

OLIWIA RAJTAR¹

Technologie deepfake w zakresie danych osobowych

1. Wstęp

Obecnie świat, w którym Internet i media społecznościowe dominują jako główne źródła informacji, staje się szczególnie niebezpieczną przestrzenią. Ludzie muszą zmagać się z ogromem informacji i stale dokonywać ich selekcji, aby uniknąć ryzyka otrzymania nieprawdziwych treści. Zjawisko fałszywych informacji, czyli fake newsów, szybko rozprzestrzenia się w sieci, natomiast od kilku lat szczególnym zagrożeniem stały się filmiki, obrazy i dźwięki wygenerowane przez sztuczną inteligencję, zwane deepfake'ami.

Manipulacja informacją w erze Internetu i mediów społecznościowych zyskała zupełnie nowy wymiar dzięki nieznanym wcześniej możliwościom rozpowszechniania treści i tworzenia wirusów informacyjnych. Zabiegi te prowadzą do kryzysu zaufania, który obecnie dotyka populację całego świata. Od dawna ludzkość korzysta z obrazowych reprezentacji informacji jako skutecznego sposobu przekazywania wiedzy. Spośród wszystkich form reprezentacji wizualnych zdjęcia i filmy są uważane za najbardziej wiarygodne źródła. Przedstawiają one bowiem obraz rzeczywistości, który wpływa na ludzkie przekonania tak samo, jak bodźce z rzeczywistego świata. Zdjęcia i filmy są uznawane za najdokładniejsze formy dokumentacji wizualnej, zarówno osób, przedmiotów, jak i wydarzeń, i są powszechnie wykorzystywane w dziedzinach takich jak dziennikarstwo, polityka, sądownictwo, marketing. Mają tak duże znaczenie, że w kryminalistyce są używane praktycznie od momentu wynalezienia fotografii.

Jednakże w drugiej dekadzie XXI w. pojawiły się nowe możliwości manipulowania rzeczywistością, np. programy do obróbki wideo, które zmieniły

1 Oliwia Rajtar – Instytut Prawa, Ekonomii i Administracji, Uniwersytet Komisji Edukacji Narodowej w Krakowie, e-mail: oliwia.rajtar@student.up.krakow.pl.

sposób, w jaki możemy postrzegać obrazy. Jedną z tych technologii jest deepfake – technologia oparta na sztucznej inteligencji, pozwalająca na tworzenie realistycznych, ale fałszywych filmów poprzez łączenie i modyfikowanie istniejących obrazów i nagrań. Deepfaki tworzy zespół algorytmów, które mogą dokładnie manipulować wizerunkiem ludzi w filmach. Dzięki temu możliwe jest spreparowanie sytuacji, które w rzeczywistości nigdy się nie wydarzyły. Osoby na ekranie mogą mówić i podejmować działania, które w rzeczywistości nie miały miejsca.

To narzędzie wykorzystuje sztuczną inteligencję do tworzenia i modyfikowania wizerunku ludzi w niespotykanym dotąd wymiarze. W przeciwieństwie do tradycyjnych metod, takich jak edycja zdjęć w Photoshopie, które były używane do fałszowania obrazów, deepfake daje możliwość manipulowania materiałem wideo w sposób znacznie bardziej realistyczny. To wzbudza poważne wątpliwości co do wiarygodności materiałów wideo, co z kolei wpływa na nasze postrzeganie informacji. Technologia deepfake stwarza nowe wyzwania, ponieważ wykrycie fałszerstw w materiale wideo staje się coraz trudniejsze – z każdym dniem oprogramowania do tego typu wideo są doskonalone, technologia ta znacząco obniża zatem wiarygodność widzianych przez nas materiałów wideo, wpływając na nasze postrzeganie informacji.

2. Czym jest technologia deepfake?

Deepfake to technologia oparta na sztucznej inteligencji, która umożliwia tworzenie realistycznych obrazów, filmów i dźwięków, przy czym uwzględnić należy, że wytwory te nie są prawdziwe, ale wygenerowane komputerowo. Nazwa „deepfake” pochodzi od połączenia słów *deep learning*, co oznacza ‘głębokie uczenie’, i *fake*, czyli ‘fałszywy, nieszczerzy’. Technologia ta wykorzystuje zaawansowane algorytmy uczenia maszynowego, aby manipulować i generować treści multimedialne trudne do odróżnienia od rzeczywistych gołym okiem².

Podstawą działania technologii deepfake są techniki głębokiego uczenia, a w szczególności sieci neuronowe zwane GAN (*Generative Adversarial Networks*). Sieci GAN składają się z dwóch głównych komponentów:

1. Generator: Sieć ta generuje fałszywe obrazy, filmy lub dźwięki, starając się naśladować prawdziwe dane.
2. Dyskryminator: Sieć ta ocenia, czy dane pochodzą od generatora, czy też są prawdziwe. Dyskryminator pomaga generatorowi doskonalić swoje umiejętności, dostarczając informacji zwrotnej na temat jakości wygenerowanych danych³.

2 A.M. Almars, *Deepfakes Detection Techniques Using Deep Learning: A Survey*, Yanbu 2021, s. 9.

3 Dane ze strony internetowej: <https://www.cvisionlab.com/cases/deepfake-gan/> (20.05.2024).

Proces ten polega na wielokrotnym doskonaleniu się obu sieci, gdzie generator staje się coraz lepszy w tworzeniu realistycznych fałszywych obrazów, a dyskryminator – lepszy w ich wykrywaniu. W rezultacie powstają obrazy i filmy, które mogą wyglądać niezwykle realistycznie, przy czym w całości są wytworzone przez sztuczną inteligencję⁴.

Fot. 1. Personalizacja wideo przy użyciu technologii deepfake



Źródło: <https://www.trymaverick.com/blog-posts/are-deep-fakes-all-evil-when-can-they-be-used-for-good> (22.05.2024).

Tworzenie deepfake'ów może być stosunkowo łatwe, zwłaszcza gdy odpowiednie narzędzia i oprogramowania do wytwarzania tego rodzaju fałszywych filmików są dostępne powszechnie. Wystarczy dostarczyć programowi odpowiednią ilość danych, aby stworzyć realistyczną manipulację wideo. Co za tym idzie, powstaje coraz więcej wyzwań związanych z deepfake'ami i ich wykrywaniem. Zaniepokojenie związane z deepfake'ami coraz częściej prowadzi do podejmowania działań prawnych. W Australii już uchwalono ustawę nakładającą surowe kary na osoby lub firmy rozpowszechniające tego rodzaju fałszywe treści⁵. Jednak masowa dostępność oprogramowania deepfake stanowi nadal poważny problem, który utrudnia odróżnienie prawdy od kłamstwa. Dzięki odpowiedniej ilości danych i czasu przeznaczanego

⁴ Dane ze strony internetowej: <https://towardsdatascience.com/deepfakes-the-ugly-and-the-good-49115643d8dd> (20.05.2024).

⁵ Dane ze strony internetowej: <https://www.gtlaw.com.au/knowledge/australian-government-targets-sexually-explicit-deepfakes> (26.06.2024).

na szkolenie komputerowe sieci GAN, nieprawdziwe filmy, tworzone nawet w zaciszu domowym, bywają bardzo przekonujące. Jednak uważniejsi obserwatorzy mogą zauważyć brak subtelnych sygnałów fizjologicznych, takich jak mruganie oczami czy brak unoszenia się klatki piersiowej, które mogą ujawnić fałszywość tych materiałów⁶.

3. Historia terminu

Sam termin „deepfake” wywodzi się od użytkownika o pseudonimie „DeepFakes”, który w 2017 r. opublikował na platformie Reddit kilka filmów pornograficznych, w których twarze aktorek zastąpił twarzami sławnych osób, jak Daisy Ridley czy Scarlett Johansson. Materiał był wysoce realistyczny, gdyż DeepFakes wykorzystał zaawansowaną technologię sztucznej inteligencji.

4. Zastosowanie deepfake’ów i ich uwarunkowania w popkulturze

Technologia deepfake znajduje zastosowanie w różnych dziedzinach – od branży rozrywkowej po system edukacyjny. Początkowo z deepfake’ów korzystano w sferze rozrywkowej, używając ich jako filtrów w aplikacjach takich jak Snapchat czy Instagram. W dziedzinie kinematografii deepfake wykorzystywany jest do tworzenia efektów specjalnych, takich jak odmładzanie twarzy lub do cyfrowe „wskrzeszanie” zmarłych aktorów⁷. Przykładem jest film *Szybcy i Wściekli 7*, gdzie za sprawą technologii CGI (deepfake jest jego tańszą pochodną, która za sprawą *deep learning* szybciej opanowuje obróbkę obrazu) na twarz dublera została nałożona cyfrowo zrekonstruowana twarz zmarłego w wypadku samochodowym aktora – Paula Walkera, odtwórcy roli Briana O’Connera w serii tych filmów. W muzyce deepfake znajduje zastosowanie w tworzeniu innowacyjnych wideoklipów oraz hologramów zmarłych artystów⁸. Przykładem tego może być hologramowy koncert Tupaca Shakura na Coachelli w 2012 r.

Deepfake może być używany do tworzenia wirtualnych „instruktorów”, którzy są realistycznymi replikami znanych ekspertów w danej dziedzinie czy nawet historycznych postaci. Te cyfrowe wersje mogą prowadzić lekcje, udzielać porad i odpowiadać na pytania uczniów w sposób bardzo zbliżony do rzeczywistego. Aktywność uczniów zwiększa się, gdy podczas zajęć mogą wchodzić w interakcje z wirtualnymi instruktorami i uczestniczyć w dynamicznych

6 Dane ze strony internetowej: <https://us.norton.com/blog/emerging-threats/what-are-deepfakes> (19.05.2024).

7 Dane ze strony internetowej: <https://screenrant.com/furious-7-brian-scenes-not-paul-walker-brothers/> (20.05.2024).

8 Dane ze strony internetowej: <https://www.cbsnews.com/news/tupac-coachella-hologram-behind-the-technology/> (19.05.2024).

scenariuszach przedstawianych w ramach podstawy programowej⁹. Pomimo wielu zalet wykorzystanie technologii deepfake w edukacji i szkoleniach wiąże się z istotnymi wyzwaniem, które należy uwzględnić, decydując się na takie rozwiązanie. Wykorzystanie sztucznej inteligencji w obszarze edukacji i szkolnictwa często godzi w etykę i autentyczność przedstawionych obrazów¹⁰. Autentyczność informacji przekazywanych przez wirtualnych instruktorów nie zawsze jest weryfikowana; nie ma zatem gwarancji, że wiedza jest rzetelna i wiarygodna. Ponadto wykorzystanie deepfake'ów wymaga uwzględnienia aspektów etycznych, jak uzyskanie zgody na użycie wizerunku danej osoby – to pozwala uniknąć nieautoryzowanego lub nieetycznego zastosowania tej technologii. Konieczne jest, aby wszelkie zastosowania deepfake'ów w edukacji były przejrzyste i respektowały prawa wszystkich zaangażowanych stron. Problem pojawia się szczególnie w sytuacji wykorzystywania wizerunku osób historycznych, zmarłych, niemogących wyrazić zgody na przetwarzanie swojego wizerunku¹¹.

Technologia deepfake znajduje szerokie zastosowanie w marketingu i reklamie – umożliwia personalizację kampanii marketingowych, np. poprzez tworzenie realistycznych, interaktywnych postaci. Dzięki tej technologii marki personalizują swoje przekazy, dzięki czemu odbiorcy otrzymują reklamę skierowaną bezpośrednio do nich. Deepfake pozwala na stworzenie wirtualnych ambasadorów marki, którzy mogą komunikować się z klientami w czasie rzeczywistym, odpowiadając na ich pytania i promując produkty w sposób bardziej angażujący¹².

W Internecie technologia deepfake jest wykorzystywana do tworzenia zabawnych filmów i obrazków (memów), które szybko zyskują popularność w mediach społecznościowych. Ta technologia umożliwia tworzenie wizerunków przedstawiających znane osoby publiczne, polityków lub celebrytów w humorystycznych sytuacjach, co przyciąga uwagę i zwiększa zaangażowanie użytkowników w prowadzone przez kreatorów tych treści konta¹³.

Jednak pomimo wymienionych zastosowań technologia deepfake wiąże się również z poważnymi zagrożeniami, i to na dużą skalę. Raport *2023 State of Deepfakes* opracowany przez Home Security Heroes dostarcza szczegółowej analizy aktualnego stanu technologii deepfake i materiałów z jej użyciem, zamieszczanych w sieci. Analiza ta bazuje na badaniu 95 820 filmów deepfake,

9 Dane ze strony internetowej: <https://www.tieonline.com/article/3632/the-rise-of-deepfakes-in-schools> (21.05.2024).

10 Dane ze strony internetowej: <https://www.edweek.org/leadership/deepfakes-expose-public-school-employees-to-new-threats/2024/05> (21.05.2024).

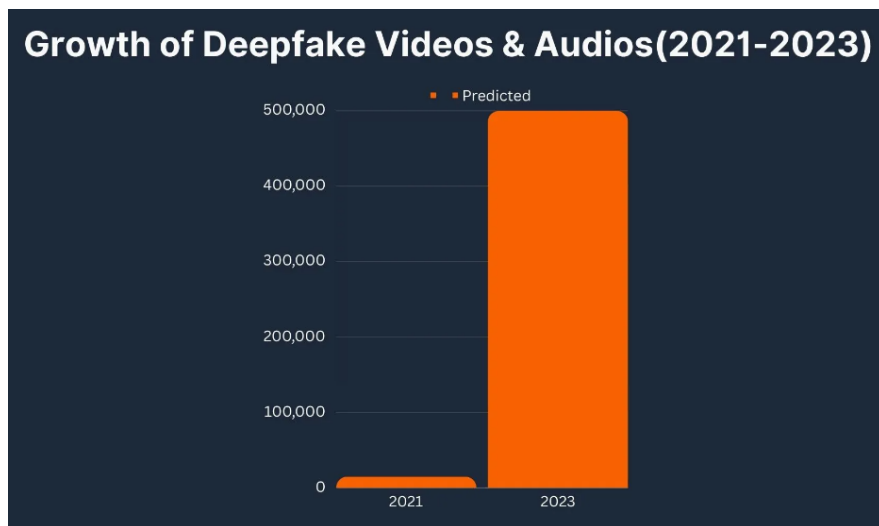
11 Dane ze strony internetowej: <https://www.restack.io/p/ai-deepfakes-answer-deep-fake-historical-figures-cat-ai> (22.05.2024).

12 Dane ze strony internetowej: <https://www.mediatropy.com/tech/how-ai-deep-fakes-are-transforming-the-world-of-marketing-and-advertising/> (22.05.2024).

13 Dane ze strony internetowej: <https://www.technologyreview.com/2020/08/28/1007746/ai-deepfakes-memes/> (22.05.2024).

85 specjalistycznych kanałów na różnych platformach internetowych oraz ponad 100 stron związanych z wykorzystywaniem deepfake'ów. Z raportu wynika, że w 2023 r. w sieci znajdowało się 95 820 filmów deepfake, co oznacza wzrost o 550% w porównaniu do roku 2019. Natomiast na rok 2023 przewidywano już pół miliona takich materiałów, które pojawić by się miały na różnych portalach internetowych i w social mediach¹⁴.

Fot. 2. Wzrost liczby deepfake'ów wideo i audio (2021–2023)



Źródło: <https://i0.wp.com/contentdetector.ai/wp-content/uploads/2023/12/deepfake-videos-Audios.jpg?ssl=1> (17.05.2024).

5. Zagrożenia i niebezpieczeństwa związane z deepfakiem

Filmy deepfake stanowią potężne narzędzie, które przestępcy wykorzystują w różnych działaniach, takich jak szantaż czy oszustwa. Dzięki technologii deepfake oszuści mogą łatwo tworzyć realistyczne nagrania wideo, które trudno rozpoznać jako fałszywe, przynajmniej bez zaawansowanej analizy prawnej – postaci ukazane w tych filmach, podszywając się pod konkretne osoby, oszukują innych, włącznie z organami ścigania. Deepfaki są nie tylko wykorzystywane w satyryczny sposób, ale mogą też stanowić poważne zagrożenie w wojnie informacyjnej¹⁵. Przeciwnik zagraniczny może wykorzystać deepfaki, aby wpłynąć na wybory poprzez publikację filmów oczerniających danego kandydata. Taka niepewność może również podważyć wiarygodność

¹⁴ Dane ze strony internetowej: <https://www.securityhero.io/state-of-deepfakes/> (23.05.2024).

¹⁵ H. Ajder et al., *Deeptrace: The State of DeepFakes. Landscape, Threats and Impact*, Amsterdam 2019, s. 15.

dziennikarzy, gdy fałszywa treść zostanie pomyślnie przedstawiona jako prawdziwa. Wykorzystanie deepfake'ów może prowadzić do wymuszeń lub fałszywych oskarżeń. Niewłaściwe użycie tej technologii stwarza środowisko, w którym trudno odróżnić prawdziwe wydarzenia od fałszywych, a to wprowadza dezorientację i chaos. Filmy deepfake przynoszą ze sobą nie tylko alternatywne fakty, ale także tworzą zupełnie nową rzeczywistość, która może mieć poważne konsekwencje dla osób i instytucji, będących ofiarami oszustw i manipulacji¹⁶.

6. Przypadki oszustw i ich konsekwencje

Deepfake to doskonałe narzędzie do oszustw. Przeszczepcy mogą tworzyć klony dźwiękowe, brzmiące dokładnie tak, jak wybrana postać. W tym celu z publicznych źródeł zbierają próbki głosu osób związanych z ofiarą, „trenują” stworzone przez siebie modele przy użyciu sztucznej inteligencji i tworzą syntetyczne głosy osób, pod które chcą się podszyć. Udając znajomych bądź członków rodziny, którzy potrzebują pomocy finansowej, nakłaniają ofiarę do przesłania pieniędzy. Oszuści kopiują także głosy pracodawców, dyrektorów firm generalnych, aby przekonać pracowników do przeniesienia środków na wskazane konta. Takie wykorzystanie technologii deepfake stanowi poważne zagrożenie dla bezpieczeństwa finansowego zarówno firm, jak i osób prywatnych¹⁷.

6.1. Kradzież tożsamości

Wiele firm prosi klientów o przesłanie zdjęć dowodu osobistego w celu potwierdzenia tożsamości. Tę sytuację wykorzystują oszuści: stosując zaawansowane programy do przekształcania zdjęć z mediów społecznościowych w realistyczne maski 3D, tworzą przekonujące duplikaty dokumentów tożsamości, dzięki czemu uzyskują dostęp do kont klientów. W miarę jak banki i inne instytucje finansowe zaczynają wdrażać biometrię głosową jako metodę weryfikacji tożsamości, pojawia się kolejne zagrożenie: deepfaki mogą omijać te środki bezpieczeństwa, podrabiając głos danej osoby. Tęgo rodzaju oszustwa stawiają przed firmami wyzwanie w zakresie ochrony danych klientów i weryfikacji tożsamości, wymagają wprowadzenia bardziej zaawansowanych technologii ochronnych oraz ciągłego doskonalenia istniejących systemów zabezpieczeń¹⁸.

16 *Ibidem*, s. 19–20.

17 Dane ze strony internetowej: <https://respeecher.medium.com/deepfake-or-synthetic-voice-whats-the-difference-43d84895b0a3> (27.05.2024).

18 Dane ze strony internetowej: <https://www.nsa.gov/Press-Room/Press-Releases-Statements/Press-Release-View/Article/3523329/nsa-us-federal-agencies-advise-on-deepfake-threats/> (28.05.2024).

6.2. Pornografia i sekstorsja

Niektóre witryny umożliwiają klientom zakup fałszywej pornografii, przedstawiającej wybrane przez nich osoby, choć nie wyraziły one zgody na wykorzystanie ich wizerunku, co stanowi poważne naruszenie prywatności i godności. W innych przypadkach przestępcy łączą publicznie dostępne zdjęcia niczego niepodejrzewających osób – np. celebrytów, nieletnich – z materiałami jednoznacznie pornograficznymi i tworzą w ten sposób realistyczne, lecz fałszywe obrazy i filmy¹⁹. Takie treści są często wykorzystywane do wymuszeń lub zemsty, powodują ogromne szkody emocjonalne i reputacyjne dla ofiar. Te praktyki są nie tylko moralnie i etycznie naganne, ale również stanowią poważne zagrożenie prawne i psychologiczne. Ich obecność zwraca uwagę na potrzebę wprowadzenia skutecznych regulacji prawnych oraz zaawansowanych technologii ochrony prywatności, aby zapobiegać takim nadużyciom i chronić ofiary przed dalszymi szkodami²⁰. Z kolei sekstorsja (ang. *sextortion*) to forma szantażu polegająca na groźeniu ofierze ujawnieniem jej nagich zdjęć lub filmów, jeśli nie spełni ona żądań sprawcy, a te często dotyczą przekazania pieniędzy, wynikają z chęci poczucia władzy czy pragnienia upokorzenia ofiary.

Przestępcy wykorzystują skradzione dane kontaktowe i zhakowane konta do wysyłania wiadomości, w których twierdzą, że przejęli kontrolę nad komputerem ofiary i wykradli jej prywatne zasoby zdjęciowe. Sekstorsja objawia się również wymuszaniem od nastolatków nagrań pornograficznych z ich własnym udziałem²¹.

6.3. Manipulacje wyborcze i spiski

Fałszywe wizerunki przywódców, kandydatów w wyborach i innych osobistości władzy pokazują, że składają oni fałszywe twierdzenia. Celem jest często zdyskredytowanie mówcy, wciągnięcie kandydatów politycznych w kontrowersje, a tym samym zmniejszenie ich szans i pozyskanie głosów dla innego kandydata. Samo istnienie deepfake'ów rodzi kulturę spiskową. Konspiranci lub przeciwnicy polityczni mogą z łatwością uznać prawdziwe wydarzenia i rzeczywiste dowody za fałszywe. Rezultatem jest rosnący sceptycyzm wobec dziennikarzy i mediów informacyjnych oraz ogólna erozja prawdy w społeczeństwie²².

19 Dane ze strony internetowej: <https://arxiv.org/abs/1905.08233> (27.05.2024).

20 Europol, *Facing Reality? Law Enforcement and the Challenge of Deepfakes. An Observatory Report from the Europol Innovation Lab*, Luxembourg 2022, s. 10.

21 S. Rathod et al., *Tracing of the Blackmailers in Sextortion Case and Tactics to Defend It – An Experimental Cybercrime Case Study*, „International Journal of Scientific Research in Science and Technology” 2021, vol. 7, z. 4, s. 3–4.

22 S. Ahmed, *Navigating the Maze: Deepfakes, Cognitive Ability, and Social Media News Skepticism*, „New Media & Society” 2021, vol. 25, z. 5, s. 4.

6.4. Celebryci i politycy

Przeciwnicy polityczni coraz częściej tworzą deepfaki znanych polityków, aby podsycać konflikty, zdobywać poparcie lub wywoływać zamieszanie wśród ludności cywilnej. Przykładem jest deepfake z marca 2022 r. przedstawiający prezydenta Ukrainy Wołodymyra Zełenskigo poddającego się Rosji; deepfake został szybko zidentyfikowany i usunięty przez władze jako zmanipulowane wideo. Nieuczciwe firmy również zaczęły wykorzystywać fałszywe wizerunki celebrytów do promowania swoich produktów, ale bez ich wiedzy i zgody. Nowa technologia pozwala na tworzenie fałszywych filmów, w których znane postacie, takie jak były prezydent USA Donald Trump czy wysocy rangą dyplomaci, są przedstawiane w kontrowersyjny sposób, mający na celu manipulację opinią publiczną²³. W eksperymencie z 2016 r. technika deepfake została zastosowana do podrobienia twarzy światowych liderów, takich jak George W. Bush czy Barack Obama²⁴. Wkrótce potem zaczęto jej używać do manipulacji wizerunkiem Obamy w sposób, który wzbudził duże kontrowersje. Eksperyment z 2016 r. miał na celu zbadanie potencjalnych zagrożeń związanych z technologią deepfake, szczególnie w kontekście manipulacji wizerunkami publicznymi i możliwości szerzenia dezinformacji. Badacze chcieli pokazać, jak łatwo można wpłynąć na opinię publiczną, tworząc realistyczne, ale fałszywe nagrania znanych postaci. Deepfake został wykorzystany do stworzenia fałszywych przemówień, które mogłyby wprowadzać odbiorców w błąd co do rzeczywistych poglądów i działań byłego prezydenta, co uwidocznilo ryzyko dla debaty publicznej i procesów demokratycznych. Tego rodzaju praktyki nie tylko podważają zaufanie publiczne, ale również naruszają prawa osób publicznych i prywatnych, co wymaga wprowadzenia surowszych regulacji prawnych oraz zaawansowanych technologii wykrywających fałszerstwa. W przypadku wykorzystania deepfake'ów względem celebrytów doskonałym przykładem jest Taylor Swift. Zwolennicy Donalda Trumpa opublikowali w lutym 2024 r. zmanipulowane materiały na platformie X (dawniej Twitter), które fałszywie przedstawiają ją jako zwolenniczkę Trumpa i osobę kwestionującą wyniki wyborów²⁵. Rozneglizowane fałszywe zdjęcia i klipy z wykorzystaniem wizerunku artystki zostały obejrzone miliony razy, co uwidocznilo problem X z kontrolą złośliwych i nieautentycznych mediów. W odpowiedzi na incydent platforma X tymczasowo zablokowała wyszukiwanie frazy „Taylor Swift” i nadal zмага się z kontrolą rozpowszechniania deepfake'ów dotyczących artystki. Niektóre deepfaki przedstawiające Swift jako osobę wspierającą Trumpa są oznaczone etykietami ostrzegającymi, że media są nieautentyczne,

23 Deepfaki były wielokrotnie używane do publikacji fałszywych wideo na różnych platformach internetowych.

24 Dane ze strony internetowej: <https://www.buzzfeed.com/craigsilverman/obama-jordan-peelee-deepfake-video-debunk-buzzfeed> (26.05.2024).

25 Dane ze strony internetowej: <https://www.yahoo.com/entertainment/fact-check-video-allegedly-shows-010400414.html> (27.05.2024).

jednak wiele udostępnień i repostów tych materiałów początkowo nie miało takich oznaczeń. Najbardziej rozpowszechniony pro-Trump deepfake Swift wykorzystuje jej nagranie z czerwonego dywanu Grammy – piosenkarka trzyma znak z napisem „Trump wygrał, Demokraci oszukali!”.

Fot. 3. Taylor Swift z transparentem, na którym jest napis: „Trump Won, Democrats Cheated!”



Źródło: <https://www.yahoo.com/entertainment/fact-check-video-allegedly-shows-010400414.html> (25.05.2024).

6.5. Marketing

W marketingu deepfaki mogą być wykorzystywane do oszustw i manipulacji, gdzie fałszywe postacie mają wprowadzać konsumentów w błąd, promując produkty lub usługi, które nie istnieją lub nie spełniają obietnic. Może to prowadzić do utraty zaufania do poszczególnych marek i ogólnego spadku wiarygodności reklam. Zwiększona liczba oszustw może również skłonić organy regulacyjne do wprowadzenia surowszych przepisów dotyczących wykorzystania tej technologii w marketingu²⁶. Ponadto wiarygodność reklam jako źródła informacji może znacząco ucierpieć i tym samym utrudnić firmom skuteczne komunikowanie się z konsumentami. Mimo potencjalnych korzyści w kreatywnych kampaniach technologia deepfake stwarza poważne zagrożenia, dlatego kluczowe jest, aby firmy i marketerzy korzystali z niej w sposób etyczny i odpowiedzialny, dzięki czemu unikną długoterminowych negatywnych skutków dla swojej reputacji i utraty zaufania konsumentów²⁷.

²⁶ Dane ze strony internetowej: <https://blog.emb.global/deepfakes-responsible-use/> (28.05.2024).

²⁷ *Ibidem*.

7. Przeciwdziałanie zagrożeniom związanym z deepfake'ami

Aby przeciwdziałać zagrożeniom związanym z deepfake'ami, konieczne są różnorodne strategie i działania. Przede wszystkim należy inwestować w rozwój zaawansowanych technologii detekcji, które mogą wykrywać fałszywe obrazy i filmy poprzez analizę subtelnych różnic w teksturach skóry, nieregularności w ruchach ust czy nienaturalnych cieni. Na przykład w USA agencja DARPA rozpoczęła projekt Medifor, który ma opracować narzędzia do automatycznej oceny autentyczności zdjęć i filmów oraz oprogramowanie wykrywające manipulacje w wideo²⁸. Podobnie firma AI Foundation pracuje nad technologią, która ma weryfikować autentyczność mediów, w tym deepfake'ów. Produkt AI Foundation Reality Defender w celu identyfikowania zmanipulowanych treści łączy moderację człowieka z uczeniem maszynowym²⁹.

Kluczowa jest również edukacja społeczna, podnosząca świadomość na temat technologii deepfake i jej zagrożeń, ucząca rozpoznawania fałszywych treści oraz krytycznego podejścia do informacji. Wprowadzenie i egzekwowanie regulacji prawnych, penalizujących nieautoryzowane tworzenie i rozpowszechnianie deepfake'ów, oraz zapewnienie ochrony prawnej dla ofiar, np. prawo do usunięcia fałszywych treści, jest niezbędne³⁰.

Współpraca międzynarodowa w zakresie opracowania standardów i strategii zwalczania deepfake'ów oraz zaangażowanie firm technologicznych w tworzenie narzędzi wykrywających fałszywe treści i wprowadzanie wewnętrznych regulacji są kluczowe. Świadomość zagrożeń oraz wdrażanie odpowiednich środków ochronnych pozwolą na minimalizowanie ryzyka i zapewnienie bezpiecznego korzystania z technologii deepfake³¹.

8. Wykorzystywanie wizerunku osoby bez jej zgody a ochrona danych osobowych

Wizerunek osoby, czyli jej wygląd zewnętrzny przedstawiony na zdjęciach, filmach czy innych nośnikach, jest integralną częścią danych osobowych, których przetwarzanie bez zgody może naruszać prywatność i dobra osobiste. Co za tym idzie, wizerunek każdej osoby jest jej prywatną własnością i podlega ochronie prawnej na mocy ustawy o prawie autorskim oraz przepisów dotyczących ochrony danych osobowych, takich jak rozporządzenie

28 Dane ze strony internetowej: <https://www.darpa.mil/program/media-forensics> (25.05.2024).

29 Dane ze strony internetowej: <https://www.realitydefender.com/> (24.05.2024).

30 Dane ze strony internetowej: <https://www.scoredetect.com/blog/posts/legal-remedies-for-deepfake-victims-guide> (25.05.2024).

31 A. Henry et al., *Deeptrace: The State of DeepFakes...*, s. 9.

o ochronie danych osobowych³². Zgodnie z RODO dane osobowe obejmują wszelkie informacje umożliwiające identyfikację osoby fizycznej, co oznacza, że przetwarzanie wizerunku podlega przepisom ochrony danych. Przetwarzanie obejmuje zbieranie, przechowywanie, rozpowszechnianie i modyfikowanie wizerunku, wymaga wyraźnej zgody osoby, której te dane dotyczą, przy czym zgoda ta musi być dobrowolna, konkretna, świadoma i jednoznaczna³³. Istnieją jednak wyjątki od wymogu uzyskania zgody, takie jak przetwarzanie niezbędne do wykonania umowy, wypełnienia obowiązku prawnego, ochrony żywotnych interesów osoby, realizacji zadania w interesie publicznym lub wynikające z prawnie uzasadnionych interesów administratora zobowiązania, o ile nie przeważają nad nimi interesy lub podstawowe prawa i wolności osoby, której dane dotyczą. W przypadku deepfake'ów zazwyczaj zgody jednostki na modyfikację owego wizerunku nie ma, co rodzi konsekwencje prawne³⁴.

9. Ochrona wizerunku w prawie polskim

W prawie polskim ochronę wizerunku zapewniają przepisy Kodeksu cywilnego³⁵ oraz ustawy o prawie autorskim i prawach pokrewnych³⁶, zgodnie z którymi rozpowszechnianie wizerunku wymaga zezwolenia osoby przedstawionej, chyba że jest ona powszechnie znana, a wizerunek utrwalono w związku z pełnieniem funkcji publicznych, zawodowych lub społecznych. Wynika to z art. 81 ustawy o prawie autorskim, który jasno mówi, że rozpowszechnianie czyjegoś wizerunku wymaga zgody samej osoby, chyba że otrzymała ona zapłatę za pozowanie. Istnienie zgody musi być udowodnione przez osobę, która chce wykorzystać czyjś wizerunek³⁷.

Rozpowszechnianie wizerunku obejmuje wszelkie formy i środki publicznego udostępniania, w tym Internet, telewizję i prasę. Naruszenie tych przepisów może prowadzić do konsekwencji prawnych i finansowych, a także do odpowiedzialności cywilnej i karnej. Osoba, której wizerunek został bezprawnie wykorzystany, może żądać zaprzestania naruszeń, usunięcia ich skutków, zadośćuczynienia pieniężnego, odszkodowania lub publicznych przeprosin. Rozporządzenie o ochronie danych osobowych przewiduje surowe sankcje

32 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych), Dz.U. UE L 119 z 4.5.2016, s. 1.

33 Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych, Dz.U. 1994, nr 24, poz. 83, art. 81.

34 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych..., art. 6 ust. 1 lit. a.

35 Ustawa z dnia 23 kwietnia 1964 r. – Kodeks cywilny, Dz.U. 1964, nr 16, poz. 93.

36 Ustawa z dnia 4 lutego 1994 r. o prawie autorskim...

37 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych..., art. 6 ust. 1 lit. a.

za naruszenia ochrony danych osobowych, w tym wizerunku, w postaci kar administracyjnych sięgających nawet 20 mln euro lub 2–4% całkowitego rocznego obrotu przedsiębiorstwa z roku poprzedzającego przewinienie, a także prawo do odszkodowania za poniesione szkody³⁸. Ochrona wizerunku jest kluczowym elementem praw osobistych, a jej naruszenie niesie za sobą poważne konsekwencje dla osoby posługującej się nielegalnie danymi osobowymi w postaci wygenerowanego filmu lub zdjęcia³⁹. Ustawa o ochronie danych osobowych penalizuje także nielegalne przetwarzanie danych osobowych – art. 107 przewiduje, że każda osoba, która przetwarza dane bez odpowiedniego uprawnienia lub w sposób niedopuszczalny, może zostać ukarana grzywną, ograniczeniem wolności lub pozbawieniem wolności do dwóch lat. Jeżeli nielegalne przetwarzanie dotyczy danych wrażliwych, takich jak informacje o pochodzeniu rasowym lub etnicznym, poglądach politycznych, przekonaniach religijnych, przynależności do związków zawodowych, danych genetycznych i biometrycznych, zdrowiu, seksualności lub orientacji seksualnej, sankcja może wynosić do trzech lat pozbawienia wolności⁴⁰.

10. Ryzyko w zakresie ochrony danych osobowych

W dobie cyfryzacji dane osobowe stały się jednym z najcenniejszych zasobów na świecie, przez firmy wykorzystywanym do profilowania konsumentów, przez rządy – do monitorowania obywateli, a przez cyberprzestępców – do różnych działań niezgodnych z prawem. Nielegalne zastosowanie danych osobowych, takie jak kradzież tożsamości, handel danymi i inwigilacja, niesie ze sobą poważne konsekwencje⁴¹. Ofiary kradzieży tożsamości mogą ponieść znaczne straty finansowe, zarówno bezpośrednie, jak i pośrednie, podczas gdy firmy mogą utracić zaufanie konsumentów i wartości rynkowej z powodu skandali związanych z wyciekiem danych. Nielegalne działania naruszają prywatność jednostek, prowadzą do stresu i problemów psychologicznych, a także pociągają za sobą konsekwencje prawne, gdyż wiele krajów ma surowe przepisy dotyczące ochrony danych osobowych, jak wspomniana wcześniej Australia⁴².

Technologia deepfake, mimo licznych korzyści, stwarza poważne zagrożenia dla prywatności i bezpieczeństwa danych osobowych, w tym stwarza szerokie pole dla oszustw, wyłudzeń, dezinformacji i manipulacji, szkodenia

38 Dane ze strony internetowej: <https://gdpr.pl/artykuly/kary-za-naruszenie-rodz> (26.05.2024).

39 Dane ze strony internetowej: <https://rsilpak.org/2024/deepfakes-a-crisis-of-human-rights/> (29.05.2024).

40 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych..., art. 9 ust. 1.

41 Dane ze strony internetowej: <https://www.unr.edu/nevada-today/news/2023/atp-deepfakes> (27.05.2024).

42 Dane ze strony internetowej: <https://www.securityhero.io/state-of-deepfakes/> (28.05.2024).

innym ludziom poprzez tworzenie kompromitujących materiałów wideo z ich udziałem⁴³. W przyszłości technologia deepfake będzie się nadal rozwijać, przynosząc nowe możliwości i wyzwania, a to wymaga zrównoważenia korzyści z jej wykorzystania z koniecznością ochrony prywatności i bezpieczeństwa danych osobowych. Sektor technologiczny i prawny będą musiały stale dostosowywać swoje strategie i regulacje, aby zapewnić bezpieczne i odpowiedzialne korzystanie z deepfake'ów⁴⁴.

11. Inicjatywy prewencyjne

Walka z nielegalnym wykorzystaniem danych osobowych wymaga zintegrowanej strategii, w której kluczowym elementem jest edukacja społeczeństwa na temat zagrożeń i metod ochrony swoich danych. Pomóc w tym mogą kampanie informacyjne i programy szkoleniowe zwiększające świadomość i umiejętności w zakresie cyberbezpieczeństwa. Państwa powinny wprowadzić i egzekwować surowsze przepisy dotyczące ochrony danych osobowych; przykładem mogą być przepisy RODO w Unii Europejskiej, nakładające na firmy obowiązek ochrony danych i dające obywatelom większą kontrolę nad swoimi informacjami⁴⁵. Wdrażanie zaawansowanych technologii zabezpieczających, takich jak szyfrowanie danych, dwuskładnikowe uwierzytelnianie czy systemy wykrywania włamań, może znacząco utrudnić cyberprzestępcom dostęp do danych osobowych. Problem ten ma charakter globalny, dlatego niezbędna jest międzynarodowa współpraca w zakresie wymiany informacji, ścigania przestępców oraz ustanawiania globalnych standardów ochrony danych. Skutki nielegalnego wykorzystania danych mogą być katastrofalne zarówno dla jednostek, jak i dla firm oraz instytucji, dlatego tylko poprzez kompleksowe i zintegrowane działania można zapewnić odpowiednią ochronę danych osobowych i zminimalizować ryzyko ich nielegalnego wykorzystania⁴⁶.

12. Wnioski

Rozwój technologii deepfake budzi coraz większe obawy związane z jej potencjalnym wpływem na debatę publiczną i bezpieczeństwo narodowe. Stała się ona bowiem narzędziem propagandy i politycznego szantażu, a realistyczne

43 Dane ze strony internetowej: <https://www.securityhero.io/state-of-deepfakes/> (28.05.2024).

44 Dane ze strony internetowej: <https://www.scoredetect.com/blog/posts/legal-remedies-for-deepfake-victims-guide> (28.05.2024).

45 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych..., art. 5.

46 Dane ze strony internetowej: <https://www.scoredetect.com/blog/posts/legal-remedies-for-deepfake-victims-guide> (28.05.2024).

manipulacje wideo mogą wprowadzać zamęt w światowym klimacie politycznym. Istnieje ryzyko poważnych konsekwencji, takich jak panika społeczna wywołana fałszywymi filmami sugerującymi ataki rakietowe lub pandemię. Dodatkowo wprowadzenie takich manipulacji do mediów społecznościowych może prowadzić do poważnej dezinformacji, wpływać na opinie społeczne i decyzje polityczne, a także odciągać uwagę od realnych problemów na arenie międzynarodowej, jak np. bombardowania niewinnej ludności w celu zajęcia nowych terenów. W erze cyfryzacji, gdzie prawda miesza się z kłamstwem, konieczna jest współpraca między władzami, firmami technologicznymi i społeczeństwem obywatelskim, aby przeciwdziałać manipulacji i dezinformacji oraz chronić fundamenty demokracji, takie jak wolność informacji i otwartość.

Ludzka skłonność do uznawania treści cyfrowych za wiarygodne potwierdza, że konieczne jest weryfikowanie ich autentyczności. Manipulacja obrazami i wideo jest prosta, co zwiększa ryzyko podejmowania błędnych decyzji i osądów na podstawie fałszywych materiałów. Szczególnie niebezpieczne są zastosowania deepfake'ów w cyberprzemocy i zemście pornograficznej – prowadzi to do upokorzenia i traumatycznych doświadczeń ofiar. W związku z tym kluczowe jest rozwijanie procedur weryfikujących autentyczność cyfrowych treści.

W kontekście ochrony danych osobowych technologia deepfake może prowadzić do poważnych naruszeń prywatności. Przykładowo, poprzez tworzenie fałszywych wizerunków osób bez ich zgody, może dojść do naruszenia prawa do prywatności i ochrony wizerunku. Fałszywe materiały mogą być wykorzystywane do szantażu, oszustw czy manipulacji, co może mieć daleko idące konsekwencje dla ofiar takich działań. W efekcie nasza wiara w autentyczność materiałów wideo zawierających, być może, dane osobowe zostaje poważnie osłabiona.

Bibliografia

- Ahmed S., *Navigating the Maze: Deepfakes, Cognitive Ability, and Social Media News Skepticism*, „New Media & Society” 2021, vol. 25, z. 5.
- Ajder H. et al., *Deeptrace: The State of DeepFakes. Landscape, Threats and Impact*, Amsterdam 2019.
- Almars A.M., *Deepfakes Detection Techniques Using Deep Learning: A Survey*, Yanbu 2021.
- Europol, *Facing Reality? Law Enforcement and the Challenge of Deepfakes. An Observatory Report from the Europol Innovation Lab*, Luxembourg 2022.
- Fajgielski P., *Prawo ochrony danych osobowych. Zarys wykładu*, Warszawa 2019.
- Rathod S. et al., *Tracing of the Blackmailers in Sextortion Case and Tactics to Defend It – An Experimental Cybercrime Case Study*, „International Journal of Scientific Research in Science and Technology” 2021, vol. 7, z. 4.
- Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem

danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych), Dz.U. UE L 119 z 4.5.2016, s. 1.

Ustawa z dnia 23 kwietnia 1964 r. – Kodeks cywilny, Dz.U. 1964, nr 16, poz. 93.

Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych, Dz.U. 1994, nr 24, poz. 83.

Young N., *DeepFake Technology: Complete Guide to Deepfakes, Politics and Social Media*, Boston 2019.

Netografia

<https://arxiv.org/abs/1905.08233> (27.05.2024).

<https://aws.amazon.com/what-is/gan/> (25.05.2024).

<https://blog.emb.global/deepfakes-responsible-use/> (28.05.2024).

<https://fal-lawyers.com.au/latest-insights/the-legal-framework-of-deepfake-technology-in-australia> (17.05.2024).

<https://gdpr.pl/artykuly/kary-za-naruszenie-rodo> (26.05.2024).

<https://respeecher.medium.com/deepfake-or-synthetic-voice-whats-the-difference-43d84895b0a3> (27.05.2024).

<https://rsilpak.org/2024/deepfakes-a-crisis-of-human-rights/> (29.05.2024).

<https://screenrant.com/furious-7-brian-scenes-not-paul-walker-brothers/> (20.05.2024).

<https://towardsdatascience.com/deepfakes-the-ugly-and-the-good-49115643d8dd> (20.05.2024).

<https://us.norton.com/blog/emerging-threats/what-are-deepfakes> (19.05.2024).

<https://www.bbc.com/news/technology-60780142> (23.05.2024).

<https://www.buzzfeed.com/craigsilverman/obama-jordan-peelee-deepfake-video-debunk-buzzfeed> (26.05.2024).

<https://www.cbsnews.com/news/tupac-coachella-hologram-behind-the-technology/> (19.05.2024).

<https://www.darpa.mil/program/media-forensics> (25.05.2024).

<https://www.edweek.org/leadership/deepfakes-expose-public-school-employees-to-new-threats/2024/05> (21.05.2024).

<https://www.gtlaw.com.au/knowledge/australian-government-targets-sexually-explicit-deepfakes> (26.06.2024).

<https://www.mediatropy.com/tech/how-ai-deepfakes-are-transforming-the-world-of-marketing-and-advertising/> (22.05.2024).

<https://www.nbcnews.com/tech/internet/taylor-swift-deepfake-x-falsely-depict-supporting-trump-grammys-flag-rcna137620> (24.05.2024).

<https://www.nsa.gov/Press-Room/Press-Releases-Statements/Press-Release-View/Article/3523329/nsa-us-federal-agencies-advise-on-deepfake-threats/> (28.05.2024).

<https://www.realitydefender.com/> (22.05.2024).

<https://www.reddit.com/r/SFWdeepfakes/?rdt=50851> (18.05.2024).

<https://www.restack.io/p/ai-deepfakes-answer-deepfake-historical-figures-cat-ai> (21.05.2024).

<https://www.scoredetect.com/blog/posts/legal-remedies-for-deepfake-victims-guide> (25.05.2024).

<https://www.smh.com.au/technology/the-deepest-fake-how-new-tech-will-test-our-belief-in-what-we-see-20180423-p4zb4w.html> (22.05.2024).

<https://www.tieonline.com/article/3632/the-rise-of-deepfakes-in-schools> (21.05.2024).

<https://www.unr.edu/nevada-today/news/2023/atp-deepfakes> (27.05.2024).

Technologie deepfake w zakresie danych osobowych

Streszczenie

Manipulacja informacją w środkach masowego przekazu nie jest zjawiskiem nowym. Wraz z rozwojem coraz nowszych technologii zjawisko to wchodzi na niebezpiecznie wysoki poziom, przez co zwiększa się ryzyko pojawienia się dezinformacji w social mediach. Możliwości, jakie obecnie daje Internet, są często wykorzystywane do rozpowszechniania nieprawdziwych, krzywdzących informacji na temat różnych organizacji i osób prywatnych. Rozwijanie nowych technologii często wykorzystywane jest w sposób niezgodny z pierwotnym zamysłem ich kreatorów. Tak też z początkiem poprzedniej dekady dużym zagrożeniem w mediach stała się technologia deepfake, która przy pomocy sztucznej inteligencji potrafi przekształcić obraz człowieka nawet do formy wideo. Proces ten polega na dokładnej analizie twarzy w celu połączenia ich lub zmienienia ze sobą; wszystko to dzieje się przy pomocy sztucznej inteligencji. Doskonalenie tego rodzaju technologii wykracza aktualnie poza sferę rozrywkową, gdyż technologie wykorzystywane są coraz częściej do podszywania się pod kogoś w Internecie lub do próby znieważenia osoby prywatnej. Przy czym należy pamiętać, że wizerunek danej osoby należy do danych osobowych tej konkretnej jednostki. W związku z tym wizerunek podlega ochronie. Technologie pozwalające na podszywanie się pod konkretną osobę lub na bezprawne wykorzystanie jej wizerunku prowadzą w konsekwencji do kradzieży tożsamości, oszustw politycznych, manipulacji wyborczych, rozszerzania się fake newsów czy też do popularyzacji formy szantażu, jaką jest *sextortion*, gdzie sprawca grozi publikacją nagich filmów lub zdjęć, które stworzone zostały przy użyciu AI.

Słowa kluczowe: fałszywe obrazy, nowe technologie, informacje pomagające przy identyfikacji osób fizycznych, manipulacja informacją, media społecznościowe, sztuczna inteligencja

Deepfake Technology in the Context of Personal Data

Abstract

Manipulation of information in the mass media is not a new phenomenon. Unfortunately, with the development of new technologies, this issue has reached a dangerously high level, significantly increasing the risk of disinformation on social media. The current opportunities offered by the Internet are often exploited to spread false and harmful information about various organizations and private individuals. Technological advancements are frequently used in ways that deviate from the original intentions of their creators. As a result, at the beginning of the last decade, *deepfake* technology emerged as a major threat in the media. This technology can transform a human image – even

into a realistic video – by using artificial intelligence. The process involves a detailed analysis of human faces to merge or alter them, all powered by AI. The refinement of such technologies has moved beyond the entertainment sphere; they are increasingly used to impersonate individuals online or to attack private persons. It is important to note, however, that a person's image constitutes personal data and, as such, is legally protected. Technologies that enable impersonation or the unlawful use of someone's likeness can lead to serious consequences, such as fraud, identity theft, political manipulation, electoral interference, the spread of fake news involving public figures, or even the rise of blackmail practices like *sextortion*, where perpetrators threaten to publish fabricated nude images or videos created using AI.

Keywords: fake images, new technologies, information that helps identify individuals, information manipulation, social media, artificial intelligence (AI)