Małgorzata Gradek Lewandowska[1],
Anna Dębska[2]
Anna Świderska[3]

O technologii deepfake i prawach własności intelektualnej

(Trzecia część cyklu: „Technologie jutra już dziś w życiu rzecznika patentowego”[4])

Wstęp

Po miesiącu trwania wojny w Ukrainie, w Internecie pojawiło się nagranie, w którym prezydent Zełenski namawia Ukrainę do poddania się. W filmie ukraiński przywódca ogłasza kapitulację i informuje o oddaniu Donbasu Federacji Rosyjskiej. Co stałby się, gdyby Ukraińcy nie byli przygotowani na taką prowokacje ze strony Rosji? W październiku 2022 r. szef tureckiej firmy produkującej drony, Haluk Bayraktar, otrzymał połączenie rzekomo od premiera Ukrainy, Denysa Szmyhala. Tak naprawdę byli to rosyjscy prowokatorzy, którzy przy użyciu deepfake chcieli doprowadzić do osłabienia relacji turecko-ukraińskich. Na szczęście próba ta została udaremniona.[5]

Prowokacje z wykorzystaniem deepfake znane były wcześniej. Przykładem może być wyemitowanie w Belgii w maju 2018 r. deepfake pokazującego Donalda Trumpa wzywającego USA do wycofania się z porozumienia klimatycznego z Paryża. Zamiarem belgijskiej partii Sp.a Party emitującej ten film, było przyciągnięcie uwagi widzów, a następnie przekierowanie ich do internetowej petycji wzywającej belgijski rząd do podjęcia pilniejszych działań na rzecz klimatu. Chociaż był to

1 Małgorzata Gradek Lewandowska - radca prawny, rzecznik patentowy

2 Anna Dębska - rzecznik patentowy

3 Anna Świderska - adwokat

4 Pierwsza część tego cyklu została opublikowana w numerze 1/2022 (str.61-72), a druga część - w numerze 2/2022 (str.73-84). Kolejny artykuł będzie dotyczył IoT (Internet of Thing)

5 https://businessinsider.com.pl/wiadomosci/deepfake-rosjanie-podszyli-sie-pod-premiera-ukrainy-zadzwonili-do-bayraktara-padlo/56l02fp

fałszywy film, niektórzy widzowie byli wstrząśnięci i żądali, aby partia potwierdziła oficjalnie, że nagranie było zmontowane[6]. Twórcy filmu powiedzieli później, że założyli, iż zła jakość deepfake wystarczy, aby widzowie byli przekonani o braku autentyczności: „z ruchów warg jasno wynika, że to nie była autentyczna przemowa Trumpa" – powiedział Politico rzecznik Sp.a Party. W połowie lutego 2020r. na WhatsApp pojawił się film Manoj Tiwari, prezydenta Indii Bharatiya Janata Party. W filmie Tiwari przemawia przekonująco w języku haryanvi, dialekcie hindi używanym przez jego potencjalnych wyborców[7]. Wideo jest fałszywe: w oryginalnym nagraniu Tiwari mówi po angielsku. Jest to kolejny przykład tego jak partia polityczna wykorzystała technologię deepfake do celów politycznych.

Co prawda bez wykorzystania warstwy wizualnej, ale oszustwa poprzez stworzenie nagrania imitującego głos innej osoby, nie są zarezerwowane jedyne dla polityków. Dla przykładu: w 2019 r. dyrektor pewnej brytyjskiej firmy z branży energetyki dokonał przelewu na ponad 200 tys. euro po tym, jak otrzymał telefon od osoby rzekomo zarządzającej spółką nadrzędną (spółką-matką). Z relacji The Wall Street Journal wynika, że dyrektor został poproszony o pilne przesłanie pieniędzy do węgierskiego dostawcy[8]. Ofiara zdarzenia twierdzi, że głos był identyczny. Zawierał nawet delikatny niemiecki akcent oraz intonacje charakterystyczną dla tej osoby. W tym przypadku, płatność z węgierskiego rachunku została natychmiast przelana do Meksyku i dalej. To ponoć pierwszy przypadek wykorzystania deepfake voice (speech synthesis) w celu popełnienia przestępstwa w Europie, ale patrząc na "udany" efekt, z pewności nie ostatni[9] Co więcej, pojawiają się już pierwsze ostrzeżenia przed nowym rodzajem wyłudzania „na wnuczka", właśnie poprzez wykorzystywanie czyjegoś głosu i podszywanie się w ten sposób przez telefon pod bliską nam osobę w celu wyłudzenia pieniędzy. Czym jest, więc deepfake? Zwłaszcza w kontekście własności intelektualnej.

O technologii deepfake możliwie najprościej

Definicję „deepfake" zaproponował pod koniec 2018 roku republikański senator Ben Sasse z Nebraski w ramach ustawy znanej jako "Malicious Deep Fake Prohibition Act of 2018"[10], jako „nagranie audiowizualne stworzone lub zmienione w taki sposób, że nagranie mogłoby fałszywie wydawać się rozsądnemu obserwatorowi autentycznym zapisem rzeczywistego przemówienia lub zachowania danej osoby". W takiej definicji podkreślony jest dezinformujący, wprowadzający w błąd charakter przekazu.

6 Schwartz, O. (2018) You thought fake news was bad? deepfakes are where truth goes to die [article] The Guardian, 12 listopada 2018

7 https://inc42.com/buzz/bjp-makes-deepfake-debut-with-manoj-tiwari-video-for-delhi-elections/

8 https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

9 https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

10 https://www.congress.gov/bill/115th-congress/senate-bill/3805/text

Deepfake to technika syntezy ludzkiego obrazu oparta na sztucznej inteligencji. Rozwijając to hasło, deepfake to zbitka wyrazowa pochodząca od słów z języka angielskiego „*deep learning*," czyli głębokie uczenie oraz „*fake"*, czyli fałszywy, imitacja. Technologia bazuje na możliwości łączenia i nakładania na wyjściowe, źródłowe obrazy kolejnych obrazów. A precyzyjniej: jest to wykorzystanie sztucznej inteligencji używanej do tworzenia lub modyfikowania obrazu odwzorowanej twarzy, do tworzenia ultra-realistycznych fałszywych filmów, w których ludzie mówią i robią rzeczy, w rzeczywistości nie mające miejsca.[11] Aby przekonać się czym jest deepfake warto obejrzeć na YouTube film prezentujący rzekome wystąpienie Baracka Obamy.[12]

Do tej pory rozróżniamy cztery rodzaje deepfake-ów:

1. **face reenactment** – z ang. face – twarz, reenactment – przywrócenie, rekonstrukcja – sterowanie mimiką danej osoby (przykładem film Hillary Clinton's face is digitally 'stitched' onto Saturday Night Live actress Kate McKinnon, źródło: Derpfakes)
2. **faceswap** – z ang. face – twarz, swap – zamiana, dosłownie zamiana twarzy jednej osoby na twarz innej osoby (przykładem film źródło: The Visual Computing Lab at TUM)
3. **face generation** – polega na tworzeniu zupełnie nowych obrazów twarzy nieistniejących w rzeczywistości, Generative Adversarial Networks jest platformą uczenia maszynowego, która działa poprzez przeciwstawienie sobie dwóch sieci neuronowych: pierwsza generuje obraz, a druga ocenia czy ten wynik jest realistyczny (przykłady na ThisPersonDoesNotExist.com – jest to popularna witryna internetowa, która za każdym razem generuje nową twarz osoby, która faktycznie nie istnieje)
4. **speech synthesis** – synteza mowy polega na stworzeniu modelu czyjegoś głosu, który może odczytać tekst w taki sam sposób, jak osoba docelowa (przykładem narzędzie Modulate.ai)

Podsumowując: deepfake to zespół algorytmów składający się na program komputerowy służący do wizualnych przeróbek głosu i twarzy.

Pojawienie się deepfake w 2017 r. nie powinno być zaskoczeniem. Skoro tworzone są oprogramowania, które nie wymagają wielu danych, aby program mógł „nauczyć się" mimiki czy sposobu poruszania osoby (do tego w zupełności wystarczą dwa pliki: docelowy, do którego chcemy wprowadzić zmiany i drugi, od którego chcemy pozyskać ruch). Jednocześnie istnieją również technologie, które po kilkusekundowej próbce głosu danej osoby mogą odczytywać dowolny tekst głosem tej osoby. Ostatni element, czyli dopasowanie ruchu ust pod daną ścieżkę audio

11 Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość, Olga Wasiuta, Sergiusz Wasiuta, Studia de Securitate 9(3) (2019) ISSN 2657-8549.

12 https://www.youtube.com/watch?v=AmUC4m6w1wo

również nie stanowi problemu. Możliwe zatem stało się stworzenie filmu, w którym dana osoba będzie mówiła to co chcemy, a czego nigdy nie wypowiedziała naprawdę. Znanym przykładem jest film wykorzystujący wizerunek Baracka Obamy[13,14].

Powstaje zatem pytanie, czy w tym przypadku mamy do czynienia z twórczością? Czy próbując dokonać analizy deepfake-ów pod kątem prawnym należy odnieść się do prawnych regulacji dotyczących procesu tworzenia, czyli możliwego powstania dób chronionych prawami własności intelektualnej. Oczywiście nie mamy obecnie przepisów prawa regulujących wprost deepfake (zmiany prawa nie nadążają za tempem postępu technologicznego ostatnich lat), ale analizując sposób powstawania filmów oraz ich wykorzystywania, można wskazać obecnie istniejące normy prawne, które mogą mieć w takim przypadku zastosowanie.

Deepfake i prawo autorskie

Na podstawie przepisów prawa autorskiego[15] możemy chronić program komputerowy do tworzenia deepfake-ów. Zgodnie z art. 74 prawa autorskiego programy komputerowe podlegają ochronie jak utwory literackie, z uwzględnieniem innych pól, zakresu dozwolonego użytku publicznego czy praw osobistych twórców programu komputerowego

Autorskie prawa majątkowe do programu komputerowego obejmują prawo do:

- reprodukcji (zwielokrotnienia) programu, w całości lub części, jakimikolwiek środkami i w jakiejkolwiek formie;
- tłumaczenia, przystosowania (adaptacji) programu, zmiany układu oraz wprowadzania innych zmian w programie
- rozpowszechniania, w tym użyczania lub najmu programu lub jego kopii.
-

Program do tworzenia deepfake-ów jest chroniony prawem autorskim i można z niego korzystać jeśli osoba, która napisała (stworzyła) ten program udostępni go np. na zasadzie otwartych licencji. Z jednej strony oznacza to możliwość udoskonalania programów do tworzenia deepfake (może dojść wtedy do powstania utworów zależnych), a tym samym do lepszej jakości tak powstających filmów – powodując ciągły rozwój a tym samym utrudnienie wykrycia deepfake. Z drugiej strony, udostepnienie programu do tworzenia deepfake, daje możliwość praktycznie każdemu korzystania z takiego programu komputerowego i stworzenia własnej postaci deepfake-owej.

13 Mack, D. (2018) This PSA About Fake News From Barack Obama Is Not What It Appears [article] Buzzfeed News, 17 kwietnia 2018.

14 https://www.youtube.com/watch?v=AmUC4m6w1wo

15 Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz. U. z 2022 r. poz. 2509).

Oznacza to, że możliwe będzie również wykorzystanie tak stworzonej postaci deepfake-owej do celów komercyjnych, a nie tylko prywatnych. Zatem, zakładając legalność korzystania z programu komputerowego, ważne jest kto, a raczej z jakim zamiarem i w jaki sposób będzie tworzył deepfake-owe filmy i postaci.

Zucconi wymienia 3 etapy powstania deepfake[16]:

- **ekstrakcja** – zebranie wystarczającej liczby obrazów do trenowania modelu zastępujący twarz
- **szkolenie** – wyszkolenie modelu zastępującego twarz przy użyciu zebranych obrazów za pomocą autoenkodera, którym jest sieć neuronowa składająca się z dwóch części: kodera i dekodera (koder pobiera obraz twarzy i kompresuje go do małego wymiaru a następnie dekoder rekonstruuje twarz do jej pierwotnej formy)
- **tworzenie** – najtrudniejszy technicznie etap polegający na wstawieniu fałszywych obrazów do wideo (dla każdej klatki wideo kąt syntetyzowanej twarzy musi pasować do kąta głowy osoby docelowej), wg Zucconiego to jedyny etap procesu, który opiera się na ręcznie napisanym kodzie, a nie na maszynowym i dlatego jest podatny na więcej błędów. Tworzenie to przypomina cyfrowy kolaż składający się z różnych elementów. Żeby go stworzyć trzeba korzystać z nagrań wizerunków osoby, jej głosu oraz nagrania osoby, która „użycza" swojego ruchu tworzonej postaci.

Czy zatem deepfake rozumiany jako film to będzie „utwór", „utwór audiowizualny"? Zgodnie z art. 1 prawa autorskiego, utworem jest każdy przejaw działalności twórczej człowieka. Trzeba zatem ustalić czy film deepfake stanowi rezultat działania człowieka. Zakładając, że co najmniej trzeci etap tworzenia deepfake wydzielony przez Zucconiego zależy od człowieka to dla dalszej analizy możemy przyjąć, iż rezultat w postaci filmu zostaje stworzony przez człowieka. Jednak to założenie będzie błędne, gdy rezultat w postaci filmu deepfake jest wyłącznym efektem działalności szablonowej, w przystosowanych do tego generatorach, automatycznie bez aktu kreacji ze strony człowieka - wówczas do powstania utworu nie dojdzie (nie zostanie spełniona przesłanka twórczej działalności człowieka).

Dla dalszej analizy skoncentrujmy się wyłącznie na pierwszym przypadku (deepfake jako efekt działalności człowieka). W tej sytuacji trzeba ustalić czy połączenie, a raczej przekształcenie zebranych obrazów twarzy, zmiana jej mimiki oraz „włożenie" w usta tak powstałej postaci innego tekstu powoduje, że praca nad tworzeniem deepfake ma twórczy, indywidualny charakter. Przyjmuje się, że „cecha twórczości jest spełniona wówczas, gdy istnieje subiektywnie nowy wytwór

16 Zucconi, A. (2018) An introduction to DeepFakes [blog] AZ website, 14 marca 2018.

intelektu”[17]. Prof. J. Błeszyński, uważa, że utwór musi być „osobistą wizją” autora, tzn. „musi polegać na uformowaniu własnego obrazu, zjawiska, procesu, idei”.[18]

Generowanie deepfake-ów odbywa się za pomocą programów komputerowych, co nie tylko wymaga pewnych umiejętności, ale również wcześniej zamierzonej wizji autora. Zgodnie z wyrokiem Sądu Najwyższego z dnia 25 stycznia 2006 r.[19] „utworem może być nawet kompilacja wykorzystująca dane powszechnie dostępne pod warunkiem, że ich wybór, segregacja, sposób przedstawienia ma znamiona oryginalności”. Podobnie Sąd Apelacyjny w Warszawie w wyroku z 26 listopada 2014 r. uznał, że „spełnienia przesłanki twórczości (oryginalności i indywidualności) utworu można doszukiwać się nawet w doborze, układzie lub uporządkowaniu składników utworu, tj. poszczególnych ujęć czy nawet doborze podkładu muzycznego”[20]. Zatem realizacja wizji autora deepfake przemawia za zakwalifikowaniem deepfake-ów do utworów, z tym zastrzeżeniem, iż być może nie każdą modyfikację lub fotomontaż, w rezultacie której powstaje deepfake, będzie można uznać za utwór z powodu braku indywidualności.

Na pewno twórcy deepfake muszą korzystać z innych nagrań z wizerunków osób i muszą mieć drugie nagranie z „podstawionym” głosem: wideogram lub fonogram. Aby skorzystać z cudzego wideogramu lub fonogramu musimy mieć na to zgodę tego komu przysługuje prawo korzystania z wideogramu lub fonogramu. W prawie autorskim przewidziane są jednak wyjątki, które zezwalają na włączanie części cudzego utworu lub jego przeróbkę.

Jedną z takich „instytucji” jest prawo cytatu. Ustawodawca wskazał w art. 29 prawa autorskiego, iż wolno przytaczać w utworach stanowiących samoistną całość urywki rozpowszechnionych utworów oraz rozpowszechnione utwory plastyczne, utwory fotograficzne lub drobne utwory w całości, w zakresie uzasadnionym celami cytatu, takimi jak wyjaśnianie, polemika, analiza krytyczna lub naukowa, nauczanie lub prawami gatunku twórczości. Z art. 100 prawa autorskiego wynika, że wykonywanie praw do artystycznych wykonań, fonogramów, wideogramów, nadań programów, podlega odpowiednio ograniczeniom wynikającym z przepisów art. 23-35. Tym samym wykorzystywanie przez twórcę deepfake fragmentu innego utworu, wideogramu lub fonogramu może być uzasadnione prawem cytatu.

Inną możliwą konstrukcją jest przyjęcie, że twórca deepfake dokonał opracowania utworu pierwotnego. I tutaj ważna jest regulacja art. 2 prawa autorskiego. Zgodnie z tym przepisem rozporządzanie i korzystanie z opracowania zależy od

17 Tak w uzasadnieniu wyroku Sądu Najwyższego z dnia 22 czerwca 2010r. IV CSK 359/09, https://sip.lex.pl/#/jurisprudence/520787965?keyword=subiektywnie%20nowy

18 J. Błeszyński „Prawo autorskie” Warszawa 1985 r. s. 40

19 Sygn.. I CK 281/05, OSN 2006, nr 11, poz. 186

20 Sygn. VI AC 212/14, Legalis

zezwolenia twórcy utworu pierwotnego (prawo zależne), chyba że autorskie prawa majątkowe do utworu pierwotnego wygasły. Dodatkowo na egzemplarzach opracowania należy wymienić twórcę i tytuł utworu pierwotnego. Czyli można zrobić „sobie" deepfake dla zabawy „do szuflady", ale na jego pokazywanie – jeśli deepfake będzie stanowił opracowanie - trzeba uzyskać zgodę twórcy materiału pierwotnego.

Najciekawsza jest jednak analiza tworzenia deepfake-ów w kontekście prawa do korzystania z cudzego utworu w ramach parodii, pastiszu i karykatury. W 2015 r. rozszerzono ten przepis o art. 29(1) prawa autorskiego podkreślając tym samym specyfikę takich praw gatunków twórczości, jak parodia, pastisz czy karykatura jako gatunków szczególnych. Nie powinno to nikogo dziwić, biorąc pod uwagę, że wraz z rozwojem technologii pojawiają się także nowe gatunki twórczości, a niektóre nabierają zupełnie innego wymiaru. Każdy z nas chyba kojarzy z wakacji artystów siedzących na deptaku i rysujących karykatury znanych osób lub nas, czyli świadome przedstawienie osoby, zjawiska lub przedmiotu w sposób zdeformowany, które ma na celu podkreślenie negatywnego charakteru elementów, właściwości, cech. Na pewno każdy przypadek trzeba będzie oceniać indywidulanie. Zapewne nie zawsze sądy będą skłonne uznawać, że ingerencja osoby produkującej deepfake w prawa autorskie twórcy utworu macierzystego mieści się w graniach parodii, pastiszu i karykatury. To jest o tyle istotne, że parodię, pastisz oraz karykaturę zalicza się głównie do utworów inspirowanych, ponieważ autor takiego dzieła chce przekazać coś nowego i odmiennego od tego, co było zawarte w utworze pierwotnym. Wówczas nie ma wymogu uzyskania zgody autora utworu macierzystego, ani obowiązku zamieszczania informacji wynikających z art. 2 ust. 5 prawa autorskiego, czyli wymieniania twórcy i tytułu utworu pierwotnego na egzemplarzach opracowania.

Zarówno w przypadku korzystania z prawa cytatu, jak i w przypadku tworzenia opracowania czyjegoś utworu konieczne jest wskazanie źródła, a z tym spotykamy się bardzo rzadko. Podawanie źródeł utworów to generalnie duży problem twórczości internetowej.

Podsumowując. nie sposób uznać, że deepfake zawsze będą traktowane jako utwory w rozumieniu art. 1 ust. 1 prawa autorskiego. Gdyby jednak uznać konkretny deepfake za utwór, to należy ocenić czy mamy do czynienia z utworem samoistnym, czy też z utworem niesamoistnym, co z punktu widzenia wykorzystania cudzej twórczości będzie miało istotne znaczenie. W tym drugim przypadku powoduje powstanie obowiązku z art. 2 ust. 2 i 5 prawa autorskiego: uzyskania zgody na korzystanie i rozpowszechnianie deepfake i wymienienia twórcy utworu pierwotnego. To wydaje się bardzo problematyczne z uwagi na rozwój narzędzi technologicznych, dzięki którym możliwe jest przerabianie, łączenie czy też miksowanie cudzych utworów oraz późniejsze ich rozpowszechnianie na masową skalę.[21]

21 K. Grzybczyk, [w:] P. Ślęzak, Ustawa…, komentarz do art. 2 PrAut, Legalis/el. 2017

Nie można wykluczyć, że utwory inspirowane mogą stanowić parodię, pastisz lub karykaturę, a wtedy nie ma potrzeby korzystania z licencji ustawowej opisanej w art. 29 (1) prawa autorskiego. Według E. Czarny-Drożdżejko „przy ocenie zachowania osoby wykorzystującej cudzy utwór istotna będzie ocena zamiaru, który jej przyświecał. Musi on bowiem podejmować eksploatację cudzego utworu w celu humorystycznym, a nie innym”.[22]

Deepfake a regulacje kodeksu cywilnego

Poza prawem autorskim, kolejna regulacja ściśle powiązana z deepfake-ami to art. 23 i 24 kodeksu cywilnego. Na podstawie tych przepisów chronione są dobra osobiste człowieka. Z deepfake-ami największy związek będą miały wizerunek i głos. Pomimo, że głos nie został wyszczególniony w art. 23 kodeksu cywilnego, to w orzecznictwie potwierdzono pogląd, że głos jest objęty ochroną kodeksową[23].

O tym, czy naruszenie dobra osobistego zostało istotnie dokonane, decyduje obiektywna ocena konkretnych okoliczności faktycznych, nie subiektywne odczucie osoby zainteresowanej. O bezprawności naruszenia zaś może być mowa wtedy, kiedy to cudze działanie narusza jakieś konkretne nakazy i zakazy. Oczywiście, w pojęciu bezprawności czynności nie mieszczą się natomiast omyłki, niedokładności, przeinaczenia i inne wadliwości ale pozbawione cech ocennych wobec określonej osoby[24].

Przy ochronie dobra osobistego jakim jest wizerunek, trzeba wrócić na chwilę do prawa autorskiego. Zgodnie z art. 81 prawa autorskiego, rozpowszechnianie wizerunku wymaga zezwolenia osoby na nim przedstawionej. Zezwolenia nie wymaga rozpowszechnianie wizerunku:

- osoby powszechnie znanej, jeżeli wizerunek wykonano w związku z pełnieniem przez nią funkcji publicznych, w szczególności politycznych, społecznych, zawodowych;
- osoby stanowiącej jedynie szczegół całości takiej jak zgromadzenie, krajobraz, publiczna impreza
- osoby, która otrzymała umówioną zapłatę za pozowanie.

Wizerunek znanej osoby utrwalony podczas wykonywania przez nią obowiązków służbowych to idealna definicja polityka. Trzeba jednak pamiętać, by nie przekroczyć przy tym granicy, która będzie naruszała nie tylko wizerunek, ale także inne dobra osobiste np. godność, dobre imię. Za przykład mogą tutaj posłużyć coraz częściej tworzone za pomocą technologii deepfake filmy pornograficzne, w których gwiazdy porno mają twarze znanych osób – aktorów, muzyków, a nawet polityków.

22 E. Czarny-Drożdżejko, Dozwolony…, s. 204.

23 Wyrok Sądu Apelacyjnego w Gdańsku z dnia 21.06.1991 r. sygn. I ACr 127/91.

24 Wyrok Sądu Apelacyjnego w Gdańsku z dnia 18.02.2020 r. sygn. V ACa 606/19.

Można zatem wyobrazić sobie, że twórcy deepfake będą powoływać się na powyższe wyjątki. Pytanie: czy faktycznie mogą? Jeśli przekroczą granice wyjątku uzyskania zgody na rozpowszechnianie wizerunku, to osoba, której to dobro osobiste zostało zagrożone czy naruszone poprzez wykorzystanie w deepfake może żądać zarówno na podstawie art. 83 i art. 78 ust. 1 prawa autorskiego jaki i art. 24 kodeksu cywilnego:

- wstrzymania tego działania
- ażeby osoba, która dopuściła się naruszenia, dopełniła czynności potrzebnych do usunięcia jego skutków, w szczególności ażeby złożyła oświadczenie odpowiedniej treści i w odpowiedniej formie;
- zadośćuczynienia pieniężnego lub zapłaty odpowiedniej sumy pieniężnej na wskazany cel społeczny;
- naprawienia szkody na zasadach ogólnych, jeżeli wskutek naruszenia dobra osobistego została wyrządzona szkoda majątkowa.

Wydaje się, że to te przepisy będą ochronić osoby, których wizerunek został wykorzystany w celu stworzenia deepfake. Trzeba jednak podkreślić, że aby móc uzyskać ewentualne odszkodowanie lub zadośćuczynienie trzeba wskazać osobę, która wyrządziła szkodę lub krzywdę. I to w przypadku twórców deepfake może być bardzo problematyczne.

To, co potocznie nazywane jest naruszeniem prawa osobistego do wizerunku, w praktyce oznacza naruszenie innego dobra (czci, prywatności, kultu pamięci osoby zmarłej, “persony”), tyle, że za pośrednictwem wizerunku. Argument, który najprawdopodobniej może być wykorzystywany przez twórców deepfake-ów, czyli wspomniany wyjątek od uzyskania zezwolenia na rozpowszechnianie wizerunku osoby powszechnie znanej. Zezwolenie takie przysługuje tylko, gdy rozpowszechniamy wizerunek osoby powszechnie znanej utrwalony w związku z wykonywaniem przez nią swojej funkcji publicznej czy zawodowej. Każde inne utrwalenie i rozpowszechnienie wizerunku, także osób znanych, które nie powstało w związku z pełnieniem przez nich ich funkcji publicznych będzie naruszeniem ich prawa do prywatności.

Jednakże nawet gdyby uznać, że w tym wypadku nie przekroczono granic prywatności, warto rozważyć, czy nie zostały naruszone inne dobra osobiste takiej osoby. Naruszeniem tych dóbr, godności może być np. narażenie tej osoby na krytykę ze strony osób trzecich, które mogłyby uznać za prawdziwe takie nagranie. Także narażenie takiej osoby na utratę zaufania publicznego w związku z wykonywanym przez nią stanowiskiem będzie godzić w jej dobra osobiste, tym samym dając jej prawo do obrony przed takim bezprawnych zachowaniem.

Karne aspekty prawa autorskiego

Sięgając do przepisów kodeksu karnego można zastanowić się, czy w niektórych przypadkach samo stworzenie deepfake będzie wypełniało znamiona przestępstw przeciwko czci i nietykalności cielesnej takich jak zniesławienie i zniewaga, czy raczej przestępstwem będzie publikowanie deepfake w Internecie.

Zniesławienie, stypizowane w art. 212 Kodeksu karnego dotyczy pomawiania innej osoby, grupy osób, instytucji, osoby prawnej lub jednostki organizacyjnej niemającej osobowości prawnej o takie postępowanie lub właściwości, które mogą poniżyć ją w opinii publicznej lub narazić na utratę zaufania potrzebnego dla danego stanowiska, zawodu lub rodzaju działalności podlega grzywnie albo karze ograniczenia wolności. Jeżeli sprawca dopuszcza się czynu zniesławienia za pomocą środków masowego komunikowania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do roku. Analiza przepisu od razu nasuwa pytanie, czy stworzenie deepfake albo jego rozpowszechnianie może stanowić pomówienie. W związku z tym, że nie istnieje precyzyjna definicja pomówienia nie sposób przesądzić czy deepfake stanie się również formą pomówienia. Poprzez pomówienie można rozumieć przypisanie komuś w publicznej wypowiedzi cech, poglądów lub czynów, które psują mu dobrą opinię, powodują utratę zaufania do niego, a przecież to nie twórca deepfake wypowiada się publicznie.

W art. 216 kodeksu karnego określono przestępstwo zniewagi, czyli: kto znieważa inną osobę w jej obecności albo choćby pod jej nieobecność, lecz publicznie lub w zamiarze, aby zniewaga do osoby tej dotarła, podlega grzywnie albo karze ograniczenia wolności, także kto znieważa inną osobę za pomocą środków masowego komunikowania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do roku. Przestępstwo zniewagi przejawia się w ubliżaniu komuś słowem lub czynem i stanowi ciężką obrazę skierowaną przeciwko godności osobistej człowieka[25].

Katalog sposobów popełnienia czynu znieważenia nie został określony, zatem można dojść do wniosku, że dotyczy to rozmaitego rodzaju zachowania, których wspólną cechą jest to, że wyrażają pogardę dla godności drugiego człowieka. Znieważające zachowanie może przybrać postać słowną (posłużenie się wulgarnym słownictwem), być wyrażone za pomocą rysunku (np. karykatura), symboli, znaków albo innego rodzaju środka przekazu (film, fotografia) czy też gestu, który wyraża brak szacunku wobec drugiej osoby. Przyjmując, że deepfake to film stworzony przez jego twórcę, który poprzez dokonanie zamiany twarzy będzie wyrażał pogardę godności osoby, której twarz została „doklejona" można pokusić się o stwierdzenie, że przestępstwo zniewagi będzie w tym wypadku odpowiednie. Jednak nie jest

25 Uchwała Sądu Najwyższego z dnia 12.10.2017 r. sygn. SNO 32/17.

pewne czy pogarda godności człowieka będzie miała charakter obiektywny, czy deepfake w większości przypadków pozostaną jedynie znieważające w ujęciu subiektywnym osoby, której wizerunek został w deepfake wykorzystany.

Zgodnie z kodeksem karnym sprawcą jest ten kto popełnia czyn niedozwolony. Ustalenie, kto jest odpowiedzialny za efekt końcowy deepfake wymaga analizy i pogłębienia wiedzy na temat tego, kto jest twórcą. Problematyczne może być korzystanie przez twórcę z programu będącego dziełem innej osoby. Powstaje wtedy pytanie, czy twórca programu komputerowego zezwalającego na tworzenie deepfake-ów może odpowiadać jako współsprawca, podżegacz albo pomocnik przestępstwa? Zgodnie z kodeksem karnym odpowiada za pomocnictwo, kto w zamiarze, aby inna osoba dokonała czynu zabronionego, swoim zachowaniem ułatwia jego popełnienie, w szczególności dostarczając narzędzie, środek przewozu, udzielając rady lub informacji.

Na pewno odpowiedzialność będzie ponosić osoba, która intencjonalnie korzysta z deepfake, np. aby wymusić niekorzystne oświadczenie czy podjęcie decyzji pod wpływem treści deepfake. Czyli może być tak, ze deepfake jest tworzony dla żartu, ale jego użycie wymyka się spod kontroli twórcy i finalnie służy popełnieniu przestępstwa.

Po pojawianiu się coraz doskonalszych deepfake-ów rozpoczęto projekty tworzenia programów wykrywających deepfake, które zajmują się oceną integralności zdjęć lub filmów, wykrywaniem manipulacji. Niestety ulepszanie oprogramowania do wykrywania deepfake-ów przyczynia się również do tworzenia coraz lepszych deepfake-ów. Innymi słowy programy uczą się od programów jak oszukiwać programy. Skoro aplikacje służące wykrywaniu fake newsów powodują, że stają się one jeszcze bardziej dopracowane nasuwa się myśl, by zawsze stosować zasadę ograniczonego zaufania. Żyjemy czasach, w których technologia idzie do przodu i nie pozostaje nam nic innego jak nauczenie się korzystania z niej. Wbrew pozorom, zaletą fake newsów może być zmniejszenie zaufania ludzi do treści znajdujących się w Internecie. Jeżeli dostajemy wiadomości z nieznanego źródła, zawsze warto jest je zweryfikować. Prawdopodobieństwo, że duży serwis przekazuje fałszywe treści jest dużo mniejsze, niż nikomu nieznany serwis, za którym nie stoi żaden człowiek.

Deepfake „dla ludzkości” i z ludzkim nadzorem

Na szczęście, technologia deepfake wykorzystywana jest także do innych celów niż tylko do popełniania przestępstw i wprowadzania celowo w błąd odbiorców. Studio badawcze należące do Disneya (Disney Research Studio), stworzyło swój własny algorytm pozwalający na w pełni automatyczną wymianę twarzy

neuronowych w obrazach i filmach w wysokiej rozdzielczości, umożliwiające wykorzystanie na dużym ekranie. To czym różni się algorytm Disneya to zwiększanie liczby megapikseli, dzięki czemu możliwe jest tworzenie filmów w rozdzielczości 1024 x 1024 pikseli czyli lepsza widoczność na większych ekranach. Disney koncentruje się na zapewnieniu stabilizacji obrazu oraz doborze oświetlenia na podstawie tła i obiektu głównego. Oczywiście, Studio jest zainteresowane tą technologią z komercyjnego względu. Stworzenie deepfake-owych postaci pozwoli na „przywrócenie na srebrny ekran aktorów, którzy zmarli bądź z powodu wieku lub choroby nie są w stanie zagrać w kolejnej części hitowej serii. Użycie wizerunku kultowego aktora w nowoczesnej produkcji mogłoby być także sposobem na oddanie mu hołdu i uczczenie jego kinowych dokonań”[26].

Dzięki temu możliwe było dokończenie prawie gotowego filmu Gwiezdne wojny: Skywalker. Odrodzenie po śmierci Carrie Fisher grającej księżniczkę Leię. Przywracanie zmarłych aktorów do życia dla sequeli filmowych zostało nazwane jako „zmartwychwstanie cyfrowe”. Są także pierwsze próby wykorzystania deepfake w szeroko rozumianej medycynie. Syntezator mowy Lyrebird.ai został wykorzystany do współpracy z organizacją charytatywną zajmującą się chorobami neuronu ruchowego. Współpraca polega na tworzeniu cyfrowych kopii głosów pacjentów w celu stworzenia sztucznego głosu – zamiennika, który ludzie będą mogli wykorzystać, gdy utracą zdolność mówienia.[27]

The Wall Street Journal podobno utworzył 20-osobowy zespół Media Forensics Committee, który ma doradzać swoim dziennikarzom, jak rozpoznać sfałszowane materiały wideo i zaprosił naukowców do wygłoszenia prelekcji na temat najnowszych innowacji w deepfake screening.[28] Jedna z form kryminalistyki mediów obejmuje badanie poszczególnych osób w materiale filmowym pod kątem pojawiających się niespójności fizjologicznych w sposobie, w jaki konstruowane są spreparowane filmy. Obejmuje to sprawdzenie, czy obiekty migają podczas nagrania oraz czy kolor i cienie na ich skórze wydają się migotać. Innym podejściem jest sprawdzenie akustyki filmu skorelowanej z nagrywaną sceną, na przykład z rozmiarem pomieszczenia lub obecnością ludzi w tle. Eksperci nadal nie są zgodni co do tego, czy kryminalistyka mediów jest zdolna do badania przesiewowego deepfake-ów. Narzędzie oparte na sztucznej inteligencji o nazwie FaceForensics zbadało zbiór danych zawierający pół miliona zmanipulowanych obrazów. Wykrywalność za jego pomocą jest zdecydowanie lepsza w porównaniu do ludzkich zespołów kryminalistycznych.[29]

26 https://www.focus.pl/artykul/disney-prezentuje-deepfake-w-megarozdzielczosci-wykorzysta-ja-na-duzym-ekranie

27 https://lyrebird.ai/work

28 Southern, L. (2019) ‘A perfect storm’: The Wall Street Journal has 21 people ‘detecting’ deepfakes [article] Digiday, 1 lipca 2019.

29 Rössler, A. et al. (2019) FaceForensics++: Learning to Detect Manipulated Face Images: https://arxiv.org/abs/1901.08971

Włączają się tez giganci internetowi. Google stworzył materiały dydaktyczne mające na celu wspieranie identyfikowania fałszywych wiadomości w Internecie, które można rozszerzyć o treści wizualne i dezinformacyjne audio[30]. Facebook niedawno ogłosił zamiar stworzenia i udostępnienia zbioru danych zsyntetyzowanych deepfake-ów dla ekspertów, którzy starają się udoskonalić swoje narzędzia do wykrywania deepfake. Także Facebook prowadzi na szeroką skalę weryfikację zamieszczanych materiałów. We wrześniu 2018 r. przedstawił autorską procedurę weryfikacji faktów na okładkach zdjęć i filmów obok tekstu[31]. Pomimo tych działań, Facebook ostatnio odmówił usunięcia wielu deepfake-ów argumentując pozostawienie ich prawem do nieszkodliwej satyry.

Legalne korzystanie z deepfake

Warto wskazać również próby legalizacji posługiwania się technologią deepfake. Wspomniana wcześniej podjęta przez senat amerykański próba senator Ben Sasse zakazania "preparowania nagrań audiowizualnych" znana jako "Malicious Deep Fake Prohibition Act of 2018"[32], nie tylko zawiera definicję „deepfake" podkreślającą oszukańczy charakter nagrania, ale przede wszystkim zakłada odpowiedzialność karną (grzywna czy kara pozbawienia wolności do 10 lat) za fakt stworzenia deepfake, z zamiarem rozpowszechniania, w celu popełnienia czynu karalnego czy innego deliktu. Projekt zakłada zwolnienie od odpowiedzialności „dostawców interaktywnych usług komputerowych" jeżeli podjęliby ograniczenia dostępu do nagrań deepfake.

Dalej idące propozycje, poza odpowiedzialnością karną, zawiera projekt "Deep Fakes Accountability Act"[33] zaproponowany w czerwcu br. przez reprezentantkę demokratów z Nowego Jorku Yvette Clarke. Według tego aktu, nagrania deepfake miałyby być rejestrowane w specjalnym rejestrze deepfake oraz wyraźnie oznaczone znakiem wodnym. Film powinien zawierać także oświadczenie pojawiające się na dole obrazu przez cały czas trwania nagrania wskazujące, że nagranie zawiera zmienione elementy dźwiękowe i wizualne oraz zwięzły opis zakresu takiej zmiany. Brak takich oznaczeń wiązałby się z zapłatą kary 150 000 USD. W przypadku braku oznaczenia, możliwe byłoby pociągniecie do odpowiedzialności karnej, przy czym odpowiedzialność byłaby zaostrzona, gdyby technologia deepfake została użyta:

- „z zamiarem poniżenia lub innego nękania osoby fałszywie przedstawionej, pod warunkiem że zaawansowany technologicznie zapis fałszywego

30 Google's Be Internet Citizens programme: https://internetcitizens.withyoutube.com/
31 Woodford, A. (2018) Expanding Fact-Checking to Photos and Videos [article] Facebook Newsroom, 13 września 2018.
32 https://www.congress.gov/bill/115th-congress/senate-bill/3805/text
33 https://www.congress.gov/bill/116th-congress/house-bill/3230/text

personifikacji zawiera treści seksualne o charakterze wizualnym i wydaje się, że taka osoba wykonuje takie czynności seksualne lub jest w stanie nagości"

- „z zamiarem spowodowania przemocy lub krzywdy fizycznej, podżegania do konfliktu zbrojnego lub dyplomatycznego lub ingerencji w oficjalne postępowanie, w tym wybory, pod warunkiem, że zaawansowane technologicznie fałszywe dane personalne faktycznie stanowiły wiarygodną groźbę podżegania do takich"
- „w trakcie postępowania karnego związanego z oszustwem, w tym oszustw związanych z papierami wartościowymi i oszustwami elektronicznymi, fałszywą personifikacją lub kradzieżą tożsamości"
- „przez obce mocarstwo lub jego przedstawiciela, z zamiarem wywarcia wpływu na krajową debatę o polityce publicznej, ingerowania w wybory federalne, stanowe, lokalne lub terytorialne lub angażowania się w inne działania, których taka władza nie może zgodnie z prawem podjąć".

Przestępstwa te, podlegałyby eksterytorialnej jurysdykcji federalnej, jeśli oskarżony lub przedstawiona w nagraniu osoba byłaby obywatelem lub stale zamieszkiwała w Stanach Zjednoczonych. Co bardzo ciekawe, tych zasad odpowiedzialności nie stosowałoby się do nagrań „wyprodukowanych przez urzędnika Stanów Zjednoczonych lub jego upoważnienia, w trosce o bezpieczeństwo publiczne lub bezpieczeństwo narodowe".

Poza Stanami, także w ustawodawstwach europejskich podejmowane są próby uregulowania tworzenia i rozpowszechniania deepfake. W 2017 roku niemiecki rząd przyjął nowe prawo nakładające kary na firmy technologiczne, które zostały zobowiązane do usuwania treści rasistowskich lub zawierających groźby, w ciągu 24 godzin od zgłoszenia. Planuje się zastosowanie tych regulacji do deepfake-ów, ale krytycy twierdzą, że mogłoby to przynieść skutki odwrotne do zamierzonych i brak skuteczności, ponieważ trudno jest zidentyfikować twórców deepfake, z których wielu mieszka poza terenem Niemiec. Szczególnie krytykowane są rozwiązania przyjęte w Singapurze, gdzie uchwalono przepisy pozwalające rządowi żądać od platform administrujących mediami społecznościowymi usunięcia treści, które uważa za fałszywe. Krytycy uważają, że te rozwiązania rząd będzie wykorzystywać do stłumienia wolność słowa pod hasłami walki z dezinformacją, co stanowić będzie „katastrofę dla wyrażania opinii online".[34]

Z naszego punktu widzenia najcenniejsze są inicjatywy podejmowane w Unii Europejskiej. Komisja Europejska w tzw. Akcie o sztucznej inteligencji[35] przewidziała w art. 52 ust. 3 konieczność oznaczania deepfake wskazując, iż „użytkownicy

34 The Guardian (2019) Singapore fake news law a 'disaster' for freedom of speech, says rights group.

systemu sztucznej inteligencji, który generuje obrazy, treści dźwiękowe lub treści wideo, które łudząco przypominają istniejące osoby, obiekty, miejsca lub inne podmioty lub zdarzenia, lub który tymi obrazami i treściami manipuluje, przez co osoba będąca ich odbiorcą mogłaby niesłusznie uznać je za autentyczne lub prawdziwe („deepfake"), ujawniają, że dane treści zostały wygenerowane lub zmanipulowane przez system sztucznej inteligencji." Nie ma przy tym dalszych wytycznych ale cel w postaci informowania o technologii deepfake został wyartykułowany. Zdaniem Komisji „jeżeli system sztucznej inteligencji jest wykorzystywany do generowania obrazów, dźwięków lub treści wideo, które w znacznym stopniu przypominają autentyczne treści, lub do manipulowania takimi obrazami, dźwiękami lub treściami wideo, powinien istnieć obowiązek ujawniania, że dane treści wygenerowano za pomocą środków zautomatyzowanych, z zastrzeżeniem sytuacji wyjątkowych dotyczących zgodnych z prawem celów (egzekwowanie prawa, wolność wypowiedzi). Pozwala to tym osobom na dokonywanie świadomych wyborów lub na wycofanie się z danej sytuacji."

Podsumowując: samo tworzenie deepfake, a także ich dystrybuowanie może być zgodne z prawem. Problem mógłby pojawić się wówczas, gdyby zmanipulowany materiał filmowy dostał się w niepowołane ręce. Można wyobrazić sobie zdestabilizowanie demokracji np. przez pokazanie przed dniem wyborów polityka zaangażowanego w działalność przestępczą albo przedstawienie żołnierza USA palącego Koran[36]. Potwierdzeniem realności takiej sytuacji są przykłady wskazane na początku tego artykułu.

Oczywiście deepfake wykorzystane do celów rozrywkowych pomagają realizować coraz to ciekawsze wizje twórców, w odpowiednich rękach mogą zdziałać wiele dobrego, jednak trafiając w te nieodpowiednie stają się zagrożeniem dla innych. Na pewno tworzenie deepfake mieści się w obowiązujących przepisach. Trudniej przypisać aktualne regulacje do skutków rozpowszechniania nagrań zawierających deepfake. Istotne jest to, w jaki sposób nagrania powstałe przy użyciu technologii deepfake będą wykorzystywane. W dalszej perspektywie korzystanie z nich będzie miało duży wpływ na ich unormowanie i to, czy ich wykorzystywanie będzie nadzorowane czego dowodem są nie tylko amerykańskie ale także europejskie regulacje. Nasuwa się pytanie, czy poza USA i Europą świat uzna potrzebę regulowania deepfake?

35 Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) i zmieniające niektóre akty ustawodawcze Unii; https://eur-lex.europa.eu/legal-content/PL/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206 [dostęp grudzień 2022]

36 Chesney, R. and Citron, D. (2019) Deepfakes and the new disinformation war. Foreign Affairs, styczeń/luty 2019

Jednakże nie tylko deepfake zagraża szeroko pojętemu cyberbezpieczeństwu. Poza dezinformacją szerzoną przez fake newsy i fałszywe konta, pojawia się nowe pojęcie tzw. „Internetu rzeczy" - (IoT – Internet of Thing). Rzeczy, które „wtłaczają" nasze dane do Internetu. Już nie tylko pojedynczy ruch myszką, kliknięcie, zalogowanie na stronie internetowej, zakup, nasze kontakty, zainteresowania, adres IP, czas dostępu, długość trwania sesji, rodzaj używanego oprogramowania, lokalizację urządzenia – wszystko w sieci jest na bieżąco rejestrowane, analizowane i oceniane – do sieci dostają się dane o naszej lokalizacji, o upodobaniach, parametrach zdrowia... przez każdą aktywnością na wyświetlaczu smartfona, wadze łazienkowej, smartwatche z pulsoksymetrem. O tych „rzeczach" w kontekście praw własności intelektualnej w kolejnym artykule tego cyklu.