JULIA BERNACKA

Problematyka prawna technologii *deepfake* – analiza legalności tworzenia i rozpowszechniania deepfake'ów po uchwaleniu AI Act

Legal Issues of Deepfake Technology – Analysis of the Legality of Creating and Disseminating Deepfakes after the Adoption of the AI Act

Abstract

Deepfake technology is almost a permanent fixture in the online world today, increasing in popularity and quality year after year. Most often expressed in the form of an image, video or sound, they present content that is not true, but looks deceptively authentic, resembling existing people, objects or places. For this reason, they can contribute to many negative consequences for people. The creation and dissemination of deepfakes can harm legally protected goods, exploiting a person's image, good name, personal information, as well as a lot of other information about the person. This is why deepfake technology can stir up a lot of controversy, contributing to questions about the legality of creating and distributing such content. Thus, in the article, the author seeks to answer the problem outlined above, namely whether or not the creation and/or dissemination of deepfake is legal in our legal system. This study was examined primarily from the perspective of the latest EU regulation on AI – the Artificial Intelligence Act (and fragments of other relevant European acts), examining whether it introduces specific regulations for the subject matter, ultimately aiming to determine the current legal situation of deepfakes in the context of the legality of their creation and/or dissemination. In order to verify possible illegality, selected potential legal violations that may occur as a result of the creation and/or dissemination of deepfake are analyzed in turn. The above is based in particular on the canvass of Polish law and, to the necessary extent, EU law. This is to enable conclusions to be drawn as to whether

JULIA BERNACKA – magister prawa, Uniwersytet Komisji Edukacji Narodowej w Krakowie, ORCID – 0009-0006-8244-9996, e-mail: ber.julia2@gmail.com

HTTPS://DOI.ORG/10.36128/PRIW.VI58.1182

there are currently regulations in our legal system that may affect the illegality of the creation and/or dissemination of deepfakes. The legal analysis has been narrowed only to the regulations pertaining to the issue under study, in the context of the distinguished legality.

KEYWORDS: deepfake, legality, illegality, AI Act, violation of law, legally protected goods

SŁOWA KLUCZOWE: *deepfake*, legalność, nielegalność, AI Act, naruszenie prawa, dobra prawnie chronione

1 Wstęp

W ostatnich latach popularność technologii *deepfake* stale rośnie, dlatego dość łatwo można natrafić na materiał tego typu, a nawet samodzielnie go wygenerować. Rozwojowi ulega także jakość tworzonych w ten sposób treści, które stają się bardziej dokładne i realistyczne, przypominając (czasem wręcz łudząco) rzeczywiste osoby czy miejsca. Perspektywa tworzenia takich materiałów w stosunkowo szybki i tani sposób może zachwycać i budzić nadzieje co do efektów rozwoju technologicznego, a także potencjalnych dróg ich wykorzystania. Obok tego pojawiają się jednak obawy i szereg niebezpieczeństw związanych z technologią *deepfake* zarówno w procesie ich tworzenia, jak i późniejszego korzystania z nich. Osoby je tworzące mogą używać rozmaitych danych dotyczących człowieka, jego wizerunku, danych osobowych, ingerować w sferę dóbr osobistych, a także działać w różnych celach, często nieetycznych, prowadzących do pokrzywdzenia kogoś czy nawet przestępstwa. To właśnie dlatego technologia *deepfake* może budzić kontrowersje już z racji specyfiki jej funkcjonowania i prawidłowości, jakie za sobą pociąga, implikując pytania o legalność tworzenia i/lub rozpowszechniania treści tego typu, a więc fałszywych z natury, ale dotyczących prawdziwych elementów rzeczywistości. Co istotne, na obszarze Unii Europejskiej sukcesywnie rozwijane są regulacje prawne, poprzez które podejmuje się problematykę sztucznej inteligencji (dalej: AI), w tym również problematykę *deepfake*, oraz wprowadza stosowne zasady czy obowiązki. Jedną z nich stanowi rozporządzenie

Unii Europejskiej w sprawie sztucznej inteligencji[1], gdzie obok innych postanowień swoiste miejsce zajmuje właśnie technologia *deepfake*.

Toteż jak zostało to już zarysowane, kwestia legalności deepfake'ów wydaje się niejednoznaczna, zaś stosunkowa nowość technologii, jej ciągły rozwój, a także pewne szczególne cechy i okoliczności wpływają na skomplikowanie tej materii, przyczyniając się do obaw związanych zarówno z bezpiecznym i legalnym tworzeniem deepfake'ów, jak i możliwością zostania ofiarą tego typu modyfikacji. Dodatkowo ze względu na fakt, iż w naszej przestrzeni prawnej pojawiła się nowa regulacja prawna, która traktuje problematykę prawną *deepfake*, niezbędne jest przeprowadzenie niniejszej analizy w kontekście zbadania przyjętych tam norm, sprawdzając ich wpływ na obrany obszar badawczy. Biorąc pod uwagę powyższe spostrzeżenia, w szczególności niepewność co do sytuacji prawnej tworzonych i rozpowszechnianych deepfake'ów, potrzeba więc rozważyć, jak wygląda obecnie stan uregulowania powyższej materii w naszym systemie prawnym[2].

Wobec tego celem artykułu jest ustalenie, czy w naszym systemie prawnym, wobec aktualnego stanu prawnego, tworzenie i/lub rozpowszechnianie deepfake'ów jest legalne, czy też nie. W głównej mierze analiza będzie prowadzona wobec regulacji AI Act (oraz w niezbędnym zakresie innych aktów europejskich), badając przyjęte tam rozwiązania prawne, relewantne dla rozważanego zagadnienia. To zaś umożliwi odpowiedź na pytanie, czy niniejsze wpływa na obrany problem badawczy. Następnie zostanie przeprowadzona analiza potencjalnych naruszeń prawnych (na kanwie wybranych przez autorkę pól), do których może dojść już na etapie tworzenia i/lub rozpowszechniania *deepfake*. Powyższe ma na celu weryfikację, czy w naszym systemie prawnym funkcjonują regulacje, których naruszenie może pociągać za sobą nielegalność wyróżnionych procesów. Dzięki powyższym badaniom zostaną wysnute wnioski na poruszony temat, w szczególności co do stanu prawnego po uchwaleniu AI Act. Na zakończenie przeprowadzone badania zostaną podsumowane, co doprowadzi

1 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz.Urz. UE L z 2024 r., 1689), dalej: akt w sprawie sztucznej inteligencji, AI Act.

2 Przez nasz system prawny należy tu rozumieć ustawodawstwo polskie oraz mające zastosowanie regulacje Unii Europejskiej.

do konkluzji w zakresie obranego problemu badawczego, jak również zauważonych potrzeb regulacyjnych.

2 *Deepfake* – czym jest i co zwraca największą uwagę

Deepfake to specyficzny rodzaj materiału (głównie w formie obrazka, wideo bądź nagrania głosu), który powstał przy użyciu AI i prezentuje treść, która w rzeczywistości nigdy się nie wydarzyła. Najczęściej wykorzystuje się przy tym wizerunki prawdziwych osób fizycznych, istniejące przedmioty czy miejsca, umieszczając je w wymyślonym kontekście. Właśnie dlatego *deepfake* nazywany jest często fałszywą rzeczywistością, ponieważ przypomina prawdziwe elementy rzeczywistości, ale w sztucznie wygenerowanym, nieprawdziwym materiale. Powyższe możliwe jest dzięki rozwojowi technik AI[3], zaś analiza literalna pojęcia oraz wykorzystywanie możliwości uczenia głębokiego pozwala mówić o „głębokim fałszerstwie", gdzie z powodu wysokiego zaawansowania technologicznego coraz trudniej wykryć fałszywość takich treści[4]. Możliwości modyfikacji rozpoczynają się od drobnych zmian, aż po całkowicie sztuczne wygenerowanie materiału[5]. Sposoby ich wykorzystania są w zasadzie nieograniczone, zaś

[3] Co do sposobu funkcjonowania oraz rozwoju technologii *deepfake*, w tym sieci GAN, por. np. Robin Chataut i Aadesh Upadhyay, "Introduction to Deepfake Technology and Its Early Foundations," w *Deepfakes and Their Impact on Business*, red. Gaurav Gupta et al. (IGI Global, 2024), 1 – 18; Min Liu i Xijin Zhang, *Deepfake Technology and Current Legal Status of It*, Proceedings of the 2022 3rd International Conference on Artificial Intelligence and Education (IC-ICAIE 2022) (AHCS 9, 2023), 1309; Jessica Ice, "Defamatory Political Deepfakes and the First Amendment," *Case Western Reserve Law Review*, vol. 70 (2) (2019): 420 – 423.

[4] Por. Berenika Kaczmarek-Templin, "Deepfake i jego walor dowodowy w procesach cywilnych," w *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, red. Kinga Flaga-Gieruszyńska, Jacek Gołaczyński i Dariusz Szostek (Warszawa: C.H. Beck, Legalis/el., 2019), 2.

[5] Metody tworzenia *deepfake* oraz rodzaje *deepfake* por. np. Ewelina Bartuzi-Trokielewicz, Michał Ołowski i Adrain Kordas, *Deepfake: Jak sztuczna inteligencja może nas oszukiwać?*, Cyberprofilaktyka NASK, https://cyberprofilaktyka.pl/blog/deepfake-jak-sztuczna-inteligencja-moze-nas-oszukiwac_i40.html [dostęp: 10.12.2024]; Oskar Bogdanka, "Naruszenie wizerunku przy wykorzystaniu technologii

efekty mogą wiązać się zarówno z pozytywnymi, jak i negatywnymi skutkami. W ten sposób powstają więc ciekawe reklamy, kampanie społeczne[6], zachwycające sceny w filmach czy śmieszne obrazki, jednak obok tego będą obraźliwe treści, sztucznie zmanipulowane sceny pornograficzne, próby zniesławienia czy oszustwa[7]. To właśnie te drugie powodują największe obawy i potrzebę pochylenia się nad tą problematyką[8].

Deepfake staje się technologią stale obecną w naszym życiu codziennym. Już teraz łączna liczba filmów *deepfake* opublikowanych online w 2023 r. wyniosła 95 820[9]. To wartość o 550% większa niż w 2019 r.[10], co wskazuje na tendencję wzrostową. Wiele narzędzi umożliwiających tworzenie *deepfake* jest łatwo dostępnych w sieci, a dla stworzenia materiału z prawdziwą osobą fizyczną wystarczy nawet niewielka próbka jej danych. Zgodnie z badaniem przeprowadzanym przez McAfee w jednym z analizowanych przypadków wystarczyły zaledwie 3 sekundy dźwięku dla wygenerowania nagrania głosu *deepfake* z 85% zgodnością z oryginałem, a dzięki

deepfake – analiza prawna i praktyczna," *Opolskie Studia Administracyjno-Prawne*, vol. 20, nr 1 (2022): 13 – 14.

6 Jako przykład tego typu wykorzystania technologii *deepfake* por. np. Zero Malaria Britain, *How We Made David Beckham Speak 9 Languages*, https://www.youtube.com/watch?v=CF_eokMCW20 [dostęp: 18.11.2024], gdzie twórcy wykorzystali możliwości sztucznej inteligencji w kampanii przeciwko malarii, tłumacząc słowa D. Beckhama na różne języki, co sprawiało wrażenie, jakby bohater sam wypowiadał te słowa.

7 Jedną z częstszych form negatywnego wykorzystania technologii *deepfake* jest oszustwo finansowe, w tym m.in. wykorzystywanie wizerunku i autorytetu osób znanych publicznie, rzekomo namawiających do wątpliwych inwestycji finansowych, czy też przypadki rozmów telefonicznych, wykorzystujących deepfake'owe nagranie głosu osoby bliskiej dla pokrzywdzonego. Por. np. *NASK ostrzega: rośnie liczba oszustw deepfake wykorzystujących wizerunki znanych osób* (NASK, 2024), https://nask.pl/magazyn/NASK-ostrzega-rosnie-liczba-oszustw-deepfake-wykorzystujacych-wizerunki-znanych-osob/ [dostęp: 10.12.2024]; Urząd Komisji Nadzoru Finansowego, *Przegląd wybranych oszust internetowych – listopad 2023* (pkt Wykorzystanie *deepfake* do „metody na wnuczka"), https://cebrf.knf.gov.pl/komunikaty/artykuly-csirt-knf/362-ostrzezenia/885-przeglad-wybranych--oszustw-internetowych-listopad-2023 [dostęp: 10.12.2024].

8 Więcej przykładów deepfake'owych treści por. np. Ilona Dąbrowska, "Deepfake – nowy wymiar internetowej manipulacji," *Zarządzanie Mediami*, t. 8 (2) (2020), 91 – 96.

9 Badanie Security Hero, *2023 State of Deepfakes, Realities, Threats and Impact*, okres badania: 15.07 – 29.08.2023, https://www.securityhero.io/state-of-deepfakes/ [dostęp: 2.12.2024].

10 Ibidem.

odpowiedniemu trenowaniu możliwe było osiąganie 95% zgodności głosu, na podstawie niewielkiej liczby plików audio[11].

Obawy o negatywne konsekwencje istnienia technologii *deepfake*[12] potęgowane są przez niektóre okoliczności ich funkcjonowania, w szczególności: wirusowe rozprzestrzenianie się *deepfake* w sieci; ryzyko dodatkowych przeróbek (pogłębiających pokrzywdzenie i inne pejoratywne następstwa); coraz lepsza jakość i dokładność; prostota tworzenia; potencjalna trudność w ustaleniu podmiotu, który stworzył *deepfake*; stosunkowa łatwość w zdobyciu danych dotyczących prawdziwych osób fizycznych, a następnie przedstawienia ich w sytuacji, w której nigdy nie brały udziału. Podstawową okolicznością, która wzbudza wątpliwości co do legalności tworzenia i rozpowszechniania deepfake'ów, jest jednak fakt ich nieprawdziwości, ale jednocześnie wrażenie, jakoby treść była autentyczna.

3 *Deepfake* a prawo. Regulacja prawna problematyki legalności *deepfake* ze szczególnym uwzględnieniem AI Act

Obecnie w naszym systemie prawnym regulacja problematyki *deepfake* jest stosunkowo wąska. Oprócz pojedynczych przepisów niedawno uchwalony AI Act wprowadza tu kilka istotnych postanowień, wobec których nie można przejść obojętnie.

Bazując na przyjętym w AI Act podziale systemów AI na te, które umożliwiają generowanie bądź manipulowanie treści typu *deepfake*, należałoby raczej kwalifikować je jako „niektóre systemy AI", implikujące dodatkowe obowiązki dla dostawców i podmiotów stosujących AI (por. np. art. 50 ust. 2 i 4). Tego typu systemy określane są jako stwarzające „ograniczone

11 Amy Bunn, *Artificial Imposters – Cybercriminals Turn to AI Voice Cloning for a New Breed of Scam* (McAfee, 2023), https://www.mcafee.com/blogs/privacy-identity-protection/artificial-imposters-cybercriminals-turn-to-ai-voice-cloning-for-a-new-breed-of-scam/ [dostęp: 10.12.2024].

12 Przykłady negatywnych sposobów wykorzystywania technologii *deepfake* por. np. Olga Wasiuta i Sergiusz Wasiuta, "Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość," *Annales Universitatis Paedagogicae Cracoviensis, Studia de Securitate*, nr 9 (3) (2019): 22 – 24.

ryzyko"[13]. W literaturze wskazuje się jednak na krytyczną lukę prawną w obszarze niektórych rodzajów *deepfake*, poprzez brak przepisów AI Act odnoszących się konkretnie do tego typu treści, jak również istotne braki w wykazie systemów AI „wysokiego ryzyka"[14]. Wskazuje się tu przede wszystkim na takie systemy, które mogą wiązać się z niepożądanym wpływem na prawa podstawowe, wymieniając przede wszystkim generowanie *deepfake* prowadzących do wymuszeń, dezinformacji wyborczej czy też prezentujących treści o wykorzystywaniu seksualnym dzieci[15]. Podkreśla się przy tym niejednolitość tej technologii, gdzie obok treści bardzo szkodzących będą pojawiać się materiały neutralne, przez co ogólne, ustandaryzowane regulacje dla wszystkich możliwych typów *deepfake* nie będą odpowiednie[16]. Przyjęte obowiązki w zakresie przejrzystości mogą okazać się niewystarczające przy niektórych rodzajach systemów, np. tych generujących treści pornograficzne[17]. Zauważa się jednak, iż szkodliwość, o której mowa, w dużym stopniu będzie zależna od kontekstu tworzonych materiałów, a nie samej natury technologii, co nie zmienia jednak faktu, iż niektóre z nich stanowią znaczne ryzyko dla człowieka, a nawet są całkowicie nielegalne, dlatego powinny zostać bezpośrednio zakazane i powiązane z konsekwencjami prawnymi[18]. Pojawiają się więc głosy, że tworzenie godnej zaufania AI wymaga działań proaktywnych, wśród których mogłoby pojawić się karanie szkodliwych form korzystania z deepfake'ów przez osoby fizyczne[19]. Brak umieszczenia systemów AI generujących *deepfake* w kategorii zakazanych praktyk czy wysokiego ryzyka można więc intepretować w ten sposób, iż technologie te nie są

13 Por. np. Lilian Edwards, *The EU AI Act: A Summary of Its Significance and Scope* (Ada Lovelace Institute, 2022), 9 – 15.

14 Felipe Romero Moreno, "Generative AI and Deepfakes: A Human Rights Approach Tackling Harmful Content," *International Review of Law, Computers & Technology*, vol. 38, nr 3 (2024): 303.

15 Ibidem, 303 – 305.

16 Por. np. Rüya Tuna Toparlak, *Criminalising Pornographic Deep Fakes: A Gender-Specific Inspection of Image-Based Sexual Abuse* (Sciences Po Law School, The 10th Graduate Conference, Paryż, 2022), 24.

17 Ibidem.

18 Por. Mateusz Łabuz, "Deep Fakes and the Artificial Intelligence Act – an Important Signal or a Missed Opportunity?," *Policy & Internet*, nr 16 (4) (2024): 7 oraz Mateusz Łabuz, "Regulating Deep Fakes in the Artificial Intelligence Act," *Applied Cybersecurity & Internet Governance*, vol. 2 (1) (2023): 262.

19 Łabuz, "Deep Fakes and the Artificial Intelligence Act – an Important Signal or a Missed Opportunity?," 13.

oceniane jako *in priori* zawsze szkodliwe, co powodowałoby konieczność wykluczenia ich z naszej przestrzeni. Mimo to nawet włączenie systemów, o których mowa, do kategorii wysokiego ryzyka nie wpłynęłoby na nielegalność generowanych w ten sposób treści, a jedynie na zwiększenie ilości obowiązków określonych podmiotów. Uznanie natomiast pewnych praktyk w zakresie *deepfake* za zakazane na gruncie art. 5 AI Act mogłoby już takie konsekwencje wywoływać, poprzez zakazanie wprowadzania do obrotu, oddawania do użytku lub wykorzystywania określonego rodzaju systemu AI.

Niezwykle istotny z punktu widzenia niniejszych rozważań jest fakt, iż legislatorzy unijni zdecydowali się na prawne wyodrębnienie pojęcia deepfake'u, proponując jego definicję legalną. Intepretując ową okoliczność, można dojść do wniosku, iż skoro wyróżniono i zdefiniowano takie pojęcie, to *deepfake* stanowi znaczący element rzeczywistości, dla którego istnieje potrzeba posługiwania się pojęciem legalnym. W ten sposób jednak tylko częściowo zawężono dyskusje nad problemem, co należy rozumieć pod tym pojęciem oraz jakie elementy świadczą o zakwalifikowaniu danego materiału jako *deepfake*. Zgodnie z AI Act będą to: „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) AI Act). Poddając niniejszą definicję analizie literalnej, należy zauważyć, iż dla powstania deepfake'u konieczne jest łączne spełnienie owych czterech cech. Są to: 1) wygenerowanie lub zmanipulowanie przez AI; 2) obraz, treść dźwiękowa lub treść wideo; 3) przypominanie istniejących osób, przedmiotów, miejsc, podmiotów lub zdarzeń; 4) odbiorca może niesłusznie uznać taki materiał za autentyczny lub prawdziwy (a więc treść nie jest prawdziwa). W związku z tym wydaje się, że jeżeli dany materiał nie będzie spełniał którejkolwiek z powyższych cech, to nie zostanie uznany za *deepfake*. Możliwe byłoby więc wygenerowanie treści, która co prawda przypomina istniejącą osobę, ale odbiorca jest w stanie bezsprzecznie stwierdzić jej fałszywość (brak cechy nr 4), bądź też materiału, który wydaje się prawdziwy, przedstawia postać człowieka, ale nie jest to żadna z istniejących osób, a twarz wygenerowana syntetycznie (brak cechy nr 3). Zgodnie z definicją legalną treści tego typu nie będą stanowiły deepfake'u, a więc nie znajdą do nich zastosowania odnośne regulacje prawne. Co istotne, *deepfake* musi powstać z udziałem AI (być wygenerowany lub zmanipulowany przez AI). Warto również zwrócić uwagę, iż wśród rodzajów *deepfake* definicja nie obejmuje treści tekstowych, a więc

nie ma możliwości uznania, jakoby wygenerowany bądź zmanipulowany przez AI tekst kiedykolwiek stanowił *deepfake*. Korzystne z punktu widzenia tak ukutej definicji wydaje się uwzględnienie w niej również innych niż człowiek elementów rzeczywistości, takich jak miejsca czy przedmioty. Nawet jeżeli *deepfake* nie prezentuje wizerunku konkretnej osoby fizycznej, wciąż może prowadzić do negatywnych konsekwencji, np. przyczyniając się do szerzenia fake newsów o sytuacji na świecie, sterując w ten sposób opinią publiczną czy zachowaniem konkretnych grup[20]. Niniejszy fakt może jednak wpłynąć na potrzebę zróżnicowania przyszłych rozwiązań regulacyjnych *deepfake* i podzielenie ich ze względu na prezentowany element rzeczywistości, co da zapewne większą ochronę człowiekowi niż innym jego przedmiotom. Podobnie może się stać w przypadku przepisów dotyczących legalności/nielegalności takich treści.

Mimo iż wyodrębnienie pojęcia deepfake'u można oceniać raczej pozytywnie, to trzeba tu zauważyć, iż przyjęta definicja rodzi pewne wątpliwości interpretacyjne ze względu na pojawiające się nieścisłości. W szczególności uwagę należy zwrócić na pojęcia „przypominają" oraz „istniejące". Zaczynając od drugiego z nich, generowana treść stanowi *deepfake*, jeżeli przypomina istniejące osoby, przedmioty, miejsca, podmioty lub zdarzenia. Nasuwa się więc dylemat, czy „istniejący" oznaczać będzie wyłącznie osobę żyjącą, czy również osoby zmarłe, a więc „istniejące kiedyś". *Deepfake* z wizerunkiem osoby zmarłej może być częstą praktyką, stąd niniejsze ustalenie wydaje się istotne. Wyróżniona przy okazji cechy nr 3 kwestia podobieństwa *deepfake* do prawdziwych elementów rzeczywistości również jest niejednoznaczna w zakresie definicyjnym. Owa niespójność zauważalna jest nawet w ramach samego AI Act, gdzie odpowiedni przepis definicyjny wskazuje wyłącznie na pojęcie „przypominania" istniejących, wyliczonych elementów (por. art. 3 pkt 60) AI Act), natomiast w innym miejscu rozporządzenia pojawia się już potrzeba „łudzącego przypominania" (por. motyw 134 AI Act). W innych aktach ustawodawstwa europejskiego, w których porusza się problematykę *deepfake*, odpowiednie opisy również zdają się wspominać o większym stopniu podobieństwa, wskazując kolejno na potrzebę „łudzącego przypominania" (por. art. 35 ust. 1 pkt k)

20 Por. np. Łabuz, "Regulating Deep Fakes in the Artificial Intelligence Act," 257 – 258; Łabuz, "Deep Fakes and the Artificial Intelligence Act – an Important Signal or a Missed Opportunity?," 5.

DSA[21]) oraz „wyraźnego podobieństwa" do istniejących naprawdę osób, przedmiotów, miejsc lub innych podmiotów (por. motyw 19 dyrektywy[22]). Biorąc pod uwagę fakt, iż kwestia „przypominania" prawdziwych, wyliczonych elementów rzeczywistości stanowi jeden z obligatoryjnych aspektów deepfake'ów, ta zbędna nieprecyzyjność może generować wątpliwości interpretacyjne[23], w szczególności przy ustaleniach, która treść kwalifikuje się jako *deepfake*, ponieważ przypomina dany element w stopniu wystarczającym, a która nie zostanie za takowy uznana, ponieważ zawiera zbyt mało podobnych części. Pojawiające się nieścisłości mogą przyczyniać się do niejednolitości, niekiedy uznając za wystarczające prezentowanie wyłącznie kilku fragmentów twarzy człowieka, a niekiedy wyraźnego wizerunku. Pomimo zauważalnych tu niejednoznaczności, to z punktu widzenia obecnie funkcjonujących regulacji prawnych w zakresie *deepfake* niniejsze nie powinno mieć wpływu na problematykę legalności ich tworzenia i/lub rozpowszechniania. W przywołanych tu aktach prawnych ukute w nich definicje czy opisy treści typu *deepfake* w pozostałym zakresie należy oceniać jako stosunkowo spójne[24]. Mimo wskazanych już różnic co do użytych sformułowań czy stopnia podobieństwa definicje zdają się zmierzać w tym samym kierunku. Niniejsze prowadzi więc do wniosku, iż w naszej przestrzeni prawnej wypracowano w miarę jednolite rozumienie pojęcia *deepfake*.

Idąc dalej, należy w końcu pochylić się nad wprowadzonymi obowiązkami w zakresie przejrzystości generowanych treści[25]. Ze względu na pewne ograniczone ryzyko dostawcy systemów AI, które generują treści w postaci syntetycznych dźwięków, obrazów, wideo lub tekstu, mają

21 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z dnia 19 października 2022 r. w sprawie jednolitego rynku usług cyfrowych oraz zmiany dyrektywy 2000/31/WE (akt o usługach cyfrowych), dalej: akt o usługach cyfrowych, DSA.

22 Dyrektywa Parlamentu Europejskiego i Rady (UE) 2024/1385 z dnia 14 maja 2024 r. w sprawie zwalczania przemocy wobec kobiet i przemocy domowej, dalej: dyrektywa.

23 Co do poruszonego problemu por. np. Łabuz, "Deep Fakes and the Artificial Intelligence Act – an Important Signal or a Missed Opportunity?," 5 – 6.

24 Por. motyw 139 AI Act, art. 3 AI Act, motyw 19 dyrektywy oraz art. 35 ust. 2 pkt k) DSA.

25 Co do wspomnianej tu zasady przejrzystości por. np. Damian Flisak, *O niełatwym wytyczaniu granic rzeczywistości przez AI Act: zasada transparentności, deepfake* (Warszawa: Prawo i Nowe Technologie 1/2024, Wydawnictwo C.H. Beck, Legalis/el., 2024), 3 – 5.

zapewnić oznakowanie takich wyników tak, aby były wykrywalne jako sztucznie wygenerowane lub zmanipulowane (art. 50 ust. 2 AI Act). AI Act kolejno określa wprost, aby „podmioty stosujące system AI, który generuje obrazy, treści audio lub wideo, stanowiące treści *deepfake* lub który manipuluje takimi obrazami lub treściami, ujawniały, że treści te zostały sztucznie wygenerowane lub zmanipulowane" (art. 50 ust. 4 AI Act). Z punktu widzenia szkodliwości potencjalnie nielegalnych treści *deepfake* należy uznać przydatność wprowadzonych tu obowiązków przejrzystości. W sytuacji gdy treść godzi w dobra człowieka, np. prezentując jego wizerunek w niekorzystnym świetle, to wyraźne jej oznaczenie jako sztucznie wygenerowana może rozwiać wątpliwości odbiorców co do prawdziwości zaprezentowanej sytuacji (co oczywiście nie jest w pełni wystarczającym środkiem w walce ze szkodliwymi formami *deepfake*). Podobny obowiązek zawarto również w akcie o usługach cyfrowych, gdzie dostawcy bardzo dużych platform internetowych i bardzo dużych wyszukiwarek mają wprowadzać środki zmniejszające ryzyko, w tym m.in. „zapewnienie, aby informację – bez względu na fakt czy stanowi ona wygenerowany lub zmanipulowany obraz, treść dźwiękową lub treść wideo, które łudząco przypominają istniejące osoby, obiekty, miejsca lub inne podmioty lub zdarzenia, przez co osoba będąca ich odbiorcą niesłusznie uznaje je za autentyczne lub prawdziwe – można było rozróżnić dzięki widocznym oznaczeniom" oraz wskazanie niniejszego, dzięki dodatkowej funkcji (art. 35 ust. 1 pkt k) DSA).

Przydatny w omawianym kontekście może być również art. 16 DSA, stanowiący o konieczności wdrażania przez dostawców usług hostingu mechanizmów, które umożliwią zgłaszanie im obecności w ich usłudze takich informacji, które dana osoba uważa za nielegalne (przykładowo taką treścią mogłoby być wideo *deepfake*, które zdaje się naruszać pewne normy prawne). Zgodnie z motywem 136 AI Act, nawet jeżeli taka treść nie będzie oznakowana jako „sztucznie wygenerowana", nie powinno to wpłynąć na kwestię oceny legalności niniejszej treści na podstawie omawianego artykułu. Należy więc skonstatować, iż brak oznaczenia treści jako sztucznie wygenerowana nie powinien sam w sobie przesądzać o jej niezgodności z prawem na podstawie art. 16 ust. 6 DSA, co „dokonywane ma być wyłącznie w odniesieniu do przepisów regulujących zgodność treści z prawem" (motyw 136 DSA)[26].

Należy również zwrócić uwagę, że omawiany tu AI Act nie będzie miał zastosowania do osób fizycznych, wykorzystujących systemy „AI w ramach

26 Por. ibidem, 1.

czysto osobistej działalności pozazawodowej" (art. 2 ust. 10 AI Act). W doktrynie wskazuje się, iż takie wykluczenie nieprofesjonalnych działań użytkowników końcowych (w tym tworzenia i rozpowszechniania deepfake'ów) wydaje się poważnym problemem[27].

Podsumowując powyższe regulacje, wydaje się jednak, iż takie obowiązki czy nawet ich zaniedbania nie wpływają bezpośrednio na kwestię legalności/nielegalności samego deepfake'u, przyczyniają się ewentualnie do pewnych konsekwencji po stronie podmiotów obowiązanych. Mimo iż brak w regulacji AI Act przepisów, które wprost zakazywałyby tworzenia i rozpowszechniania określonych treści *deepfake* (czy też ograniczały niniejsze w jakiś sposób), w ustawodawstwie unijnym podejmuje się pewne kroki, które mogą mieć znaczenie w tym zakresie. Wśród nich m.in. wspominany już mechanizm zgłaszania nielegalnych treści oraz zwalczania ich przez dostawców usług hostingu (art. 16 i 17 DSA) czy uznanie, iż publiczne udostępnianie za pomocą technologii ICT treści *deepfake porn*, bez zgody osoby, stanowi przestępstwo[28] (motyw 19, art. 5 dyrektywy). Takie działania prowadzą więc do wniosku, iż na poziomie unijnym zauważa się negatywne aspekty funkcjonowania technologii *deepfake* oraz stopniowo zmierza do ich zwalczania.

Analizując obowiązujący w naszym systemie stan prawny w zakresie dotyczącym badanego problemu, można zauważyć, iż brak obecnie regulacji, które wprost podejmowałyby tematykę legalności deepfake'u, określając chociażby zasady ich tworzenia lub rozpowszechniania, kryteria czy granice ich legalności, czy też sytuacje, kiedy *deepfake* będzie nielegalny, wprowadzając przy tym szczególne zasady postępowania czy swoiste konsekwencje prawne. Na bazie obowiązujących regulacji prawnych nie można *a priori* wskazać, jakoby *deepfake* stanowił technologię prawnie zakazaną, a więc należy domniemywać, iż ich tworzenie i rozpowszechnienie jest prawnie dopuszczalne. Powyższe nie oznacza jednak, iż każdy *deepfake* będzie legalny i zgodny z prawem, co uzależnione jest obecnie od oceny prawnej każdego konkretnego przypadku. Naruszenie określonych norm prawnych w momencie tworzenia i/lub rozpowszechniania deepfake'u może prowadzić do nielegalności takiego procesu oraz wiązać się z konsekwencjami prawnymi dla osoby tworzącej bądź rozpowszechniającej. Można więc mówić o względnej neutralności owej technologii, gdzie

27 Por. Łabuz, "Deep Fakes and the Artificial Intelligence Act – an Important Signal or a Missed Opportunity?," 12.

28 Więcej w tym temacie por. s. 687 – 689 niniejszego artykułu.

w konkretnym przypadku poprzez niewłaściwe użycie może dojść do naruszenia prawa[29]. Podobne stanowisko przedstawił K. Szpyt, zauważając, że:

> samo w sobie posługiwanie się technologią *deepfake* nie wypełnia znamion czynu zabronionego i *ex lege* nie rodzi ani negatywnych, ani pozytywnych skutków prawnych. Określone konsekwencje natury prawnej mogą natomiast powstać, w zależności od okoliczności danej indywidualnej sprawy, z uwagi na powstanie i/lub naruszenie praw do konkretnych «składników» wykorzystanych do stworzenia deepfake'owego nagrania[30].

Zdaje się, iż w świetle prędkości rozwoju omawianej technologii, wciąż zwiększającej się liczby powstających deepfake'ów i skomplikowania problematyki, gdzie istnieje ryzyko wystąpienia negatywnych konsekwencji, regulacje w tym zakresie powinny się rozwijać, dostosowując systemy prawne do pojawiających się potrzeb. Trzeba więc zgodzić się z poglądem, że AI Act należy intepretować raczej jako punkt odniesienia (wstęp) dla budowania silniejszych zabezpieczeń aniżeli docelowe działanie regulacyjne Unii Europejskiej w zakresie AI (tu w szczególności *deepfake*)[31].

4 Potencjalne naruszenia – kiedy tworzenie i/lub rozpowszechnianie *deepfake* staje się nielegalne?

Jak zostało to już wyrażone, w obowiązujących w naszym systemie regulacjach prawnych nie podjęto raczej wprost problemu nielegalności deepfake'ów, co nie oznacza jednak, iż obowiązujące prawo nie stawia tu stosownych ograniczeń. Granice legalności, o których mowa, zostają

29 Podobnie por. Łabuz, "Deep Fakes and the Artificial Intelligence Act – an Important Signal or a Missed Opportunity?," 2.

30 Kamil Szpyt, "Sztuczna inteligencja i nowe technologie (nie zawsze) w służbie ludzkości, czyli cywilnoprawna problematyka rozwoju i popularyzacji technologii deepfake," w *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, red. Kinga Flaga-Gieruszyńska, Jacek Gołaczyński i Dariusz Szostek (Warszawa: C.H. Beck, Legalis/el., 2019), 7.

31 Łabuz, "Deep Fakes and the Artificial Intelligence Act – an Important Signal or a Missed Opportunity?," 1 – 3, 15.

zakreślone w szczególności przez dobra prawnie chronione, ale także inne regulacje sektorowe. W niniejszej części zostaną częściowo wyliczone oraz poddane analizie obszary, do których naruszeń może dojść przy okazji tworzenia i rozpowszechniania deepfake'ów. Wykazanie, iż faktycznie funkcjonują takie regulacje, będzie prowadzić do wniosku, iż w określonych sytuacjach może to stanowić działanie nielegalne. Oprócz tego taki przejaw może również zostać wykorzystany dla celów sprzecznych z prawem (np. dla popełnienia przestępstwa). Poniższemu badaniu podlega jednak wyłącznie kwestia legalności/nielegalności deepfake'u na etapie jego tworzenia/rozpowszechniania, a więc to właśnie ten aspekt będzie przedmiotem poniższego przeglądu. Głównym zamierzeniem nie jest więc szczegółowa analiza wskazanych obszarów, ale zwrócenie uwagi na możliwe pola naruszeń poprzez technologię *deepfake*, skutkujące nielegalnością tworzenia i rozpowszechniania takich treści.

4.1. Ochrona danych osobowych

Nie ulega wątpliwości, iż deepfake'i mogą zawierać dane osobowe człowieka, w szczególności dane biometryczne, jak wizerunek czy próbkę głosu, ale także imię, nazwisko czy adres. To zaś wiąże się z koniecznością ich przetwarzania na etapie tworzenia i rozpowszechniania (chociażby poprzez ich zbieranie, łączenie czy rozpowszechnianie) (por. art. 4 pkt 2) RODO). Zgodnie z normami regulacyjnymi RODO[32] takie przetwarzanie musi odbywać się z poszanowaniem zasad określonych w tym akcie, w szczególności przez zapewnienie zgodności z prawem, rzetelności i przejrzystości. W związku z tym, jeżeli osoba przetwarzająca nie uzyskała odpowiednich zgód (oraz nie spełniła innego warunku legalizującego) lub nie zapewnia realizacji wskazanych zasad, dopuszcza się naruszenia względem ochrony danych osobowych, generując lub rozpowszechniając treść w sposób sprzeczny z prawem. Powyższe nie dotyczy jednak sytuacji, gdy osoba fizyczna przetwarza takie dane w ramach czynności o charakterze czysto osobistym lub domowym, co również może stanowić częste przypadki (art. 2 pkt c) RODO). Jednak i wówczas konieczne będzie przestrzeganie

32 Rozporządzenie Parlamentu Europejskiego i Rady (UE) z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych), dalej: RODO.

innych norm prawem obowiązanych, np. ochrony wyznaczonej przez dobra osobiste człowieka bądź prawnoautorską ochronę wizerunku.

4.2. Ochrona dóbr osobistych

Każdy człowiek posiada szereg dóbr dotyczących jego samego, które należałoby określić jako dobra osobiste, pozostające pod ochroną prawa. Pod tym pojęciem należy rozumieć „wartości niemajątkowe, związane z osobowością człowieka, [...] powszechnie uznawane w danym społeczeństwie", jednak wyłączyć należy tu przypadki na tyle błahe, że nie zasługują na ochronę w przekonaniu społecznym[33]. Dobra osobiste pozostają w szczególności pod ochroną prawa cywilnego (art. 23 k.c.[34]). Co istotne, aby możliwe było stwierdzenie tu niezgodności z prawem, działanie musiałoby być bezprawne oraz dążyć do zagrożenia bądź naruszenia dobra osobistego. Bezprawność działania może zostać wyłączona m.in. przez zgodę uprawnionego i to wyłącznie na określonych zasadach[35]. Przy tworzeniu i rozpowszechnianiu deepfake'ów może dojść do ingerencji przede wszystkim w takie dobra jak: wizerunek, cześć, dobre imię czy też nazwisko lub pseudonim człowieka. Przykładowo możliwe byłoby rozpowszechnianie fałszywie wygenerowanych materiałów wideo, które przedstawiają wizerunek człowieka w szkodzącym mu kontekście, czy też sytuacji, która wpływa na negatywny odbiór jego osoby w opinii społecznej, co naruszałoby tym samym jego dobre imię. Zdaje się, iż uprzednie uzyskanie zgody raczej nie jest częstą praktyką.

Zgodnie z utrwalonym orzecznictwem ochrona wizerunku jest niezależna od tego, „czy wskutek posłużenia się nim w sposób bezprawny, a więc bez zgody zainteresowanego, przez osobę trzecią, doszło do naruszenia innych dóbr osobistych człowieka, jak cześć czy godność"[36]. Samo więc

---

33 Anna Sylwestrzak, "art. 23," w *Kodeks cywilny. Komentarz*, wyd. 2, red. Małgorzata Balwicka-Szczyrba i Anna Sylwestrzak (Warszawa: Wolters Kluwer, Lex/el., 2024), 2.

34 Ustawa z dnia 23 kwietnia 1964 r. – Kodeks cywilny (t.j. Dz.U. z 2024 r., poz. 1061 ze zm.), dalej: k.c.

35 Więcej na temat wyłączenia bezprawności poprzez zgodę uprawnionego por. np. Justyna Regan-Balcarczyk, "art. 4," w *Kodeks cywilny. Komentarz*, wyd. 4, red. Mariusz Załucki (Warszawa: Wydawnictwo C.H. Beck, Legalis/el., 2024), 3 – 4.

36 Wyrok SA w Warszawie z dnia 13 stycznia 1999 r., I ACa 1089/98, Wokanda 2000, nr 3, poz. 42.

bezprawne wykorzystanie wizerunku dla stworzenia deepfake'u może grozić konsekwencjami prawnymi. W literaturze podnosi się jednak liczne problemy, jakie pojawiają się podczas próby kwalifikacji prawnej naruszenia wizerunku poprzez *deepfake* oraz ustalenia odpowiedzialności. W szczególności mówi się tu o trudnościach przy: wyborze właściwego dla rozstrzygnięcia porządku prawnego (na jakim terytorium doszło do naruszenia?) oraz samych konkretnych przepisów prawa (które regulacje będą odpowiednie?); wskazaniu podmiotu odpowiedzialnego oraz podmiotów legitymowanych do wystąpienia z powództwem w celu ochrony wizerunku osoby zmarłej; wykazaniu przesłanek zagrożenia lub naruszenia dóbr osobistych, jak również ewentualnego określenia właściwego dobra. Zgodnie ze stanowiskiem wyrażonym przez Sąd Najwyższy głos człowieka również jest jednym z jego dóbr osobistych, zaś za ingerencję można by uznać „przekształcenie go za pomocą narzędzi cyfrowych w taki sposób, iż barwa dźwięku pozwala odbiorcy na identyfikację głosu innej osoby, zaś z wykorzystaniem tak utrwalonego dźwięku dochodzi do fałszywego przypisania tej osobie wypowiadania przez nią określonych treści – zjawisko tzw. *deep fake*"[37]. Tworzenie i rozpowszechnianie materiałów *deepfake*, które w sposób bezprawny uderzają w dobra osobiste człowieka, można by więc uznać za naruszenie, zaś sam materiał za niezgodny z prawem.

4.3. Prawo autorskie

Przy tworzeniu nowych materiałów, takich jak obrazki, widea czy nagrania głosu, w dość oczywisty sposób może dojść do naruszeń prawnoautorskich, poprzez nielegalne wykorzystanie treści prawnie chronionych. Osoba tworząca *deepfake*, w celu jego wykreowania, może potrzebować dostępu do innych materiałów tego typu, przykładowo dążąc do ich manipulacji, łączenia, kopiowania czy nakładania na siebie. W takiej sytuacji, jeżeli użyta treść stanowi utwór (bądź jego fragment) w rozumieniu prawa autorskiego, istnieje ryzyko naruszeń autorskich praw majątkowych oraz autorskich praw osobistych do niego. Będzie to miało miejsce przede wszystkim w sytuacji naruszania art. 17 upapp[38] (poprzez nieuprawnione wykorzystanie treści, w szczególności jeżeli nie uzyskano zgody autora

37 Wyrok SN z dnia 7 marca 2023 r., II CSKP 659/22, OSNC-ZD 2024, nr 1, poz. 7.

38 Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (t.j. Dz.U. z 2022 r., poz. 2509), dalej: upapp.

na korzystanie z utworu i rozporządzanie nim, a nie zachodzą przesłanki dozwolonego użytku) oraz art. 16 upapp (w szczególności w wyniku braku oznaczenia autorstwa użytego utworu, naruszenia jego treści bądź formy, nierzetelnego wykorzystania czy też utraty nadzoru autora nad sposobem korzystania z utworu). Ponadto obok powyższego regulacje prawnoautorskie chronią również wizerunek osoby fizycznej, co wydaje się tu dość często wykorzystywanym elementem. Zgodnie z regulacją rozpowszechnianie wizerunku wymaga co do zasady zezwolenia osoby na nim przedstawionej (art. 81 ust.1 upapp)[39]. Można jednak wątpić, że każdy przypadek *deepfake* tworzony jest po jego uzyskaniu. Co więcej, w odróżnieniu od ochrony na gruncie prawa cywilnego do naruszenia prawnoautorskiego może dojść już w momencie nieuprawnionego rozpowszechnienia wizerunku, a więc ocena, czy doszło do ingerencji w dobro osobiste, czy też nie, nie ma w tym przypadku znaczenia. Należy podkreślić, iż do naruszenia dochodzi tu dopiero w momencie rozpowszechnienia (a więc samo stworzenie deepfake'u z wizerunkiem konkretnej osoby, bez uzyskania jej zgody, nie będzie jeszcze godziło w niniejszą ochronę). Odpowiedzialność za powyższe będzie spoczywać na osobie rozpowszechniającej, a nie tworzącej deepfake'owy materiał.

4.4. Prawo karne

Z perspektywy prawnokarnej jednym z większych problemów w ramach omawianego tematu jest tworzenie i rozpowszechnianie *deepfake* o charakterze seksualnym/pornograficznym, tzw. *deepfake porn*. Tego typu materiały prezentują treści pornograficzne, które w rzeczywistości nigdy nie miały miejsca. Najczęściej wizerunek pewnej osoby fizycznej zostaje w nich podstawiony pod inny materiał wideo. W efekcie powstaje sztuczny materiał, prezentujący, jakoby dana osoba faktycznie brała udział w czynności o charakterze seksualnym. O skali tego problemu świadczą statystyki, w tym ilość publikowanych treści tego typu. Według badań 98% wszystkich filmów *deepfake* dostępnych w sieci to treści pornograficzne; w 2023 r. w sieci pojawiło się 21 019 *deepfake porn* (co daje wzrost o 464% w porównaniu z 2022 r.); jedno na trzy narzędzia AI do tworzenia *deepfake* umożliwia

[39] Chyba że zachodzą wyjątki określone w ustawie, jak wizerunek osoby powszechnie znanej bądź stanowiącej jedynie szczegół całości, jak również zapłata za pozowanie (art. 81 ust. 1–2 upapp).

tworzenie pornografii; zaś dla stworzenia 60 sekund filmu pornograficznego z wizerunkiem dowolnej osoby wystarcza jeden wyraźny obraz jej twarzy i mniej niż 25 minut[40]. Powyższe wyniki wskazują więc na gwałtowny rozwój sektora *deepfake porn*, przyczyniając się do licznych obaw na tym polu. Rozpowszechnianie (a nawet samo tworzenie) takich sztucznych treści może wiązać się z wieloma negatywnymi konsekwencjami, w tym przede wszystkim ze szkodami oraz krzywdami, jakich doznaje osoba fizyczna, której wizerunek wykorzystano w ten sposób.

Zgodnie z polskim ustawodawstwem rozpowszechnianie wizerunku nagiej osoby lub osoby w tracie czynności seksualnej bez jej zgody stanowi przestępstwo oraz podlega karze (art. 191a k.k.[41]). Wydaje się jednak, iż zazwyczaj *deepfake porn* będzie sztucznie wygenerowaną lub zmanipulowaną treścią, w której podstawiono jedynie wizerunek twarzy konkretnej osoby fizycznej pod materiał o charakterze seksualnym. W takiej sytuacji nie zawiera on wizerunku tejże osoby jako nagiej czy w trakcie czynności seksualnej, co utrudnia subsumcję takiego stanu faktycznego pod omawianą normę[42]. W doktrynie zauważa się więc fragmentaryczność obecnej ochrony prawnej w tym zakresie oraz istnienie pewnej luki prawnej w zestawieniu z rozwojem *deepfake porn*, sugerując potrzebę rozwoju regulacji prawnych na tym polu[43].

Mimo to na gruncie europejskim stosunkowo nowa dyrektywa w sprawie zwalczania przemocy wobec kobiet i przemocy domowej zawiera regulację, gdzie czyn umyślny, w postaci wytwarzania lub manipulowania obrazami, nagraniami wideo lub podobnymi treściami, które sprawiają wrażenie, że osoba uczestniczy w czynnościach o wyraźnie seksualnym charakterze, „a następnie publiczne udostępnianie ich za pomocą technologii ICT, bez zgody tej osoby, gdy takie czyny mogą wyrządzić jej poważną szkodę",

40 Badanie Security Hero, *2023 State of Deepfakes, Realities, Threats and Impact*, okres badania: 15.07 – 9.08.2023, https://www.securityhero.io/state-of-deepfakes/ [dostęp: 2.12.2024].

41 Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny (t.j. Dz.U. z 2024 r., poz. 17), dalej: k.k.

42 Można by jednak rozważać dopuszczenie się niniejszego przestępstwa w stosunku do osoby, której wizerunek w trakcie czynności seksualnej wykorzystano do stworzenia deepfake'u z wizerunkiem twarzy innej osoby.

43 Agata Ziobroń, "Deepfake a prawo karne. Uwagi *de lege lata* i *de lege ferenda* dotyczące fałszywej pornografii," w *Studenckie Prace Prawnicze, Administratywistyczne i Ekonomiczne 37. Wyzwania dla państwa prawa i gospodarki w dobie pandemii*, red. Sebastian Jakubowski i Daria Kostecka-Jurczyk (Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 2021), 235.

powinien stanowić przestępstwo i podlegać karze (art. 5 ust. 1 pkt b) dyrektywy)[44]. Wydaje się więc, iż wiele spośród *deepfake porn* należałoby uznać za przejaw takowego, a przez to za przestępstwo (a więc działanie nielegalne)[45]. Odrębnym tematem jest kwestia utrwalania, posiadania, produkowania, rozpowszechniania, prezentowania i przechowywania treści pornograficznych z wytworzonym albo przetworzonym wizerunkiem małoletniego, uczestniczącego w czynności seksualnej, swoiście karana w naszym systemie prawnym[46].

Oprócz powyższego należy również zwrócić uwagę na inne możliwe prawnokarne konsekwencje tworzenia i rozpowszechniania *deepfake*, jak chociażby naruszenie art. 212 k.k. poprzez zniesławienie człowieka (w szczególności art. 212 § 2 k.k. za pomocą środków masowego komunikowania) czy art. 190a § 2 k.k. poprzez wykorzystanie wizerunku bądź innych danych człowieka, aby podszyć się pod inną osobę, w celu wyrządzenia jej szkody majątkowej lub osobistej. Możliwe byłoby również popełnienie przestępstwa oszustwa (art. 286 k.k.), np. poprzez popularne w ostatnim czasie rozmowy telefoniczne, gdzie sprawca podszywa się pod osobę bliską pokrzywdzonemu (dodatkowo realizacja znamion przestępstwa z art. 190a k.k.) i doprowadza ją do niekorzystnego rozporządzenia własnym lub cudzym mieniem, za pomocą wprowadzenia jej w błąd[47].

B. Kaczmarek-Templin, badając *deepfake* w kontekście dowodowym, wskazała również, iż „posługiwanie się sfałszowanym dokumentem

---

44 Można się jednak zastanawiać, czy taki rodzaj regulacji nie stanowi zbyt słabej ochrony, penalizując wyłącznie publiczne udostępnianie wskazanych treści, pomijając szkodliwość samego ich tworzenia bez zgody osoby. Dodatkowo wymóg „poważnej szkody" również będzie ograniczał zakres niniejszej normy, nie uwzględniając zwykłej szkodliwości, która również może wpływać negatywnie na sytuację ofiary oraz jej stan psychiczny.

45 Łatwość tworzenia i rozpowszechniana takich treści, a także ryzyko daleko idących konsekwencji nasuwają liczne obawy, przede wszystkim jednak z perspektywy osoby fizycznej, której wizerunek może zostać wykorzystany w takich materiałach. Warta rozważenia jest więc regulacja wprost zakazująca i penalizująca tworzenie *deepfake* o charakterze seksualnym/pornograficznym, dopasowana do specyfiki technologii, wprowadzająca szczególny tryb postępowania w takich przypadkach.

46 Por. np. Marcin Niedbała, "Dylemat odpowiedzialności karnej z tytułu generowania pornografii przy wykorzystaniu sztucznej inteligencji," *Krytyka Prawa. Niezależne Studia nad Prawem*, t. 15, nr 4 (2023), 59 – 63.

47 Por. np. Malwina Kuśmierek, *Tak działa nowe oszustwo na wnuczka. Stare oszustwo, nowe metody* (Spider'sWeb/Tech, 2024), https://spidersweb.pl/2024/10/oszustwo-na-wnuszka-deepfake-glos.html [dostęp: 2.12.2024].

w postaci zmodyfikowanego filmu (*deepfake*) rodzi konsekwencje prawne w postaci odpowiedzialności karnej", wskazując tu w szczególności na art. 270 k.k., który penalizuje podrabianie lub przerabianie dokumentu w celu jego użycia jako autentyczny bądź samo używanie go jako autentycznego[48]. Modyfikowanie materiału z użyciem AI, w celu użycia jako autentyczny (tu: dowód w sprawie), jak najbardziej mogło spełnić powyższe przesłanki, przyczyniając się do nielegalności już na etapie tworzenia deepfake'u.

4.5. Zwalczanie nieuczciwej konkurencji

Technologia *deepfake* z powodzeniem może być również wykorzystywana do promowania usług czy produktów. Prawdopodobne jest tworzenie treści, które będą zakłamywać rzeczywistość, np. wskazując na wyjątkową jakość własnych ofert bądź osłabiając opinię o konkurencji. Jeżeli *deepfake*, zawierający nieprawdziwe lub wprowadzające w błąd wiadomości o swoim lub innym przedsiębiorcy albo przedsiębiorstwie, zostanie rozpowszechniony w celu przysporzenia korzyści lub wyrządzenia szkody, owo działanie może zostać uznane za czyn nieuczciwej konkurencji oraz grozić konsekwencjami prawnymi (art. 14 uznk[49]). Stworzenie deepfake'u prezentującego takie informacje nie będzie jeszcze stanowić naruszenia samo w sobie, natomiast rozpowszechnienie tych treści, i to koniecznie mające na celu przysporzenie korzyści lub wyrządzenie szkody, mogłoby zostać za takowe uznane.

Mnogość regulacji prawnych, do których naruszenia może dość przy okazji tworzenia i rozpowszechniania deepfake'ów, relatywnie delikatne granice, które wydają się łatwo przekraczalne, a także specyfika technologii *deepfake* (w szczególności przykłady wykorzystujące informacje o prawdziwych osobach) prowadzą do konstatacji, iż legalne stworzenie i rozpowszechnianie deepfake'ów może okazać się dużym wyzwaniem, podczas którego wielokrotnie może dojść do naruszenia różnych norm prawnych, narażając się na konsekwencje prawne. Powyższe nie oznacza jednak, iż legalne stworzenie deepfake'u jest w naszym systemie prawnym niemożliwe, jednak wymaga ostrożności i poruszania się w ramach

48 Kaczmarek-Templin, "Deepfake i jego walor dowodowy w procesach cywilnych," 5.

49 Ustawa o zwalczaniu nieuczciwej konkurencji z dnia 16 kwietnia 1993 r. (t.j. Dz.U. z 2022 r., poz. 1233), dalej: uznk.

obwiązujących norm prawnych. Obniżenie standardów ochrony dóbr człowieka, celem ułatwienia tworzenia i rozpowszechniania deepfake'ów nie wydaje się jednak słusznym rozwiązaniem, prowadząc raczej do nieusprawiedliwionego wykorzystywania informacji o człowieku, osłabiając jego poczucie bezpieczeństwa, prywatności, realnego wpływu na kreowanie swojego wizerunku, dobrego imienia czy opinii, a także innych negatywnych konsekwencji.

Powyższe wyliczenie nie jest oczywiście wyczerpującym katalogiem możliwych naruszeń prawnych przy okazji tworzenia/rozpowszechniania deepfake'ów. Wymienione obszary wydają się jednak szczególnie narażone w obliczu rozwoju technologii *deepfake* oraz pozwalają na zaobserwowanie pewnych prawidłowości, gdzie możliwe jest wskazanie sytuacji niezgodności omawianego procesu z prawem. Nie powinno to jednak przesądzać o tezie nielegalności technologii *deepfake* w ogóle, bowiem możliwe jest wskazanie takich przypadków, gdzie ich tworzenie i rozpowszechnianie będzie prawnie dopuszczalne.

5 Wnioski

Obecnie problematyka prawna technologii *deepfake* dopiero zaczyna się rozwijać, poprzez wprowadzanie kolejnych regulacji prawnych. Jako zdecydowany plus należy oceniać wypracowanie definicji *deepfake*, uwzględnienie w niej także innych elementów rzeczywistości jak przedmioty czy miejsca, obowiązki przejrzystości, wyróżnienie przestępstwa tworzenia i rozpowszechniania za pomocą technologii ICT *deepfake porn*, obowiązki wprowadzenia mechanizmów zgłaszania treści nielegalnych dostawcom usług hostingu oraz zwalczania ich. Wciąż brakuje jednak jasnych wytycznych, kiedy konkretne *deepfake* będą stanowić treść nielegalną w świetle regulacji prawnych, uwzględniając specyfikę technologii, dostosowując przyjęte zasady i normy.

Bazując na ustaleniach poczynionych w niniejszym artykule, z punktu widzenia obecnie funkcjonujących regulacji prawnych wykazanie, czy dany materiał stanowi *deepfake*, czy też nie nie będzie miało większego znaczenia dla ustalenia legalności takiej treści. Podstawę dla tego wciąż stanowić będzie zgodność tworzenia i/lub rozpowszechniania danego materiału z normami obowiązującego prawa. Sam AI Act, mimo odniesienia

się w kilku miejscach do technologii *deepfake*, nie wprowadził nowych zasad dla wyróżnionego tu problemu badawczego. To zaś prowadzi do wniosku, iż tworzenie/rozpowszechnianie takich materiałów będzie prawnie dopuszczalne, o ile nie narusza obowiązujących przepisów prawnych. Zgodnie z poczynioną wyżej analizą istnieje cały szereg możliwych pól naruszeń, co może skutkować niezgodnością z prawem, która następuje już w momencie stworzenia i/lub rozpowszechnienia. Z punktu widzenia osoby tworzącej *deepfake* konieczność wzięcia pod uwagę tak wielu różnych przepisów przy ocenie legalności procesu tworzenia może powodować nieczytelność i przyczyniać się do generowania treści z ich pominięciem.

Wobec tak przeprowadzonej analizy oraz uzyskanych tu wniosków wyróżniło się kilka potrzeb regulacyjnych, których wprowadzenie mogłoby wpłynąć pozytywnie na kwestię ustalania nielegalności określonych, szkodliwych treści *deepfake* oraz zwalczanie ich. Wśród nich można wskazać na potrzebę: 1) ujednolicenia zauważonych niejednoznaczności w zakresie definicyjnym *deepfake* (por. s. 679–680) i stworzenia jednolitego stanowiska, niezbędnego dla właściwej kwalifikacji; 2) wyodrębnienia grup *deepfake*, które ze względu na pewne kwestie (np. charakter pornograficzny) powinny stać się przedmiotem wzmożonej regulacji prawnej, obarczone dodatkowymi obowiązkami i ograniczeniami, w interesie ochrony praw człowieka; 3) prawnego wyróżnienia sytuacji, gdy określone typy *deepfake*, sposoby wykorzystania technologii czy konkretne przejawy ich użycia stanowią działanie nielegalne, powiązane z określonymi konsekwencjami prawnymi (oczywiście głównie co do wyraźnych sytuacji naruszeń, o poważnych następstwach dla człowieka). Warte rozważenia jest także wypracowanie zasad nielegalności *deepfake ex lege*, ze względu na określone przesłanki jego powstania bądź funkcjonowania, np. nielegalny cel, działanie wprost na niekorzyść przedstawionego podmiotu, manipulowanie opinią publiczną. Takie zbiorcze określenie przesłanek nielegalności przyspieszałoby proces zwalczania szkodliwych treści (naruszenie zasady – bezpośrednie konsekwencje) oraz w ten sposób ułatwiało dochodzenie odpowiedzialności samym poszkodowanym.

Rozwój sektora *deepfake* rodzi więc wyzwania nie tylko dla aspektów technologicznych, ale także prawnych i etycznych. Pojawiające się możliwości stwarzają pole dla wykorzystania *deepfake* także w negatywnych celach, co wiąże się z zagrożeniem dla człowieka i jego dóbr. Istniejące w naszym systemie regulacje prawne jedynie częściowo mogą stanowić podstawę dla oceny tej technologii, jednak w obliczu wielokrotnie wspominanych już aspektów często wydają się niewystarczające, co implikuje

potrzebę wypracowania bardziej dopasowanych norm. Właściwa regulacja prawna dla technologii *deepfake* jest więc niezbędnym krokiem prawidłowego i etycznego rozwijania tej technologii.

Bibliografia

Bartuzi-Trokielewicz, Ewelina, Michał Ołowski i Adrain Kordas. *Deepfake: Jak sztuczna inteligencja może nas oszukiwać?* Cyberprofilaktyka NASK. https://cyberprofilaktyka.pl/blog/deepfake-jak-sztuczna-inteligencja-moze-nas-oszukiwac _i40.html.

Chataut, Robin i Aadesh Upadhyay. "Introduction to Deepfake Technology and Its Early Foundations." W *Deepfakes and Their Impact on Business*, red. Gaurav Gupta, Sailaja Bohara, Raj K. Kovid i Kapil Pandla, 1 – 18. USA: IGI Global, 2024.

Edwards, Lilian. *The EU AI Act: A Summary of Its Significance and Scope*. Ada Lovelace Institute, 2022.

Flisak, Damian. *O niełatwym wytyczaniu granic rzeczywistości przez AI Act: zasada transparentności, deepfake*. Warszawa: Prawo i Nowe Technologie 1/2024, Wydawnictwo C.H. Beck, Legalis/el., 2024.

Ice, Jessica. "Defamatory Political Deepfakes and the First Amendment." *Case Western Reserve Law Review*, vol. 70 (2) (2019): 417 – 454.

Kaczmarek-Templin, Berenika. "Deepfake i jego walor dowodowy w procesach cywilnych." W *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, red. Kinga Flaga-Gieruszyńska, Jacek Gołaczyński i Dariusz Szostek. Warszawa: C.H. Beck, Legalis/el., 2019.

Liu, Min i Xijin Zhang, *Deepfake Technology and Current Legal Status of It*. Proceedings of the 2022 3rd International Conference on Artificial Intelligence and Education, IC-ICAIE 2022, AHCS 9, 2023.

Łabuz, Mateusz. "Deep Fakes and the Artificial Intelligence Act – an Important Signal or a Missed Opportunity?" *Policy & Internet*, nr 16 (4) (2024): 1 – 18. https://doi.org/10.1002/poi3.406.

Łabuz, Mateusz. "Regulating Deep Fakes in the Artificial Intelligence Act." *Applied Cybersecurity & Internet Governance*, vol. 2 (1) (2023): 252 – 291. https://doi.org/10.60097/ACIG/162856.

Niedbała, Marcin. "Dylemat odpowiedzialności karnej z tytułu generowania pornografii przy wykorzystaniu sztucznej inteligencji." *Krytyka Prawa. Niezależne*

Studia nad Prawem, t. 15, nr 4 (2023): 56 – 68. https://doi.org/10.7206/kp.2080-1084.638.

Regan-Balcarczyk, Justyna. "art. 4." W *Kodeks cywilny. Komentarz*, wyd. 4, red. Mariusz Załucki. Warszawa: Legalis/el., 2024.

Romero Moreno, Felipe. "Generative AI and Deepfakes: A Human Rights Approach Tackling Harmful Content." *International Review of Law, Computers & Technology*, vol. 38, nr 3 (2024): 297 – 326. https://doi.org/10.1080/13600869.2024.2324540.

Sylwestrzak, Anna. "art. 23." W *Kodeks cywilny. Komentarz*, wyd. 2, red. Małgorzata Balwicka-Szczyrba i Anna Sylwestrzak. Warszawa: Wolters Kluwer, Lex/el., 2021.

Szpyt, Kamil. "Sztuczna inteligencja i nowe technologie (nie zawsze) w służbie ludzkości, czyli cywilnoprawna problematyka rozwoju i popularyzacji technologii deepfake." W *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*. red. Kinga Flaga-Gieruszyńska, Jacek Gołaczyński i Dariusz Szostek. Warszawa: C.H. Beck, Legalis/el., 2019.

Wasiuta, Olga i Sergiusz Wasiuta. "Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość." *Annales Universitatis Paedagogicae Cracoviensis, Studia de Securitate*, nr 9 (3) (2019): 19 – 30. https://doi.org/10.24917/26578549.9.3.2.

Ziobroń, Agata. "Deepfake a prawo karne. Uwagi *de lege lata* i *de lege ferenda* dotyczące fałszywej pornografii." W *Studenckie Prace Prawnicze, Administratywistyczne i Ekonomiczne 37. Wyzwania dla państwa prawa i gospodarki w dobie pandemii*, red. Sebastian Jakubowski i Daria Kostecka-Jurczyk, 225 – 237. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 2021.