

Moderacja „Facebooka” bez cenzury

3 czerwca 2017

Portale społecznościowe stały się dla użytkowników sieci kluczowymi pośrednikami w dostępie do informacji. Warto pamiętać jednak, że obiegiem tych informacji w dużej mierze sterują dziś firmy internetowe. To one decydują o blokowaniu i filtrowaniu treści, które udostępniasz, wykorzystując stworzoną przez nie „infrastrukturę”. Niestety szczegółowe kryteria i procedury, którymi w tym zakresie kierują się cyberkorporacje, są często niedostępne lub co najmniej nieprzejrzyste dla użytkowników. Trochę więcej światła na zasady moderacji treści w mediach społecznościowych rzuca niedawna publikacja dziennika „The Guardian”. Ujawnia ona wewnętrzne wytyczne stworzone dla pracowników „Facebooka”, dzięki którym po raz pierwszy możemy przyjrzeć się bliżej kulisom pracy moderatorów tego serwisu.



BLOKOWANIE WPISÓW – PORADNIK PRAKTYCZNY

Zapewne większość z nas traktuje media społecznościowe jako platformę swobodnej wymiany opinii i poglądów rozpowszechnianych w ramach korzystania z wolności słowa. Od

dawna wiemy jednak, że serwisy te mogą być także zręcznym narzędziem dystrybucji treści, które są fałszywe, obraźliwe, a w skrajnych przypadkach stanowią mowę nienawiści, nawoływanie do terroryzmu czy przemocy. Dokumenty opublikowane przez „The Guardian” ujawniają, w jaki sposób moderatorzy „Facebooka” powinni reagować na zgłoszenia dotyczące tego typu treści, a także innych kontrowersyjnych publikacji zawierających nagość, pornografię czy akty autoagresji. Wytyczne te mają pomóc im w ocenie, które materiały mieszczą się jeszcze w granicach swobody wypowiedzi, a które wykraczają już poza jej ramy.

W sumie „The Guardian” dotarł do ponad 100 dokumentów zawierających szczegółowe instrukcje dla moderatorów. Niektóre z nich budzą pewne wątpliwości. Wynika z nich np., że moderatorzy powinni ostrzej reagować na nawoływanie do przemocy wobec osób publicznych niż prywatnych (np. wypowiedź „Niech ktoś zastrzeli Trumpa” powinna zostać zablokowana, ale wpis „Mam nadzieję, że ktoś Cię zabije” w odniesieniu do osoby prywatnej może pozostać w serwisie). Dodatkowo, jak podaje brytyjski dziennik, moderatorzy mierzą się z niemal niewykonalnym zadaniem, mając zazwyczaj na dokonanie odpowiedniej kwalifikacji kontrowersyjnych treści (tj. np. oddzielenie znieważenia od uprawnionej satyry) jedynie...10 sekund. Podobne decyzje sądy podejmują po wielomiesięcznych procesach.

WŁADZA W PARZE Z ODPOWIEDZIALNOŚCIĄ

Nie chodzi jednak o to, aby odebrać portalom społecznościowym kompetencje do blokowania treści czy ustalania w tym zakresie własnych reguł. Administratorzy tych serwisów są zresztą bardzo ważnym ogniwem walki z bezprawnymi treściami w Internecie. Są w stanie najszybciej zareagować na nieuprawnione treści, a biorąc pod uwagę specyfikę rozprzestrzeniania się informacji w sieci, czas reakcji ma ogromne znaczenie w kontekście minimalizowania krzywdzących skutków takich publikacji. Co więcej, administratorzy portali są prawnie zobowiązani, aby reagować na bezprawne treści, pod

groźbą poniesienia odpowiedzialności, jeśli pomimo wiedzy o naruszeniu, pozostaną wobec niego bezczynni (te kwestie reguluje w Polsce ustawa o świadczeniu usług drogą elektroniczną).

Jeśli jednak oddajemy portalom społecznościowym „władzę” nad informacją, możemy żądać, aby korzystały z niej w odpowiedzialny sposób, z poszanowaniem praw swoich użytkowników. Wymagajmy więc, aby portale te były naszymi sprzymierzeńcami w walce z bezprawnymi treściami w sieci, ale jednocześnie nie stosowały arbitralnie „prywatnej cenzury”. Presja sprostania tym niekiedy trudnym do pogodzenia oczekiwaniom, powoduje, że błędy w moderacji (takie jak brak odpowiedniej reakcji na mowę nienawiści czy wręcz przeciwnie – niezasadne zablokowanie uprawnionego wpisu – najpewniej zawsze będą się zdarzały. Nie wyeliminują ich nawet najlepsze wytyczne, ani zaangażowanie większej liczby moderatorów (obecnie według „The Guardian” „Facebook” zatrudnia ich ok. 4,5 tys.). Ryzyko i skutki błędnych decyzji może ograniczyć jednak wprowadzenie większej przejrzystości zasad funkcjonowania serwisów społecznościowych, która umożliwi poddanie ich skuteczniejszej społecznej kontroli.

MNIEJ TAJEMNIC, WIĘCEJ PRZEJRZYSTOŚCI

Jak ta przejrzystość wygląda dzisiaj? Poziom szczegółowości ujawnionych przez „The Guardian” wytycznych istotnie różni się od tego, co możemy przeczytać w regulaminie „Facebooka” i w „Standardach społeczności” na temat kryteriów dopuszczalności treści. „Facebook” nie publikuje też pełnych danych statystycznych dotyczących otrzymywanych zgłoszeń i blokowanych informacji. W tzw. „Raporcie przejrzystości” serwis podaje jedynie dane związane z otrzymywaniem tego rodzaju żądań od instytucji publicznych poszczególnych państw. Zgłaszając daną treść do moderacji, nie wiemy też kto, na jakich zasadach i w jakim czasie zajmie się naszym zgłoszeniem. Jeśli natomiast portal zablokował nasz wpis, nie ma jasnych reguł określających procedurę odwołania się od tej

decyzji, ani oczywistych środków prawnych, z których można byłoby skorzystać, aby ją zakwestionować.

Ważne jest zatem zwiększenie transparentności kryteriów podejmowania decyzji o blokowaniu treści, co m.in. umożliwiłoby poddanie ich publicznej dyskusji. Równie istotne jest stworzenie odpowiednich mechanizmów proceduralnych pozwalających użytkownikom skutecznie weryfikować zasadność decyzji podjętych przez moderatorów, zarówno na drodze wewnętrznej, jak i przed niezależnym organem. Co więcej, przejrzyste zasady zarządzania treścią powinny dotyczyć nie tylko ludzkich zachowań, ale również funkcjonowania automatycznych systemów, które mogą być stosowane do identyfikacji i blokowania określonych treści.

Podsumowując, dokumenty ujawnione przez „The Guardian” uświadamiają przede wszystkim, jak niewiele wiemy o zasadach zarządzania treścią w portalach społecznościowych i jak bardzo jesteśmy zdani w tym zakresie na łaskę i niełaskę firm internetowych. Jako użytkownicy tych serwisów, możemy próbować to zmienić, wywierając presję bezpośrednio na cyberkorporacje, ale także na organy państwa. Te ostatnie są przecież zobligowane do ochrony praw jednostek przed ich naruszaniem przez inne podmioty prywatne, zwłaszcza gdy dotyczy to obszaru mającego tak istotny wpływ na realizację podstawowych praw i wolności, takich jak swoboda wypowiedzi czy prawo do prywatności w Internecie.

Autorstwo: Dorota Głowacka

Zdjęcie: [Simon](#) (CC0)

Źródło: Panoptikon.org