CYBERODPORNOŚĆ

DeepFake a odporność systemów opartych na wideoweryfikacji tożsamości

dr Agnieszka Besiekierska, dr Kamil Czaplicki

XVI Konferencja Bezpieczeństwo w Internecie - Cyberodporność

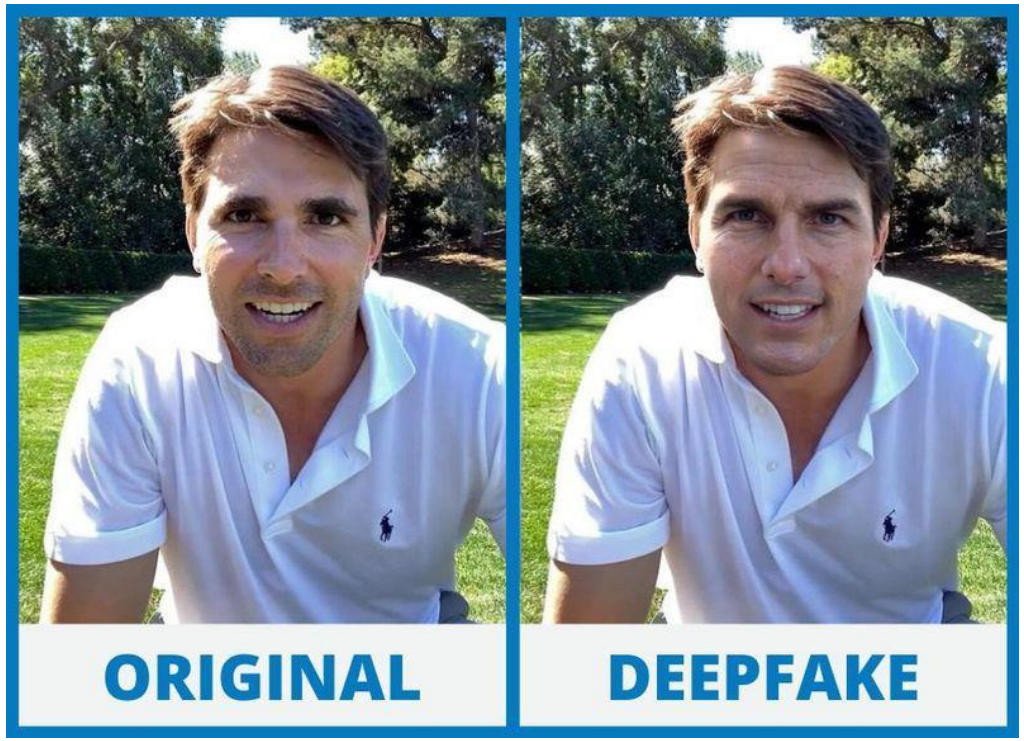
Warszawa 5 grudnia 2024 r.

Organizatorzy: UKSW, Ministerstwo Cyfryzacji, Naukowe Centrum Prawno-Informatyczne

Partner wspierający: NASK

Partner merytoryczny: SAMSUNG

Deepfake - wygenerowane przez AI lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo, które przypominają istniejące osoby, przedmioty, miejsca, podmioty lub zdarzenia, które odbiorca mógłby niesłusznie uznać za autentyczne lub prawdziwe (Art. 3 pkt 60 rozp. akt w sprawie SI)



https://www.trymaverick.com/blog-posts/are-deep-fakes-all-evil-when-can-they-be-used-for-good

[wideoidentyfikacja jako system identyfikacji tożsamości] **„system zdalnej identyfikacji biometrycznej w czasie rzeczywistym”** oznacza system zdalnej identyfikacji biometrycznej, w którym zbieranie danych biometrycznych, ich porównywanie i identyfikacja odbywają się bez znacznego opóźnienia, i który obejmuje nie tylko natychmiastową identyfikację, ale także ograniczone krótkie opóźnienia w celu uniknięcia obchodzenia przepisów (Art. 3 pkt 42 rozp. akt w sprawie SI)

Relacja cyberbezpieczeństwo - sztuczna inteligencja

1. Cyberbezpieczeństwo systemów SI;
2. **SI wykorzystana do cyberataków** – deepfake'i;
3. **SI wspiera cyberbezpieczeństwo** – systemy SI wykrywania deepfake'ów.

Wykorzystanie wideoweryfikacji w bankowości

Stanowiska UKNF
- dotyczące identyfikacji klienta i weryfikacji jego tożsamości w bankach oraz oddziałach instytucji kredytowych w oparciu o metodę wideoweryfikacji z dnia 5 czerwca 2019
- dotyczące identyfikacji klienta instytucjonalnego i weryfikacji jego tożsamości w sektorze finansowym (...) w oparciu o metodę wideoweryfikacji z dnia 3 marca 2022

Regulamin wideoweryfikacji tożsamości jednego z banków podkreśla udział (**weryfikację przez) „wykwalifikowanego agenta ludzkiego.” =>**

Ważne w świetle art. 22 ust. 3 RODO czyli „administrator wdraża właściwe środki ochrony praw, wolności i prawnie uzasadnionych interesów osoby, której dane dotyczą, a co najmniej **prawa do uzyskania interwencji ludzkiej ze strony administratora, do wyrażenia własnego stanowiska i do zakwestionowania tej decyzj**i.”

Wykorzystanie wideoweryfikacji w administracji publicznej

Osoba upoważniona do potwierdzania profilu zaufanego w sposób może w trakcie transmisji audiowizualnej zażądać od wnioskodawcy wykonania gestu, który ułatwi wykrycie działania oprogramowania zakłócającego teletransmisję audiowizualną uniemożliwiającego lub utrudniającego potwierdzenie tożsamości wnioskodawcy.

(§ 6 ust. 7 Rozporządzenie MC z dnia 29 czerwca 2020 r. w sprawie profilu zaufanego i podpisu zaufanego).

Powrót

Wideospotkanie i masz profil zaufany

17.06.2021

Od dziś profil zaufany można potwierdzić także w trakcie wideospotkania z urzędnikiem. To nie wszystko - tymczasowe PZ stają się tradycyjnymi. W skrócie: jeśli nie masz profilu zaufanego, od teraz jeszcze łatwiej go założyć.

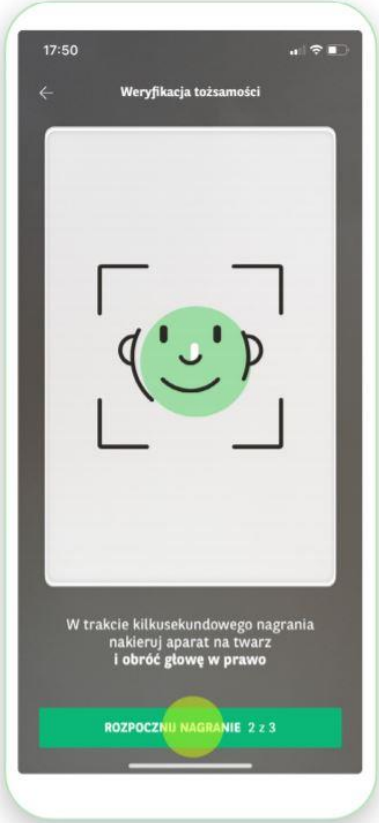
Wykorzystanie wideoweryfikacji w administracji publicznej

Hipotetyczne wykorzystanie systemów SI do wideoweryfikacji w przypadku szczególnych rodzajów usług - usługi opieki zdrowotnej, świadczenia z tytułu macierzyństwa, choroby, czy utraty zatrudnienia, a także z tytułu pomocy społecznej. *„Wskazuje się, iż w tym wypadku należy mieć świadomość, iż systemy SI mogą naruszać ich prawa podstawowe, takie jak prawo do ochrony socjalnej, niedyskryminacji, godności człowieka lub skutecznego środka prawnego i dlatego też systemy te należy zaklasyfikować jako **systemy wysokiego ryzyka**. (motyw 58).”*

Obowiązek nadzoru człowieka art. 14 rozp. akt w sprawie SI i związane z tym wyzwania

Wideoweryfikacja tożsamości a deepfake

Według szacunków KPMG dostępnych jest obecnie ok. 100 tys. modeli AI pozwalających na generowanie lepszych lub gorszych deepfake'ów. Jednocześnie tylko niewielka część z nich – ok. 3 tys. pozwala na wykrywanie manipulacji w materiałach audiowizualnych. Na szczęście istnieją rozwiązania komercyjne, często dostarczane przez największe firmy na rynku IT, które umożliwiają identyfikowanie deepfake'ów. Również takie narzędzia wykorzystują algorytmy AI, które pozwalają na wykrywanie zmanipulowanych materiałów ze skutecznością przekraczającą 95 proc. – i to w czasie zaledwie kilku sekund.



https://biznes.t-mobile.pl/pl/deepfakes-realne-zagrozenie-dla-cyberbezpieczenstwa

CYBERODPORNOŚĆ

Dziękujemy

UKSW

Ministerstwo Cyfryzacji

Naukowe Centrum Prawno-informatyczne

SAMSUNG

NASK