Ludzie mogą nie rozpoznać sztucznie wygenerowanej mowy

15 sierpnia 2023

Ludzie są w stanie wykrywać sztucznie generowaną mowę tylko w 73 proc. przypadków, niezależnie od języka – informuje „Plos One”.

Deepfake to technika obróbki obrazu lub dźwięku przeprowadzona przez sztuczną inteligencję. W przypadku dźwięków polega ona na generowaniu próbek mowy czy komunikatów językowych, które mają przypominać głos prawdziwej osoby. Opiera się na uczeniu maszynowym, które trenuje algorytm w celu poznania wzorców i cech z dostarczonego zbioru danych.

Podczas gdy wczesne algorytmy tzw. deepfejka audio wymagały dostarczenia tysięcy próbek głosu danej osoby, aby móc wygenerować przypominające jej mowę dźwięki, to najnowsze algorytmy mogą osiągnąć ten efekt używając zaledwie trzysekundowego fragmentu wypowiedzi.

Naukowcy z University College London wykorzystali taki algorytm na dwóch publicznie dostępnych zestawach danych, jednym w języku angielskim i jednym w mandaryńskim, w celu wygenerowania 50 fałszywych próbek mowy w każdym z tych języków.

Otrzymane próbki odtworzono następnie 529 uczestnikom badania, aby sprawdzić, czy będą oni w stanie rozróżnić prawdziwą mowę od deepfejka. Okazało się, że udało się to tylko w 73 proc. sytuacji, co oznacza, że przeciętny człowiek nie potrafi rozpoznać ponad jednej czwartej przypadków fałszywej mowy. Bardzo nieznaczna poprawa wyników nastąpiła po przeszkoleniu uczestników w temacie identyfikowania sfabrykowanego głosu.

„Nasze odkrycie potwierdza, że ludzie nie są w stanie skutecznie wykrywać fałszywej mowy, niezależnie od tego, czy

przeszli szkolenie w tym zakresie, czy też nie” – mówi dr Kimberly Mai, główna autorka badania. „Warto również zauważyć, że próbki, których użyliśmy w badaniu, zostały utworzone przy użyciu stosunkowo starych algorytmów, co rodzi pytanie, czy gdybyśmy użyli najbardziej wyrafinowanej technologii, czyli tej, jaką dysponujemy teraz, sytuacja nie wyglądałaby jeszcze gorzej”.

Kolejnym krokiem naukowców będzie opracowanie skuteczniejszych, automatycznych detektorów mowy w celu przeciwdziałania zagrożeniu deepfejkami.

Jak podkreślają autorzy publikacji, chociaż istnieją pewne korzyści z generowanej przez AI mowy, np. dla osób, które utraciły głos z powodu choroby – to rosną obawy, że technika ta może być wykorzystywana do celów przestępczych.

Udokumentowano już takie przypadki. Jednym z nich jest incydent z 2019 r., w którym dyrektor generalny brytyjskiej firmy energetycznej został przekonany przez fałszywy głos swojego przełożonego do przekazania oszustowi setek tysięcy funtów.

„Dzięki coraz bardziej wyrafinowanej generatywnej sztucznej inteligencji i otwartemu dostępowi do tego typu narzędzi jesteśmy coraz bliżej różnych korzyści oraz różnych zagrożeń” – podsumowuje prof. Lewis Griffin, współautor badania. „Rozsądne byłoby więc, aby rządy i organizacje opracowały strategie radzenia sobie z nadużywaniem tych narzędzi. Choć ważne jest też, abyśmy nie zapominali o korzyściach, które takie technologie oferują”.

Autorstwo: Katarzyna Czechowicz (PAP)
Źródło: NaukawPolsce.pl