

TWITTER WALCZY Z TERRORYSTAMI. ALGORYTMY TO ZA MAŁO

Od połowy 2015 r. Twitter zawiesił 125 tys. kont dżihadystów, którzy na swoich kanałach propagowali terrorystyczne treści i udostępniali filmiki z egzekucji. To jednak kropla w morzu potrzeb, bo rosnąca fala niedozwolonych wpisów rośnie. Grupy słabo opłacanych studentów i mieszkańców krajów trzeciego świata nie są w stanie wyłapać wszystkiego, a żaden algorytm nie jest doskonały.

Kulisy walki Twittera ze zwolennikami Daesh ujawniają dziennikarze The Independent. Ich artykuł pokazuje, że amerykańska firma starając się być bastionem wolności słowa ma problem z nielegalnymi treściami zamieszczanymi przez terrorystów. Błędem jest wyobrażenie o wojownikach Daesh jako o ludziach pustyni odciętych od cywilizacji. To zazwyczaj osoby świetnie poruszające się w cyfrowym świecie, z ogromną wiedzą i biegłością w korzystaniu z internetowych narzędzi. Doskonale wykorzystują je do swojej propagandy.

Ciężko z tym walczyć za pomocą komputerowych algorytmów. Te narzędzia nie są jeszcze tak rozwinięte, by wyłapać mowę nienawiści czy obraźliwe treści publikowane w sieci. Programy nie są w stanie odróżnić zakazanych treści od legalnych politycznych komentarzy.

Terrorystyci Daesh w swojej pracy czasami wykorzystują automatyczne konta, które publikują w sieci nienawistne komentarze. Takie informacje są jeszcze względnie łatwe do wyłapania i zablokowania – na podobnej zasadzie, na jakiej programy są w stanie odseparować spam. Gorzej, gdy wpisy tworzy prawdziwa osoba, a nie automat. Wtedy ryzyko wykrycia przez komputerowe narzędzia spada do zera.

Dlatego Twitter nie polega tylko na algorytmach i zatrudnia armię ludzi, którzy każdego dnia przeczesują sieć w poszukiwaniu nielegalnych treści. Problem w tym, że to fatalnie opłacani pracownicy, zazwyczaj studenci lub mieszkańcy biedniejszych krajów, jak Filipiny. Zarabiają mniej niż 500 dolarów miesięcznie, a ich praca to codzienne oglądanie filmów z dziecięcą pornografią i terrorystycznymi egzekucjami. Lepiej zarabiające zespoły z USA i Irlandii, które zajmują się czytaniem twitterowych profili oznaczonych jako „szerzące treści terrorystyczne”. Praca tych osób też nie jest łatwa. Muszą sami interpretować wpisy i podejmować decyzje, które treści są legalne, a które nie. Jak zapewniają fachowcy zatrudnieni przez amerykańskiego molocha, nie jest to łatwe zadanie.

Wszystkie te zespoły opierają swoją pracę na zgłoszeniach od użytkowników Twittera. Bez aktywności internautów wyłapanie nienawistnych treści byłoby niemal niemożliwe. W ten sposób to sami użytkownicy wskazują, co ich zdaniem jest dozwolonym politycznym dyskursem, a co przekracza granicę przyzwoitości.

Okazuje się, że internetowa aktywność staje się kwestią życia i śmierci. Przekonał się o tym pochodzący z Wielkiej Brytanii hacker Daesh Junaid Hussain, który zginął w ataku dronów w 2015 r. Wojskowi swoją decyzję o ataku tłumaczyli właśnie jego aktywnością na Twitterze. Z kolei szef

Twittera Dick Costolo otrzymał śmiertelne pogróżki po tym, jak udostępnił zainteresowanym organom informacje o kontaktach związanych z Daesh.

Źródło: The Independent