

Program sprawdzisz człowieczeństwo rozmówców online

21 lipca 2024

Prototypowy program do sprawdzania, czy przez kamerę online rozmawialiśmy z prawdziwą osobą, a nie deepfejkem, opracowali naukowcy z ośrodka badawczo-rozwojowego IDEAS NCBR. Rozwiązanie ma służyć walce z oszustwami i ochronie wizerunku w sieci.

Coraz częściej słyszymy o przypadkach oszustw wykorzystujących deepfejk – zaawansowaną technikę manipulacji materiałami wideo lub audio przy użyciu sztucznej inteligencji (AI). Technologia deepfejk pozwala na tworzenie realistycznych i trudnych do odróżnienia od oryginału fałszywych materiałów multimedialnych – nagrań wideo lub audio. Deepfejk można wykorzystywać w branży rozrywkowej albo edukacji, jednak często służy również dezinformacji, manipulacji i oszustwom.

W Polsce w marcu 2024 r. fałszywe wideo z udziałem CEO InPost Rafała Brzoski, prezydenta Andrzeja Dudy albo piłkarza Roberta Lewandowskiego promowały oszukańcze „programy inwestycyjne” służące wyłudzeniu pieniędzy” – przypomniano w komunikacie nadesłanym PAP we wtorek przez IDEAS NCBR. Autorzy dokumentu podali też, że w styczniu br. pracownik banku w Hongkongu wypłacił 25 milionów dolarów oszustom, którzy – stosując deepfejk – podczas wideokonferencji podszyli się pod dyrektora finansowego firmy.

„Deepfejki stają się coraz bardziej zaawansowane, a ich wykrycie coraz trudniejsze, co stawia przed nami nowe wyzwania w zakresie bezpieczeństwa cyfrowego” – uważają badacze z IDEAS NCBR, ośrodka badawczo-rozwojowego w obszarze sztucznej inteligencji.

Naukowcy z tej instytucji stworzyli eksperymentalny projekt o nazwie ProvenView. Twórcami rozwiązania są postdoc Shahriar Ebrahimi i doktorantka Parisa Hassanizadeh. To badacze z Iranu pracujący w grupie „Bezpieczeństwo systemów i prywatność danych” kierowanej przez prof. dr hab. Stefana Dziembowskiego. Opracowanie wykorzystuje Zero-Knowledge Proof (ZKP) – technikę kryptograficzną, która pozwala na udowodnienie prawdziwości pewnej informacji bez ujawniania szczegółów o niej.

„Technologia ZKP jeszcze dekadę temu uważana była za dziedzinę wyłącznie teoretyczną, bliższą czystej matematyce niż zastosowaniom informatyki [...] Rozwiązanie ProvenView stanowi jeszcze jeden spektakularny przykład tego, jak ZKP może być wykorzystane w walce z zagrożeniami współczesnego świata cyfrowego” – powiedział cytowany w komunikacie prof. Dziembowski.

Jak zapewniają twórcy tego programu, ProvenView może znaleźć zastosowanie w interakcjach online i w ochronie treści. Na przykład po zakończeniu spotkania online użytkownik będzie mógł sprawdzić, czy wideo jego rozmówcy rzeczywiście pochodziło z jego kamery internetowej. „To rozwiązanie może się przydać zarówno użytkownikom indywidualnym, którzy mogliby go użyć do autoryzacji spotkań online, jak i twórcom treści, którzy dzięki temu mogliby wzmocnić ochronę swojego wizerunku opublikowanego w filmie na YouTube przed kradzieżą i wykorzystaniem np. w pornografii” – wyjaśnił zacytowany w dokumencie Shahriar Ebrahimi.

Naukowcy podali uproszczony przykład działania ProvenView. W pierwszym kroku autor oryginalnego wideo tworzy dowód jego autentyczności – cyfrowy zapis potwierdzający, że materiał multimedialny jest jego autorstwa. Oprogramowanie umożliwiające taki zapis może być wbudowane w aplikację do nagrywania wideo lub być osobnym programem. Następnie, kiedy ktoś chce sprawdzić autentyczność wideo, może zobaczyć ten cyfrowy dowód.

„I właśnie tu pomaga nam Zero-Knowledge Proofs: dzięki zaawansowanej matematyce dowód ten można sprawdzić bez potrzeby oglądania samego wideo. Każdy może generować takie dowody autentyczności na swoim własnym urządzeniu” – tłumaczył Shahriar Ebrahimi. Dodał, że w przyszłości to rozwiązanie mogłoby stać się dodatkiem do przeglądarki internetowej albo programu do edycji wideo.

ProvenView to rozwiązanie w fazie eksperymentalnej, które nie jest jeszcze gotowe do wdrożenia. Ebrahimi podkreślił, że w tej chwili dowody autentyczności można generować dopiero po zakończeniu tworzenia wideo, a uzyskać je – by sprawdzić pochodzenie materiału multimedialnego – dopiero ok. godziny po odbyciu rozmowy. Naukowcy z IDEAS NCBR mają nadzieję, że w przyszłości możliwe będzie udowadnianie autentyczności wideo w czasie rzeczywistym.

W maju 2024 r. ProvenView zostało nagrodzone podczas hackathonu ZK Hack Kraków.

IDEAS NCBR został powołany w 2021 r. przez Narodowe Centrum Badań i Rozwoju. Jego misją jest wsparcie rozwoju tej technologii w Polsce poprzez stworzenie platformy łączącej środowisko akademickie z biznesowym. Ośrodek kształci również nowe pokolenie naukowców, ukierunkowanych na praktyczne zastosowanie opracowanych algorytmów oraz ich późniejszą komercjalizację.

Autorstwo: PAP

Źródło: [NaukawPolsce.pl](https://naukawpolsce.pl)