

Jak KE chce walczyć z treściami terrorystycznymi w sieci?

25 września 2018

Wszystkie portale będą musiały w ciągu godziny usunąć treści wskazane przez służby, a ponadto samodzielnie przeczesać swoje zasoby (też za pomocą algorytmów) i się z tego sprawozdawać – to najważniejsze założenia zaprezentowanego kilkanaście dni temu projektu rozporządzenia, którego celem jest zwalczanie treści terrorystycznych w sieci.

O pomysł Komisji Europejskiej na regulację treści terrorystycznych pisaliśmy przy okazji wydanych w marcu rekomendacji. Wtedy nie podobało nam się, że służby miały uzyskać nowe uprawnienia do zgłaszania treści terrorystycznych poza jakąkolwiek ścieżką legislacyjną, a platformy miały pełnić rolę arbitrów rozstrzygających w ciągu godziny, czy służbom uwierzyć, czy nie. W ubiegłym tygodniu Komisja zdecydowała się uregulować tę kwestię w drodze rozporządzenia, które – w przeciwieństwie do rekomendacji – podlega normalnej procedurze legislacyjnej i będzie powszechnie obowiązujące.

Zacznijmy od tego, z jakimi treściami chce walczyć Komisja. Treści terrorystyczne to te, które pochwalają czyny terrorystyczne i wzywają do ich popełniania. Za terrorystyczne zostaną też uznane instrukcje, jak dokonać zamachu, czy zachęcanie do dołączenia do grup terrorystycznych. Oceniając, czy dana treść ma charakter terrorystyczny, organ lub firma internetowa mają wziąć pod uwagę kontekst, w jakim treść się pojawiła; to, jakim językiem jest sformułowana i jakie jest ryzyko, że pozostawienie jej na portalu doprowadzi do szkodliwych konsekwencji. Komisja podkreśla, że za terrorystyczne nie powinny być uznawane treści rozpowszechniane w celach edukacyjnych, dziennikarskich lub

badawczych, a także komunikaty zawierające radykalne, polemiczne lub kontrowersyjne poglądy wyrażane w debacie publicznej dotyczącej drażliwych problemów politycznych.

Zgodnie z projektem walka z treściami terrorystycznymi ma być prowadzona na trzy sposoby.

1. Organ krajowy zwraca się do portalu z nakazem usunięcia treści, który musi mieć formę decyzji administracyjnej lub wyroku sądu – w tej sytuacji portal nie ma nic do powiedzenia i musi usunąć treść w ciągu godziny.

2. Organ krajowy lub unijny (np. Europol) przekazuje portalowi zgłoszenie, które ten ocenia samodzielnie. To portal decyduje, czy treść usunąć, czy nie.

3. Portale prowadzą działania zmierzające do wykrycia treści terrorystycznych, również za pomocą automatycznych filtrów.

W marcu pisaliśmy, że Facebook, Twitter i inne, mniejsze portale nie będą w stanie zweryfikować zgłoszenia pochodzącego od służb w ciągu godziny, co naturalnie będzie prowadziło do automatycznych blokad, niezależnie od tego, jakie uzasadnienie przedstawią służby. Na szczęście Komisja w projekcie rozporządzenia zaproponowała, żeby treści zdejmowane były automatycznie tylko wtedy, gdy służba wystąpi do portalu nie z luźnym zgłoszeniem, ale z decyzją administracyjną (która jest poprzedzona postępowaniem) lub z wyrokiem sądu. Ta propozycja jest niezwykle podobna do już istniejących w Polsce rozwiązań wprowadzonych ustawą antyterrorystyczną. Na jej mocy Szef ABW może – po uzyskaniu zgody Prokuratora Generalnego i sądu – zarządzić zablokowanie dostępności „danych informatycznych” mających związek ze zdarzeniem o charakterze terrorystycznym. Konieczność uzyskania akceptacji dwóch instytucji odrobinę zmniejsza zagrożenia związane z samym tworzeniem mechanizmów cenzorskich, jednak ryzyko nadmiernego ograniczania wolności słowa jest aktualne – zwłaszcza że ABW nie chce ujawnić nawet tego, ile stron zablokowało (sprawa czeka na rozstrzygnięcie

sądu).

Drugi pomysł Komisji na zwalczanie treści terrorystycznych przypomina konstrukcję notice and takedown. Z założenia to nie jest zły pomysł, zwłaszcza jeśli weźmiemy pod uwagę, że użytkownicy, których treść została usunięta, będą się mogli od tej decyzji odwołać. Nikt jednak nie będzie czuwał nad tym, jak te procedury odwoławcze będą wyglądać – projekt rozporządzenia nie ustanawia wspólnych standardów, tylko pozostawia to w całości portalom, co rodzi ryzyko dla wolności słowa, zwłaszcza w sytuacji rażącej dysproporcji sił: czy ktoś wierzy, że portale będą się masowo sprzeciwiać „sugestiom” potężnych służb w imię obrony swojego pojedynczego użytkownika?

Nie tak dawno świat obiegła informacja, że algorytm „You Tube” flaguje jako treści terrorystyczne filmy, które dokumentują działania wojenne w Syrii i niekiedy stanowią jeden z kluczowych dowodów zbrodni wojennych. Tymczasem w projekcie przedstawionym przez Komisję zobowiązanie portali do podejmowania działań, w tym filtrowania treści za pomocą automatycznych narzędzi, stało się faktem. Komisja zdaje się dostrzegać podnoszony przez nas wielokrotnie problem pomyłek wpisanych w takie narzędzia i próbuje ograniczać arbitralność platform, zobowiązując je do publikowania raportów przejrzystości (dowiemy się z nich m.in., ile treści zostało usuniętych, a ilu użytkowników się temu sprzeciwiło i jak te sprzeciwy zostały rozpatrzone). Portale miałyby także w przystępny sposób tłumaczyć użytkownikom, jak działają automatyczne filtry i inne metody wyłapywania treści terrorystycznych. Nie zmienia to jednak faktu, że arbitralne podejmowanie decyzji przez portale – czy to poprzez moderatorów, czy „rękami” algorytmów – rodzi poważne ryzyko dla wolności słowa i dostępu do informacji.

Nie kwestionujemy konieczności walki z treściami terrorystycznymi, które zagrażają porządkowi publicznemu, bezpieczeństwu i demokracji. Jednak propozycja Komisji wpisuje

się w szerszą tendencję wzmacniania firm w roli prywatnych cenzorów, którzy podejmują decyzje o tym, jakie informacje do nas dotrą, a jakie nie. A to oznacza jeszcze więcej władzy nad obiegiem informacji w sieci dla firm i jeszcze mniej praw dla użytkowników.

Autorstwo: Karolina Iwańska

Na podstawie: EC.Europa.eu, Panoptikon.org, Fastcompany.com

Źródło: Panoptikon.org