ALGORYTMY TWITTERA AUTOMATYCZNIE USUWAŁY CO DRUGĄ OBRAŹLIWĄ WIADOMOŚĆ

Automatyczne narzędzia monitorujące Twittera skutecznie rozpoznawały i usuwały co drugą wiadomość zawierającą obraźliwe treści w pierwszej połowie 2019 r. Jeszcze rok temu algorytmy odpowiadały za co piątą interwencję - podała agencja Reutera.

W raporcie dot. przejrzystości firma odpowiedzialna za platformę społecznościową stwierdziła, że bada możliwości proaktywnych mechanizmów monitoringu treści, by zmniejszyć zależność od ręcznych zgłoszeń nieodpowiednich wiadomości przez użytkowników.

Twitter poinformował, że w przeciągu półrocza o 105 proc. wzrosła liczby zamkniętych lub zawieszonych kont, które łamały zasady korzystania z serwisu. W porównaniach rok do roku o 48 proc. wzrosła także liczba kont moderowanych ze wzgledu na łamanie polityki dot. mowy nienawiści.

W opisywanym okresie blisko 116 tys. kont zamknięto w związku z podejrzeniem o działalność terrorystyczną. To spadek o 30 proc. w porównaniu z pierwszą połową ubiegłego roku.

W pierwszym półroczu 2019 r. o 67 proc. wzrosły ponadto motywowane prawnie wnioski o usunięcie zawartości. Spłynęły one z 49 krajów. Za 80 proc. ich całkowitej liczby odpowiedzialne były Japonia, Rosja i Turcja.

Firmy z Doliny Krzemowej w ostatnich miesiącach zadeklarowały zaostrzenie zasad dotyczących niewłaściwej zawartości i dzielenie się informacjami o niej z innymi podmiotami. Samoregulacja ma być sposobem na uniknięcie wprowadzenia odgórnych przepisów prawnych w różnych krajach świata.