FACEBOOK: ALGORYTMY NIE WYCHWYCIŁY NAGRANIA Z ZAMACHU

Algorytmy do wykrywania łamiących regulamin Facebooka treści są wciąż niedoskonałe, a materiał szkoleniowy dla nich - niewystarczający, dlatego nie oflagowały nagrania transmitowanego na żywo przez zamachowca z Christchurch - tłumaczy koncern na swoim blogu.

Jak napisał wiceprezes Facebooka ds. integralności Guy Rosen "natychmiastowość transmisji w ramach opcji Facebook Live (pozwalającej na przekaz materiałów wideo na żywo - PAP) niesie z sobą szczególne wyzwania", i w ciągu kilku ostatnich lat firma skoncentrowała się na tym, by zespół moderacji szybciej wykrywał kluczowe materiały".

Rosen wyjaśnił, że Facebook wykorzystuje narzędzia sztucznej inteligencji do wykrywania filmów wymagających podjęcia działań przez moderatorów oraz do określania stopnia ich ważności. Jak dodał, chodzi przede wszystkim nagrania zawierające sceny przedstawiające przemoc bądź samobójstwa.

Wiceprezes koncernu stwierdził, że sztuczna inteligencja wspomagająca pracę moderatorów Facebooka "nie jest doskonała". Wyjaśnił, że systemy te działają w oparciu o "dane szkoleniowe", co oznacza, że muszą wpierw nauczyć się rozpoznawać dane treści na tysiącach przykładowych obrazów, nagrań i tekstów. Rosen podkreślił, że narzędzia te "sprawdziły się bardzo dobrze w przypadku treści takich jak nagość, propaganda terrorystyczna i graficzne przedstawienia przemocy, gdyż istnieje duża liczba przykładowych materiałów, mogących posłużyć do wyszkolenia systemów" Facebooka.

Guy Rosen podkreślił jednak, że wideo transmitowane przez zamachowca z Christchurch nie zostało wychwycone przez sztuczną inteligencję, gdyż system nie dysponował odpowiednio dużą ilością treści tego rodzaju, co umożliwiłoby mu oflagowanie tej treści jako łamiącej regulamin.

Jednocześnie przedstawiciel Facebooka podkreślił, że systemy sztucznej inteligencji działają coraz lepiej, gdy chodzi o wykrywanie treści powiązanych z terroryzmem na platformie. "Ich wydajność wzrasta, choć nigdy nie będą doskonałe. Ludzie zawsze będą elementem tego równania, niezależnie od tego, czy pracują w zespole moderatorów, czy też jako użytkownicy zgłaszają treści, które moderacji powinny podlegać". Jak wyjaśnił Rosen, dlatego Facebook w ostatnim roku zwiększył liczbę osób pracujących w zespołach bezpieczeństwa Facebooka do ponad 30 tys. W gronie tym znajduje się również około 15 tys. moderatorów treści.

Rosen zapowiedział, że w związku ze sprawą zamachu w Nowej Zelandii koncern współpracuje z krajową policją i ma zamiar wspierać ją w kolejnych działaniach, również tych mających na celu

wyjaśnienie roli platform internetowych w wydarzeniach w Christchurch. Jak napisał wiceprezes Facebooka, chodzi o znalezienie najbardziej efektywnych środków regulacyjnych i technologicznych do zapobiegania takim wydarzeniom w przyszłości, m.in. poprzez usprawnienie procesów moderacji i walkę z mową nienawiści na platformie, a także wzmocnienie współpracy z Międzynarodowym Forum Internetowym na rzecz Przeciwdziałania Terroryzmowi (GIFCT).

W strzelaninie w Christchurch 15 marca zginęło 50 osób, a kilkadziesiąt zostało rannych. 28-letni Australijczyk Brenton Tarrant transmitował przez 17 minut swoje działania właśnie za pośrednictwem platformy Facebooka. Koncern usunął jego konto oraz materiał po otrzymaniu informacji na ten temat od nowozelandzkiej policji. Nadawane na żywo przez zamachowca wideo zdążyło obejrzeć mniej niż 200 użytkowników Facebooka, żaden z nich jednak nie zgłosił filmu do moderacji. Facebook dowiedział się o nim 29 minut po rozpoczęciu transmisji. Do chwili usunięcia materiału z platformy obejrzano go cztery tysiące razy.