

## Jakie działania w sprawie systemów rekomendacyjnych wielkich platform podejmuje Komisja Europejska?

Na podstawie aktu o usługach cyfrowych (Digital Services Act, DSA) Komisja Europejska od 2023 r. prowadzi **postępowania przeciwko wielkim platformom internetowym, badając m.in. ich uzależniające funkcjonalności i potencjalnie szkodliwy wpływ algorytmów rekomendujących treści**. Między innymi:

- w postępowaniu wszczętym w lutym 2024 r. przeciwko TikTokowi Komisja bada wykorzystywanie manipulacyjnych, uzależniających interfejsów (systemów rekomendacyjnych, mechanizmów powodujących wpędzanie osoby w tzw. królicze nory, czyli pętle monotematycznych treści, stymulowanie uzależnień behawioralnych, problem radykalizacji, w szczególności osób małoletnich);
- w postępowaniu wszczętym w grudniu 2024 r. przeciwko TikTokowi Komisja bada, czy platforma właściwie zarządziła ryzykiem związanym z dyskursem obywatelskim i procesem wyborczym w trakcie ostatnich wyborów w Rumunii, zwłaszcza jeśli chodzi o systemy rekomendacyjne TikToka i ryzyko skoordynowanej akcji manipulacyjno-dezinformacyjnej lub zautomatyzowanego wykorzystania platformy; według ujawnionych raportów służb, działania te były inspirowane przez Rosję i miały na celu zmanipulowanie debaty w mediach społecznościowych tak, żeby wypromować niezależnego kandydata o poglądach nacjonalistycznych, Călina Georgescu – omijając w ten sposób prawo chroniące w Rumunii wolność wyborów;
- w postępowaniu wszczętym w maju 2024 r. przeciwko Mecie Komisja bada, w jaki sposób projektowanie interfejsów Facebooka i Instagrama może eksploatować słabości i brak doświadczenia małoletnich użytkowników, a także powodować uzależnienia behawioralne i wciągać użytkowników w tzw. królicze nory; Komisja sprawdza też, czy Meta stosuje wystarczające środki zapobiegania ryzyku dla prywatności i bezpieczeństwa małoletnich użytkowników, zwłaszcza w zakresie domyślnych ustawień.

2 października 2024 r. Komisja Europejska **wezwała YouTube'a, Snapchata i TikToka do złożenia wyjaśnień dotyczących sposobu zaprojektowania i funkcjonowania ich systemów rekomendacyjnych**. Komisja domaga się informacji o parametrach, według których algorytmy tych platform „rekomendują” użytkownikom treści, a także jaka jest ich rola w zwiększaniu systemowych ryzyk, m.in. związanych z procesami wyborczymi i dyskursem obywatelskim, zdrowiem psychicznym użytkowników (w tym zachowań prowadzących do

uzależnienia i efektu króliczych nor) czy ochroną małoletnich”. Komisja chce też dowiedzieć się, w jaki sposób platformy tym ryzykom przeciwdziałają.

Jeśli platformy nie dostarczą tych informacji i wyjaśnień, Komisja będzie mogła wszcząć formalne śledztwo (na podstawie art. 66 DSA), w wyniku którego na platformy mogą zostać nałożone kary za złożenie nieprawdziwych, niepełnych lub wprowadzających błąd informacji (na podstawie art. 74(2) DSA).

W tej kadencji komisarze zapowiadają nie tylko twarde egzekwowanie aktu o usługach cyfrowych, ale też przyjęcie **nowych przepisów chroniących konsumentów**, ze szczególnym uwzględnieniem sytuacji dzieci w sieci (Digital Fairness Act, DFA). Komisja Europejska prowadzi też **zaawansowane prace w grupie roboczej z przedstawicielami państw członkowskich nad wprowadzeniem europejskiego standardu weryfikacji wieku dla usług internetowych**.

W pierwszym kwartale 2025 r. mają się pojawić **wytyczne Komisji dla wielkich platform w zakresie ochrony małoletnich** (doprecyzowanie standardu wynikającego z DSA).

Spodziewamy się, że jedną z rekomendowanych praktyk będzie weryfikacja wieku użytkowników za pomocą narzędzi wypracowanych przez instytucje UE w porozumieniu z państwami członkowskimi.

## Co w sprawie wielkich platform może zrobić polski rząd?

Na gruncie przepisów UE polski rząd i polskie instytucje publiczne mają dość ograniczone kompetencje, jeśli chodzi o regulowanie działalności wielkich platform internetowych. Co do zasady, **Polska nie może tworzyć przepisów, które nakładałyby nowe obowiązki na wielkie platformy, ponieważ jest to sfera kompetencji UE**. Wiodącą rolę w egzekwowaniu prawa, które już reguluje działanie wielkich platform (DSA i DMA), ma Komisja Europejska. Nadal jednak polski rząd i instytucje publiczne (UODO, RPO, RPD, UOKiK, UKE) mogą być ważnymi sojusznikami na tym polu.

Punktami zaczepienia do rozmowy o społecznym wpływie mediów społecznościowych w Polsce mogą być:

- prace nad ustawą o ochronie małoletnich przed treściami szkodliwymi w Internecie, która ma wprowadzić obowiązkową weryfikację wieku dla platform (w tym wielkich platform internetowych – VLOP-ów) serwujących treści pornograficzne, ale nie wprowadzi nowych obowiązków dla pozostałych platform (takich jak TikTok czy YouTube) [[komunikat o założeniach projektu ustawy](#)];

- prace nad strategią cyfryzacji do 2035 r., której projekt przewiduje m.in. kampanie edukacyjne w zakresie higieny cyfrowej i środki publiczne na badanie oddziaływania mediów społecznościowych na zdrowie i dobrostan ludzi w Polsce;
- wdrażanie nowych procedur, jakie przewiduje akt o usługach cyfrowych (DSA), w tym indywidualnych skarg na działania/zaniechania wielkich platform i wniosków zweryfikowanych badaczy o udostępnienie danych;
- prace nad reformą systemu edukacji, w ramach której rząd będzie musiał zająć stanowisko w sprawie zasad korzystania ze smartfonów w szkołach.

## Czego dowiedzieliśmy się z ocen ryzyka opublikowanych przez wielkie platformy?

W listopadzie 2024 r. wielkie platformy internetowe (Very Large Online Platforms, VLOP-y) opublikowały długo oczekiwane raporty z wyników oceny ryzyka. Dowiadujemy się z nich, jak same firmy oceniają wpływ „sposobu projektowania lub funkcjonowania” ich usług (w tym systemów algorytmicznych) na m.in. prawa podstawowe ludzi, fizyczny i psychiczny dobrostan, ochronę zdrowia publicznego i małoletnich oraz dyskurs obywatelski i procesy wyborcze. Konieczność przeprowadzania (cyklicznie!) ocen ryzyka wynika z aktu o usługach cyfrowych (DSA), który w stosunku do VLOP-ów obowiązuje od sierpnia 2023 r.

**Wszystkie platformy przyznają, że ich systemy rekomendacyjne są źródłem systemowych ryzyk (m.in. dla praw podstawowych, praw dzieci, zdrowia psychicznego, debaty publicznej i procesów wyborczych), ale w swoich sprawozdaniach kładą zdecydowany nacisk na ryzyka związane z funkcjonowaniem tych systemów (a więc działaniami samych użytkowników i użytkowniczek i możliwymi nadużyciami), pomijając ryzyka wynikające z tego, jak te systemy zostały zaprojektowane.**

Oto kilka przykładów:

„Systemy rekomendacyjne to kluczowa funkcja naszej platformy i mamy szeroki wachlarz narzędzi do zarządzania nimi, ale niektóre potencjalnie problematyczne treści, które nie naruszają naszych wytycznych, mogą zostać zarekomendowane, zanim je wykryjemy i usuniemy, a niektóre naruszające zasady konta mogą uniknąć wykrycia przez filtry i zostać zarekomendowane, zanim je wykryjemy i usuniemy, podobnie niektóre konta łamiące regulamin mogą przemknąć przez nasze filtry i zostać zarekomendowane. Ten czynnik może mieć wpływ na wszystkie obszary systemowego ryzyka”. (Meta)

„Nasze systemy rekomendacyjne zaprojektowane są w taki sposób, żeby zapobiegać rekomendowaniu, redystrybuowaniu i wzmacnianiu zasięgów treści potencjalnie naruszających regulamin i w innych sposób problematycznych”. (Meta)

„Nieodłącznym wyzwaniem dla każdego systemu rekomendacyjnego jest zapewnienie, żeby zakres wyświetlanych treści był zróżnicowany, a nie zbyt wąski lub powtarzalny. TikTok stosuje środki zapobiegające ryzyku w celu dywersyfikacji treści, aby młodszy użytkownicy i użytkowniczki nie byli narażeni na powtarzające się treści, co jest szczególnie ważne, jeśli odkrywają treści związane z bardziej złożonymi tematami, ale które nie naruszają regulaminu lub Wytycznych dla społeczności TikToka“.

„Osoby małoletnie mogą uzyskiwać dostęp lub przeglądać na Platformie treści nieodpowiednie dla ich wieku. Związane z tym ryzyko obejmuje m.in. negatywny wpływ na samopoczucie fizyczne lub psychiczne”. (TikTok)

Na liście ryzyk zidentyfikowanych przez Google dla platformy YouTube znalazły się:

- ryzyko, że prezentacja treści na platformie internetowej negatywnie wpływa na różnorodność głosów, opinii i perspektyw w mediach,
- ryzyko, że prezentacja treści na platformie internetowej promuje polaryzację lub silosową segmentację głosów, opinii i perspektyw w mediach,
- ryzyko, że prezentacja treści na platformie internetowej nie sprzyja treściom pochodzącym od grup historycznie niedoreprezentowanych w mediach.

Nadal jednak jako główne przyczyny systemowych ryzyk platformy wskazują (nieodpowiedzialne) zachowania użytkowników i użytkowniczek (np. publikujących szkodliwe treści) i nadużywanie platformy przez podmioty zewnętrzne (np. do świadomego rozpowszechniania dezinformacji).

**Skupiając się na przeciwdziałaniu dystrybucji treści niezgodnych z regulaminem, autorzy raportów pomijają ryzyka związane z hiperpersonalizacją rekomendowanych treści,** które – chociaż nie są groźne (szkodliwe) same w sobie – stają się takie, gdy w nadmiernej ilości zalewają osoby wrażliwe (np. przeżywające lęki lub frustracje) na jakimś punkcie (np. zdrowia, bezpieczeństwa, wyglądu).

**Żadna z platform nie przyznaje, że sama zasada działania systemów rekomendacyjnych (tj. premiowanie treści mocno angażujących, dopasowanych do profilu behawioralnego użytkownika) jest źródłem systemowych ryzyk (np. może zwiększać ekspozycję**

użytkowników i użytkowniczek na problematyczne treści; może przyczyniać się do problemów psychicznych i uzależnień behawioralnych).

Google, w raporcie opisującym różne serwisy (w tym YouTube), przyznaje co prawda, że „źle zaprojektowane” lub „słabo kontrolowane” systemy rekomendacyjne mogą zwiększać widoczność szkodliwych treści, ale wyraźnie odcina się od tego problemu, tak jakby nie dotyczył tej platformy:

„Nasze algorytmy sortują setki miliardów treści, aby znaleźć najbardziej trafne i przydatne wyniki. **Algorytmy pozwalają nam podnosić jakość i trafność, jednocześnie zmniejszając ryzyko systemowe dla użytkowników/-czek i społeczeństwa. (...) Źle zaprojektowane lub kontrolowane systemy rekomendacyjne mogą zwiększać ryzyko, że szkodliwe treści uzyskają popularność.** Prawidłowo działające systemy rekomendacyjne powinny jednak zmniejszać to ryzyko poprzez zwiększanie widoczności treści rzetelnych i jakościowych oraz poprzez promowanie różnorodnej tematyki i zdywersyfikowanych źródeł, z którymi użytkownicy i użytkowniczki mogą się zapoznać. (...) [N]iektóre rodzaje ryzyka są łagodzone poprzez zapobieganie, usuwanie lub zwiększanie widoczności niektórych rodzajów treści (tj. moderację treści), a **inne są łagodzone poprzez zmiany w projektowaniu lub funkcjonowaniu usługi** lub sposobu interakcji użytkowników i użytkowniczek z usługą (tj. projektowanie usług)”.  
[podkreślenia nasze]

Problem systemów rekomendacyjnych, które zwiększają widoczność szkodliwych/problematycznych treści, pojawia się za to między wierszami, w opisach konkretnych strategii ograniczania ryzyka:

„Rutynowo ewaluujemy, czy sygnały, na których bazują rekomendacje treści odpowiednich dla użytkowników/-czek, mogą prowadzić do ekspozycji na problematyczne treści. Redukujemy to ryzyko poprzez limitowanie znaczenia udostępnień i komentarzy w dystrybucji wrażliwych tematów”. (Meta)

„Decyzja, jakie treści zostaną polecane, determinowana jest przez czynniki takie, jak: polubienia, udostępnienia, komentarze, wyszukiwania, różnorodność treści i popularne filmy. TikTok stosuje standardy kwalifikowalności treści dla FYF [Kanał dla Ciebie, For You Feed], które priorytetowo traktują bezpieczeństwo i są oparte na różnorodności społeczności TikTok oraz normach kulturowych. TikTok uniemożliwia umieszczenie niektórych treści w FYF, ponieważ mogą nie być odpowiednie dla szerokiego grona odbiorców. Utrudnia także wyszukiwanie takich treści”.

„TikTok proaktywnie wykrywa szkodliwe treści z pomocą narzędzi automatycznych i ludzkich moderatorów, zmniejszając w ten sposób ryzyko, że treści nieodpowiednie dla wieku będą rekomendowane lub pojawią się w FYF”.

„TikTok ręcznie przegląda treści wideo, które osiągnęły określony poziom popularności pod względem liczby wyświetleń, zmniejszając w ten sposób ryzyko wyświetlania w FYF lub innego sposobu szerokiej dystrybucji treści naruszających prawo”.

„Korzystanie z systemów rekomendacyjnych w celu uporządkowania prezentowanych treści, w tym poprzez podbijanie treści rzetelnych i jakościowych, jest często bardziej proporcjonalnym podejściem do rozwiązywania problemu szkodliwych treści niż całkowite ich usuwanie, które mogłoby stanowić zagrożenie dla wolności wypowiedzi i informacji”. (YouTube)

„Gdy dochodzi do ważnych zdarzeń, rzetelne źródła informacji potrzebują czasu na sprawdzenie faktów, dlatego wyprzedzają je spekulacje i dezinformacja. *Złe podmioty* mogą publikować treści na bazie niezweryfikowanych informacji, żeby wprowadzać w błąd lub przyciągnąć uwagę i zwiększyć ruch. Aby chronić się przed tym ryzykiem, YouTube i wyszukiwarka są zaopatrzone w zautomatyzowane systemy zaprojektowane do promowania wiarygodnych treści”. (YouTube)

W tabeli poniżej zebraliśmy **przykłady najciekawszych narzędzi i procedur, jakie wielkie platformy wprowadziły, żeby ograniczać szkodliwe skutki działania ich systemów rekomendacyjnych**. Na papierze te działania wyglądają całkiem dobrze, choć **na podstawie samych raportów nie jesteśmy w stanie ocenić ani ich skuteczności, ani skali zastosowania**. Potrzebujemy więcej danych pochodzących bezpośrednio od platform albo z niezależnych badań. W prawej kolumnie podajemy też przykłady dalej idących środków zarządzania ryzykiem, jakich oczekujemy od wielkich platform.

Samodzielne kontrolowanie zawartości feedu	
<p>Meta daje swoim użytkownikom i użytkownicy:</p> <ul style="list-style-type: none"> <li>• możliwość kontrolowania poziomu wyświetlania w swoim feedzie treści</li> </ul>	<p>Oczekujemy, że wielkie platformy wykażą, że opisywane w ich raportach funkcjonalności rzeczywiście działają i (według samych</p>



<p>„niskiej jakości”, „nieoryginalnych i problematycznych” oraz „kontrowersyjnych”. Użytkownicy i użytkowniczki mogą zaznaczyć, czy chcą widzieć ich więcej, czy mniej (albo pozostać przy domyślnych ustawieniach poziomu ich wyświetlania);</p> <ul style="list-style-type: none"><li>• przyciski „Ukryj post i pokazuj mniej takich treści” oraz „Nie interesuje mnie”, które powinny spowodować, że „podobne” treści będą rzadziej pojawiały się w feedzie, a także przycisk „Interesuje mnie”, która ma spowodować, że zobaczymy więcej „podobnych” materiałów.</li></ul> <p>TikTok daje swoim użytkownikom i użytkowniczkom:</p> <ul style="list-style-type: none"><li>• możliwość zablokowania wyświetlania filmów według określonych słów kluczowych;</li><li>• możliwość zresetowania historii feedu, żeby móc korzystać z aplikacji „z czystą kartą”, bez obciążającego profilu;</li><li>• możliwość sprawdzenia, dlaczego algorytm pokazał użytkownikowi/-czce określony film;</li><li>• przycisk „nie interesuje mnie”, który powinien spowodować, że „podobne” treści będą się rzadziej pojawiały w feedzie.</li></ul>	<p>użytkowników/użytkowniczek) dają pożądaną efekt.</p> <p>Oczekujemy też dalej idących rozwiązań takich jak:</p> <ul style="list-style-type: none"><li>• suwaki przeznaczone do ustawiania preferencji doboru treści („pokaż mi więcej informacji” albo „pokaż mi więcej rozrywki”);</li><li>• przycisk twardej blokady, zatrzymujący przyszłe polecenia treści związanej z konkretnym tematem, hasztagiem albo użytkownikiem (zgodnie z tym sygnałem algorytm doboru treści powinien odfiltrowywać treść na tematy oflagowane przez użytkownika jako niepożądane);</li><li>• przycisk resetowania feedu (wszystkie informacje wykorzystywane do personalizacji feedu zostają usunięte) na wszystkich platformach;</li><li>• funkcje pozwalające użytkownikom/-czkom na sprawdzenie, w jaki sposób ich informacja zwrotna (np. sygnał „pokazuj mniej”) wpłynęła na dobór rekomendowanych treści;</li><li>• częste zachęty do określania rzeczywistych zainteresowań i preferencji co do treści, które użytkownicy chcą (albo których nie chcą) oglądać;</li></ul>
---	--

	<ul style="list-style-type: none"> <li>• zachęty do sprawdzenia, jakie czynniki (sygnały, dane) wpływają na ranking rekomendowanych treści oraz możliwość ich zmiany.</li> </ul>
<h2>Nieprofilowany feed jako alternatywa</h2>	
<p>Meta deklaruje, że „użytkownicy i użytkowniczki mogą skorzystać z łatwo dostępnych opcji nieprofilujących. Mogą zdecydować, że nie chcą używać systemów rekomendacyjnych opartych na wnioskowaniu, którego systemy Mety dokonują na podstawie danych użytkownika/-czki”.</p> <p>TikTok również oferuje nieprofilowany feed jako alternatywę, ale nie poświęca temu rozwiązaniu miejsca w raporcie. Najwyraźniej nie traktuje tego rozwiązania jako środka zarządzania ryzykiem.</p> <p>YouTube nie oferuje możliwości łatwego przełączenia się na nieprofilowany feed. Nieprofilowany feed można osiągnąć wyłączając zapisywanie historii oglądania i kasując wcześniejszą historię (a więc pozbawiając system rekomendacyjny danych behawioralnych).</p>	<p>Nie oczekujemy, że wielkie platformy zaczną serwować swoim użytkownikom/-czkom nieprofilowany (układany chronologicznie albo według aktualnych trendów) feed w opcji domyślnej.</p> <p>Oczekujemy natomiast, że w opcji domyślnej wielkie platformy zrezygnują z wykorzystywania danych behawioralnych do personalizowania rekomendacji (a więc będą korzystały jedynie z danych dobrowolnie i świadomie przekazywanych przez użytkowników i użytkowniczki w celu personalizowania rekomendacji).</p>



## Zachęty i ostrzeżenia sprzyjające refleksji użytkowników/użytkowniczek

TikTok deklaruje, że oznacza i maskuje niektóre filmy, dając użytkownikom i użytkowniczkom możliwość zdecydowania, czy chcą je obejrzeć – zanim te treści się wyświetlą.

Do treści, których partner odpowiedzialny za fact-checking nie jest w stanie zweryfikować (zakwalifikować jako dezinformację wyborczą lub wykluczyć, że nią nie są), TikTok dodaje baner zachęcający do zapoznania się z faktami (Know the Facts).

TikTok zachęca osoby, które chcą udostępnić na platformie niezweryfikowane treści polityczne, do przemyślenia tej decyzji.

TikTok wprowadził zachętę „zastanów się zanim udostępnisz”, która przypomina użytkownikom i użytkowniczkom o regulaminie serwisu i daje szansę edycji komentarza przed opublikowaniem.

TikTok deklaruje, że w kontekście ważnych wydarzeń (np. wyborów) odsyła użytkowników i użytkowniczki szukające informacji do sprawdzonych źródeł (np. przed wyborami w Hiszpanii, w ramach aplikacji działało centrum informacji o wyborach zasilane z rzetelnych źródeł). Nie wiemy jednak, dla jakich innych tematów i na jaką skalę TikTok podejmuje takie działania.

Dalej idące rozwiązania, jakich oczekujemy od wielkich platform:

- dodatkowy krok wymuszający refleksję przed podaniem dalej („pomyśl, zanim podasz dalej”) dla wszystkich treści;
- zachęty dla użytkowników i użytkowniczek do ocenienia wiarygodności treści przed podaniem jej dalej;
- ustalenie limitu możliwości podania dalej tego samego postu, żeby spowolnić tzw. wirale; po osiągnięciu limitu, dalsze rozpowszechnianie treści wymagałoby stworzenia nowego postu;
- zachęty dla użytkowników i użytkowniczek do skorzystania z wyszukiwarki treści po przekroczeniu ustalonego (przez platformę) limitu interakcji z treściami rekomendowanymi („Czy na pewno chcesz dalej to oglądać?”), ew. blokowanie rekomendowanego feedu w ramach danej sesji („Zobaczyłeś/aś już wszystkie rekomendowane treści dla twojego zapytania. Jeśli chcesz zobaczyć nowe

<p>YouTube zapewnia, że „poza zwiększaniem widoczności jakościowych informacji i zwalczaniem dezinformacji udostępnia użytkownikom i użytkowniczkom narzędzia przydatne w ewaluowaniu informacji”. Niestety, w raporcie brakuje konkretów. Nie wiemy, ani jak to narzędzie działa, ani na jaką skalę i dla jakich tematów zostało wprowadzone.</p>	<p>rekomendacje, wpisz nowe zapytanie”);</p> <ul style="list-style-type: none"> <li>• promowanie treści pochodzących ze sprawdzonych (wiarygodnych) źródeł w odpowiedzi na wszystkie zapytania użytkowników i użytkowniczek dotyczące zdrowia, finansów, bezpieczeństwa i tematów politycznych.</li> </ul>
<h3><b>Systemowe zabezpieczenia i ustawienia domyślne</b></h3>	
<p>Dla tematów uznanych przez platformę za wrażliwe (w tym dotyczących odchudzania lub ekstremalnej sprawności, uproszonych informacji dotyczących zdrowia psychicznego i treści nacechowanych seksualnie), TikTok wprowadził politykę dywersyfikacji wyświetlanych treści. W ten sposób pomaga użytkownikom/użytkowniczkom w wychodzeniu z tzw. króliczych nor (pętli monotematycznych, silnie angażujących treści). Niestety raport nie wchodzi w szczegóły tego rozwiązania.</p> <p>Na TikToku treści, które platforma uznaje za „nieodpowiednie dla szerokiej publiczności” (związane m.in. z zaburzeniami odżywiania, stereotypami kulturowymi, materiałami dotyczącymi seksualności, materiały, których prawdziwości nie jest w stanie zweryfikować, zachęcające do niebezpiecznych zachowań czy treści publikowane przez użytkowników i użytkowniczki poniżej 16 roku życia) nie są</p>	<p>Dalej idące rozwiązania, jakich oczekujemy od wielkich platform:</p> <ul style="list-style-type: none"> <li>• rezygnacja z hiper-personalizacji rekomendowanych treści w wersji domyślnej (system nie wykorzystuje danych behawioralnych);</li> <li>• rezygnacja z sygnałów skorelowanych z krótkoterminowym zaangażowaniem użytkowników i użytkowniczek;</li> <li>• priorytetyzowanie sygnałów skorelowanych z trafnością i wiarygodnością rekomendowanych treści,</li> <li>• wprowadzenie różnorodności jako jednej z miar sukcesu dla systemów rekomendacyjnych, tak by</li> </ul>

rekomendowane w ramach Kanalu Dla Ciebie (choć są nadal możliwe do wyszukania na platformie przez zdeterminowanych użytkowników i użytkowniczki, o ile mają oni powyżej 18 lat).

TikTok klasyfikuje treści w oparciu o dojrzałość, jakiej dana treść wymaga od odbiorcy (Content Levels) i nie daje młodszym użytkownikom/użytkowniczkom dostępu do treści wymagających większej dojrzałości. Chodzi o treści, które nie naruszają regulaminu platformy ani przepisów prawa, ale mogą być nieodpowiednie dla młodszych użytkowników/użytkowniczek (przedstawiające np. operacje plastyczne, aktywności, które zachęcają do naśladowania, a mogą prowadzić do uszkodzeń ciała, znaczne odsłanianie ciała osób dorosłych, ich uwodzicielskie zachowania lub aluzje do aktywności seksualnej, krew ludzką lub zwierzęcą, spożywanie nadmiernych ilości alkoholu i konsumpcję tytoniu).

Na TikToku treści publikowane przez użytkowników lub użytkowniczki poniżej 16 roku życia w ogóle nie trafiają do rekomendowanych (Kanał dla Ciebie). A posty udostępnione przez użytkowników i użytkowniczki w wieku 13-15 lat mają ograniczoną możliwość komentowania.

TikTok wykrywa treści, które mogą być dezinformacją wyborczą. W przypadku gdy nie mogą one zostać zweryfikowane przez fact-checkerów, algorytm oznacza je jako „niezweryfikowane”. Tak oznaczone treści nie są rekomendowane w ramach Kanalu Dla Ciebie. Możliwość ich wyszukiwania również

personalizowany feed zawierał treści na różne tematy/pochodzące z różnych źródeł (np. jedna na dziesięć rekomendacji mogłaby dotyczyć innej tematyki wybranej spośród tego, co użytkownik/czka określił/a jako interesujące),

- umożliwienie użytkownikom/-czkom komfortowego korzystania z zewnętrznych dostawców algorytmów rekomendujących treści.

jest ograniczona (niestety, raport nie precyzuje jak).

W raporcie Google czytamy, że systemy rekomendacyjne „wykorzystują sygnały takie jak znaczenie, trafność, jakość, użyteczność i kontekst, aby wyżej wyświetlić wyniki, które wydają się najbardziej pomocne, w szczególności treści, które wydają się bazować na specjalistycznej wiedzy, doświadczeniu, autorytatywności i wiarygodności. Sygnały te są szczególnie ważne w przypadku tematów „Twoje pieniądze lub Twoje życie” (YMYL), zdefiniowanych jako te, które „mogą znacząco wpływać na zdrowie, stabilność finansową lub bezpieczeństwo poszczególnych osób lub dobrostan społeczeństwa”.

I dalej: „Jakościowe informacje z wiarygodnych źródeł wyświetlamy wyżej w wynikach wyszukiwania, rekomendacjach i panelach informacyjnych, pomagając w ten sposób ludziom znaleźć dokładne i przydatne informacje (...) W przypadku zapytań w serwisie YouTube dotyczących tematów podatnych na dezinformację, takich jak COVID-19 czy zmiany klimatyczne, wyświetlamy panele informacyjne, które dostarczają treści pochodzące z niezależnych serwisów. Te panele zapewniają użytkownikom i użytkowniczkom dodatkowy kontekst i pomagają im podejmować bardziej świadome decyzje dotyczące tego, co oglądają”.


W 16 państwach UE YouTube wprowadził „Breaking News Shelf” – zakładkę na stronie głównej serwisu, w której promuje rzetelne źródła wiadomości na gorące, newsowe tematy dla danego kraju. Dodatkowo, kiedy takie

gorące wydarzenia mają miejsce, system rekomendacyjny w wynikach wyszukiwania podbija wiarygodne źródła. Niestety, w naszym rozumieniu te interwencje dotyczą tylko wyników wyszukiwania w serwisie YouTube, a nie treści aktywnie podsuwanych przez algorytm („Proponowane do obejrzenia w następnej kolejności”).

YouTube zaktualizował swoje podejście do treści związanych z zaburzeniami odżywiania, które obecnie obejmuje: zakaz publikowania treści dotyczących zaburzeń odżywiania, ograniczenie wiekowe dla niektórych filmów oraz wyświetlanie przy filmach o takiej tematyce paneli informacyjnych, odsyłających do rzetelnych publikacji o zaburzeniach odżywiania.

YouTube wprowadził ograniczenia w dostępie do treści, które nie naruszają regulaminu, ale są nieodpowiednie dla osób poniżej 18 roku życia. Obejmuje to filmy pokazujące dorosłych uczestniczących w niebezpiecznych czynnościach, które dzieci mogą chcieć naśladować, filmy pokazujące substancje zakazane dla dzieci, treści o charakterze seksualnym lub treści zawierające przemoc i wulgaryzmy. Filmy z ograniczeniami wiekowymi nie są dostępne dla wylogowanych użytkowników i użytkowniczek.

Algorytmy YouTube’a analizują sygnały, które mogą świadczyć o tym, że właścicielem kanału jest osoba poniżej 13 roku życia. W przypadku takiego podejrzenia, zawartość kanału jest oceniana przez wyspecjalizowany zespół. W stosunku do użytkowników i użytkowniczek zaklasyfikowanych jako „potencjalny małoletni”, YouTube uruchamia wymuszony

<p>proces odzyskiwania konta. Taki użytkownik ma dwa tygodnie na udowodnienie, że ukończył 13 lat, lub uzyskanie zgody rodziców i konfigurację kontroli rodzicielskiej.</p>	
<h2>Higiena cyfrowa</h2>	
<p>Meta zachęca swoich użytkowników i użytkowniczki do korzystania z „narzędzi zarządzania czasem i innych zasobów, żeby mieć lepszą kontrolę i czuć większą sprawczość”. Niestety, w raporcie brakuje konkretnych przykładów.</p> <p>TikTok oferuje wszystkim użytkownikom i użytkowniczkom panel umożliwiający zarządzanie „Codziennym czasem przed ekranem”. Pozwala on ustawić przypomnienie o przekroczeniu dziennego limitu czasu ekranowego. Dla wszystkich młodszych użytkowników i użytkowniczek domyślnie ustawiony limit wynosi 60 minut. Kontrolki dotyczące czasu ekranowego są dostępne w panelu rodzica (Family Pairing).</p> <p>YouTube dostosowuje przypomnienia o przerwie (Take a Break) i nocnym odpoczynku (Bedtime) do różnych grup wiekowych. Ustawienia rodzicielskie umożliwiają określenie limitu czasu, jaki dziecko może spędzić w aplikacji YouTube Kids.</p> <p>W serwisie YouTube funkcja autoodtwarzania jest domyślnie wyłączona dla użytkowników</p>	<p>Oczekujemy dalszych środków wspierających wychodzenie z uzależnień behawioralnych:</p> <ul style="list-style-type: none"> <li>• rezygnacji z funkcji, które niezależni badacze klasyfikują jako „uzależniające”, w szczególności: automatycznych powiadomień, niekończącego się feedu, funkcji auto-odtwarzania oraz liczników sygnałów społecznej walidacji (takich jak lajki) dla wszystkich kont, bez względu na wiek;</li> <li>• ostrzeżeń, że użytkownik/-czka spędził już 30 minut w danym serwisie;</li> <li>• zachęt do ustawienia własnych limitów korzystania z aplikacji;</li> <li>• zachęt do ograniczenia powiadomień do niezbędnych i wybrania „higienicznego” sposobu ich dostarczania (np. powiadomienia o nowych wiadomościach lub komentarzach przychodzą w paczkach po upływie określonego</li> </ul>



i użytkowników, którzy zadeklarowali, że mają mniej niż 18 lat, dla kont nadzorowanych oraz w aplikacji YouTube Kids. Rodzic ma możliwość wyłączenia kontrolki autoodtworzenia, aby dziecko nie mogło samodzielnie włączyć tej opcji.

W listopadzie 2022 r. YouTube uruchomił kampanię edukacyjną "Hit Pause (Zatrzymaj)". Firma deklaruje, że będzie „ nadal badać możliwości opracowania dodatkowych kampanii”.

czasu lub w czasie wskazanym przez użytkownika/-czkę);

- zachęt do ustawienia dnia bez social mediów;
- automatycznego blokowania aplikacji po upływie ustawionego czasu lub w czasie zdefiniowanym przez użytkownika/-czkę jako czas offline;
- spowalniania tempa, w jakim można przewijać feed, proporcjonalnie do czasu spędzonego na platformie (tak, by nie nagradzać nadmiernego korzystania z platformy),
- prowadzenia na platformie/ w aplikacji kampanii uświadamiających ryzyka związane z nadmiernym korzystaniem z social mediów.