie natomiast opierają się na odczuciach i refleksjach osoby, która je wygłasza. Nie można też dowiedzieć ich prawdziwości. Opinie bazują na emocjach i własnych doświadczeniach, co w skrajnych przypadkach przyczynia się do podziału na zwolenników lub przeciwników danego stanowiska. Natomiast fakt to stan prawdziwy i jako taki może być poparty dowodami. Często spotykamy się również ze spekulacjami, czyli treściami opartymi na błędnym wnioskowaniu o tym, co się stało lub ma się stać. W działaniach dezinformacyjnych często wykorzystywane są także plotki, czyli historie (lub ich fragmenty), które są na tyle niedookreślone co do czasu, miejsca i osób biorących udział w zdarzeniu, że stają się całkowicie nieweryfikowalne.

Opinie, spekulacje i zwykłe plotki są często wykorzystywane w działaniach dezinformacyjnych i w wielu przypadkach mogą doprowadzić do szkód zdrowotnych odniesionych w realnym życiu.

Wyrażanie własnej opinii jest kluczowym prawem każdego człowieka. Wykorzystywanie opinii, spekulacji i plotek w celu wprowadzania w błąd co do stanu faktycznego, powinno być wskazywane i zwalczane. Wskaźnikiem i sygnałem ostrzegawczym są silne, negatywne emocje, jakie wywołuje treść i jej nakierowanie na konkretne osoby lub grupy. Treści skutkujące okazywaniem agresji i nienawiści w relacjach społecznych powinny być sygnalizowane i w miarę możliwości, zgodnie z prawem i poszanowaniem konstytucyjnej wolności słowa, zwalczane.

## Jak informować o pojawieniu się treści dezinformacyjnych?

**Nieodpowiednia, nieprofesjonalna komunikacja dotycząca wykrytych treści dezinformacyjnych może przynieść więcej szkód niż pożytku. Kluczową kwestią jest forma prezentowanego przez nas komunikatu i zwracanie uwagi na detale.**

Należy unikać udostępnienia informacji o treściach dezinformacyjnych bez stosownego, wyraźnego oznaczenia wskazującego na szkodliwość pojawiającej się treści. Warto używać screenów zamiast aktywnych hiperłączy, które dodatkowo zwiększają zasięgi fałszywej narracji. Jeśli sytuacja i konstrukcja serwisu weryfikującego treści wymaga podawania linków, to należy je opatrywać atrybutem "nofollow" w kodzie, co stanowi informację wykluczającą dla robotów wyszukiwarkowych.

Każdorazowo należy podawać wiarygodne, rzetelne źródła, które wskazują fakty dotyczące treści. Jeżeli treść odnosi się do opinii będących manipulacjami wprowadzającymi w błąd co do stanu faktycznego, plotki lub spekulacji, warto szerzej wyjaśnić kontekst szkodliwej treści, mechanizm manipulacji stanem faktycznym, jaki w niej zachodzi i przedstawić potencjalne skutki jej rozpowszechniania. W każdym przypadku należy osadzić treść w kontekście informacyjnym.

Ważnym czynnikiem przy wykryciu treści dezinformacyjnej jest czas reakcji. Im szybciej podejmie się działanie, tym mniejsze szkody zostaną wyrządzone. Warto zgłaszać szkodliwe wpisy - tak do podmiotów zajmujących się weryfikacją treści w Internecie, jak i samych portali społecznościowych. Tego typu działania mogą przynieść realny skutek i prowadzić do usunięcia części szkodliwych treści, choć należy podkreślić, że w przypadku portali społecznościowych wciąż nie są to wystarczające reakcje. Należy więc wywierać nacisk, by skuteczniej blokowały szkodliwe treści. Dostępne są publiczne listy zgłoszonych kont, które rozprzestrzeniały treści o charakterze dezinformacyjnym.

Ważnym aspektem są również proaktywne działania ostonowe - na bazie trendów w dyskusji możliwe jest zbudowanie scenariuszy rozwoju narracji szkodliwych i wcześniejsze przygotowanie treści informacyjno-edukacyjnych.

Warto pamiętać, że często przeciwnie jest bezpośrednio odnoszenie się do treści niskozasięgowych, gdzie reakcja bardzo często powoduje wzmocnienie zasięgu szkodliwych treści.

## Jak rozpoznawać zjawisko dezinformacji?

**Jak już pokazywaliśmy wyżej - dezinformacja ma wiele twarzy. Najskuteczniejszą metodą na jej rozpoznanie jest zachowanie czujności i ustalenie możliwie największej liczby faktów.**

Działania dezinformacyjne są bardzo często nacechowane silnymi emocjami. Najbardziej szkodliwe treści są skonstruowane w sposób uniemożliwiający prostą weryfikację prezentowanych informacji. Ostrożność warto zachować również względem tych treści, które podburzają nas przeciwko np. innej grupie społecznej bądź próbują zaznajomić nas z "utajnionymi", "niepublikowanymi wcześniej" informacjami, których tylko my mamy być powiernikiem. Silnie emocjonalne opinie są bardzo często oparte na spekulacjach. Zamierzonym skutkiem działań dezinformacyjnych są wątpliwości co do stanu faktycznego.

Dezinformacja często opiera się na zmanipulowanych źródłach. Odpowiednio wykadrowane filmy, grafiki lub zrzuty ekranu, wyjęte z kontekstu cytaty mogą skutecznie manipulować odbiorcami. Umieszczane przy treściach dezinformujących odsyłacze do źródeł budzą zaufanie swoim profesjonalnym wyglądem lub domniemaną wiarygodnością, niekiedy o rzekomo naukowym charakterze. Nierzadko występuje powoływanie się na portale kreujące się na lokalne czy branżowe - to prosta metoda na uwiarygodnienie narracji dezinformacyjnych. Podobnie jest w przypadku odwołań do stron publikowanych w innym niż łaciński alfabetcie lub mało popularnym języku. Warto pamiętać, że nie każde pismo wyglądające na pierwszy rzut oka jak periodyk naukowy (istnieją również tzw. predatory journals, publikujące treści za opłatą).

Każdorazowo warto zweryfikować tego typu źródła poprzez poświęcenie chwili na pozyskanie informacji na temat źródła i zdarzenia.

# Jak przeciwdziałać zjawisku dezinformacji?

## ● Fact-checking

**Definicja:** Etapowa weryfikacja informacji polegająca na sprawdzeniu, czy wszystkie stwierdzenia zawarte w pracy pisemnej, artykule informacyjnym, materiałach graficznych/video lub przemówieniu są zgodne ze stanem faktycznym. Działanie ma na celu weryfikację lub falsyfikację prezentowanych informacji na drodze analizy i porównania z wiarygodnymi źródłami.

**Metody:** Porównywanie źródeł, metadanych, wyszukiwanie odwrócone (materiały audiowizualne), analiza propagacji treści w mediach społecznościowych; sprawdzanie treści, obrazu, domeny strony; weryfikacja prawdziwości źródła, autora, sięganie do pierwotnych źródeł. Proces powinien zakończyć się niezwłoczną informacją o wyniku weryfikacji wraz z oznaczeniem autora/źródła informacji.

## ● Edukacja użytkowników

**Definicja:** Edukacja jest kluczowym działaniem dla długofalowej walki z dezinformacją. Wpływa także na zwiększenie kompetencji społeczeństwa w zakresie umiejętności rozróżniania informacji fałszywych od prawdziwych oraz faktów od opinii, selekcji źródeł informacji, wiedzy na temat technik dezinformacji; umiejętności krytycznego myślenia, poddawania informacji analizie. Korzystanie z wielu źródeł jest jedynym sposobem na utrwalenie działań zwalczających dezinformację. Działania w tym zakresie powinny być przekrojowe: od dzieci w wieku szkolnym po osoby dorosłe i seniorów.

**Opis metody:** Kampanie informacyjne, działania NGO-sów i państwa, wprowadzenie treści na temat dezinformacji do szkół, miejsc pracy; szkolenia, warsztaty skierowane do grup najbardziej narażonych na dezinformację; współpraca z mediami. Działanie powinno zawierać moduły uświadamiające o istnieniu zjawiska celowej dezinformacji, uczące, jak rozpoznawać szkodliwe treści i jak się przed nimi chronić.

## ● Wysokie standardy działań medialnych

**Definicja:** Wysokie standardy działań medialnych to „dziennikarstwo u źródeł”, czyli podawanie wyważonych, rzetelnych, opierających się na faktach informacji. Konieczne jest wyraźne oddzielenie informacji od publicystyki, która z natury zawiera opinię autora.

**Opis metody:** Potwierdzanie informacji w przynajmniej dwóch lub trzech źródłach (crosschecking); sięganie do źródeł pierwotnych i jak najmniej przetworzonych (np. instytucje/podmioty/informator); **zasada pytania obu stron w kwestiach, gdzie znaczenie ma opinia - jednocześnie zdecydowane trzymanie się zasady prezentowania tylko faktów w kwestiach, w których ich prawdziwość nie budzi wątpliwości (np. kwestie związane z badaniami naukowymi);** stanowcza reakcja moderatora na nieprawdziwe tezy; powstrzymanie się od kierowania się wyłącznie priorytetem czasu i szybkości publikacji, a wiarygodności źródła; nieopieranie się na celu wzbudzania kontrowersji i emocji; niestosowanie clickbaitów wprowadzających w błąd już na etapie tytułu;

**Propozycja we wzorcowym wariancie:** W każdej redakcji zatrudniona jest osoba odpowiedzialna za fact-checking, która decyduje o publikacji w razie wątpliwości, edukacji w kwestii dezinformacji, która odpowiada za kontakt z NGO'sami i instytucjami walczącymi z dezinformacją; każdy stażysta i początkujący członek redakcji zostaje odpowiednio przeszkolony z weryfikacji informacji. Koniecznym komponentem wydaje się być również edukacja w zakresie ryzyka wynikającego z przekazywania nieprawdziwych informacji.

## ● Zgłaszanie/blokowanie treści/trust flagging kont/treści dezinformacyjnych

**Definicja:** Zgłaszanie treści i kont propagujących fałszywe treści do platform internetowych, na których działają oraz, jeśli to potrzebne, do właściwych organów i organizacji zajmujących się zwalczaniem dezinformacji.

**Opis metody:** Zgłoszenia poprzez stosowne procedury na platformach, kontakt z instytucjami zajmującymi się walką z dezinformacją.



## ● Rola platform społecznościowych

**Stan obecny:** Platformy podejmują pewne kroki, ale nie są one wystarczające. Oprócz realnych programów, takich jak Niezależny Program Weryfikacji Informacji Facebooka przy współpracy z fact-checkerami zrzeszonymi w IFCN, trudno znaleźć inne, skuteczne działania mające na celu ograniczenie szerzenia się treści dezinformujących. Potrzebne są regulacje na poziomie legislacyjnym, presja społeczna i transparentność platform na poziomie działania algorytmów. Media społecznościowe deklarują, że podjęły kroki, by walczyć z dezinformacją. Jednak od początku inwazji na Ukrainę reagowały zbyt późno – przykładowo, dopiero po kilku dniach zablokowały dostęp do propagandowych mediów państwowych, np. RT czy Sputnik, czy też zaczęły oznaczać, że informacja ta pochodzi z oficjalnego konta rosyjskiego rządu/medium (np. Twitter). Działania podjęte przez właścicieli platform w naszej ocenie są niewystarczające – wiele analiz wskazywało, że np. Facebook promuje nienawistne komentarze, które polaryzują społeczeństwo, a fałszywe wpisy są promowane jako wzbudzające emocje. Na popularnym TikToku algorytm sam podsuwa treści dezinformacyjne i fake newsy w propozycjach „dla Ciebie” na równi z rzetelnymi informacjami bez wskazywania, jakie wideo są fałszywe, a jakie nie. Jest to bardzo często forma modelu biznesowego dla platformy społecznościowej, która zyskuje na rozpowszechnianych nieprawdziwych informacjach, które są wysokozasięgowe.

### **Każda z platform społecznościowych powinna:**

- Bezwzględnie oznaczać źródło informacji, np. treść pochodzi z oficjalnej strony państwa X;
- Umożliwić zgłaszanie podejrzanych treści przez użytkowników, co do których mają wątpliwości, bezpośrednio do zespołu fact-checkerów;
- Ścisłe współpracować z zewnętrznymi fact-checkerami z wielu państw (niezatrudnianie jedynie fact-checkerów anglojęzycznych);
- Nie pozostawiać zgłaszanych treści bez odpowiedzi - użytkownik otrzymuje odpowiedź, czy dana informacja/komentarz/opinia zawierała fake news;

- Przejrzyście informować co do: liczebności zespołów specjalistów odpowiedzialnych za walkę z fake newsami, funduszy przeznaczanych na walkę z dezinformacją/edukacją użytkowników;
- Rzetelnie informować o konkretnych powodach usunięcia/ograniczenia treści;
- Wdrożyć faktyczną transparentność o tym, w jaki sposób dezinformacja krąży na platformie oraz publicznie przedstawiać planowane działania w walce z fake newsami;
- Opublikować pełną politykę moderacji dezinformacji, w tym stosowania sztucznej inteligencji oraz wykorzystywanych przez nią danych;
- Poza usuwaniem treści, w trosce o zgodność z prawem, platformy powinny koncentrować się na dostarczaniu szerokiego kontekstu i demaskowaniu przekłamań – w formie dodatkowej informacji wyraźnie nakładanej na treści;
- Walczyć z recydywistami, którzy tworzą treści regularnie oceniane jako dezinformacja (w szczególności z tymi, którzy zarabiają na tych treściach na platformie i poza nią). Platformy powinny uniemożliwić algorytmom rekomendacje i promowanie treści z takich źródeł;
- Zwiększyć aktualne i przyszłe wysiłki przeciwko dezinformacji w językach innych niż angielski. Platformy muszą dostarczać dane o poszczególnych krajach i językach, zapewnić usługę transkrypcji dla każdego języka na platformie.

Ponadto powyższe zalecenia nie powinny być traktowane jako „inicjatywa platform”, a jako obligatoryjne działanie. Algorytmy platform nie powinny promować nienawistnych, polaryzujących treści bądź komunikatów, co do których istnieją wątpliwości w zakresie ich prawdziwości.

