Kwartalnik Krajowej Szkoły Sądownictwa i Prokuratury
Quarterly of the National School of Judiciary and Public Prosecution
Zeszyt 3 (51)/2023

ARTYKUŁY

Deepfake jako narzędzie do przekazywania informacji fałszywej i domniemanej. Analiza prawnokarna i cybernetyczna

DOI: 10.53024/5.3.51.2023

KS. KRZYSZTOF MAREK KIEŁPIŃSKI*

STRESZCZENIE

Od 2017 r. zjawisko *deepfake* pojawiło się w przestrzeni wirtualnej. Bardzo szybko zostało wykorzystane w różnych obszarach ludzkiego życia. Rozwój środowiska cyfrowego, *social mediów*, różnych form rozrywki online – wszystko to sprawiło, że wymiana informacji między ludźmi została przeniesiona do sieci. Nowe zjawisko wymusiło na badaczach podjęcie próby jego charakterystyki. Niestety, w literaturze polskiej i obcojęzycznej brakuje monografii i artykułów, które kompleksowo opisywałyby zagrożenia w świecie wirtualnym, z których jednym jest technologia *deepfake*. Polega ona na przekazywaniu informacji fałszywej lub domniemanej. Dokonuje się to za pomocą zdjęć lub filmów. Niekiedy nie jest możliwe zweryfikowanie autentyczności materiałów. Z uwagi na różnorodność technologii *deepfake'ów* użytkownicy sieci oraz wielu aplikacji nie zdają sobie sprawy, z jakiego typu zagrożeniem się spotykają. Artykuł opisuje *deepfake'i* jako narzędzie do przekazywania informacji fałszywej i domniemanej. Analiza została przeprowadzona w dwóch wymiarach: prawnokarnym i cybernetycznym. W tym celu użyto wielu metod badawczych, m.in. dogmatycznoprawnej, historycznej, porównawczej i filologicznej. Wśród wielu wniosków najważniejszy z nich to ten, że *deepfake* jest nową formą manipulacji i dezinformacji. Technologia tego typu umożliwia popełnienie wielu przestępstw, które podczas toczących się postępowań karnych są trudne do udowodnienia sprawcy. Tylko współpraca wielu instytucji państwowych, organizacji pozarządowych, producentów oprogramowania lub aplikacji pomoże wygrać walkę z tego typu zjawiskiem.

* Doktor nauk prawnych; ORCID: 0000-0001-8168-2514.

Słowa kluczowe: *deepfake*, manipulacja, informacja, dezinformacja, przestępstwo, cybernetyka

WSTĘP

Vitae semper reformanda – tę starożytną paremię należy również odnieść do świata cyfrowego. Wirtualna rzeczywistość nie tylko się zmienia, a jej cechą charakterystyczną jest zawrotne tempo w określaniu ważnych dla ludzkości wydarzeń i wzywań cywilizacyjnych. Obserwując zmieniający się świat online, nie można przejść obojętnie obok zjawiska *deepfake*, które w ostatnich latach stało się punktem zainteresowania niektórych ekspertów. Warto podkreślić, że tego typu zjawisko jest wszechobecne w przestrzeni wirtualnej. Zmieniająca się technologia oraz powstanie wielu aplikacji przyczyniły się do ukształtowania wspomnianego zjawiska. Polega ono na kreowaniu nieprawdziwych informacji lub obrazów, a jednak łudząco podobnych do oryginału, dzięki czemu można zmienić wiele ważnych obszarów społecznych. Celem niniejszego artykułu będzie analiza wspomnianej nowości wirtualnej. Przedmiotem rozważań i badań będzie *deepfake* w podwójnym ujęciu: prawnokarnym i jakościowej teorii informacji. Argumentem przemawiającym za przeprowadzeniem rozważań w tym obszarze jest luka badawcza. Należy podkreślić, że w polskiej przestrzeni naukowej brakuje kompleksowej monografii i artykułów naukowych na ten temat. Podobny problem występuje na międzynarodowym gruncie, chociaż niektórzy autorzy opisywali zjawisko *deepfake* w różnym kontekście[1].

DEZINFORMOWANIE I PARAINFORMOWANIE – PODSTAWY DLA TWORZENIA *DEEPFAKE'ÓW*

Narrację dotyczącą *deepfake'ów* należy rozpocząć od wyjaśnienia dwóch kluczowych sposobów przekazywania informacji. Zalicza się do nich dezinformowanie i parainformowanie. Ich celem jest przekazanie odbiorcy dwóch różnych informacji: fałszywej oraz domniemanej. Charakterystyką i zdefiniowaniem tych terminów najpełniej zajęła się jakościowa teoria informacji. Na uwagę zasługuje analiza, którą przeprowadził M. Mazur – polski ekspert z zakresu nauk cybernetycznych. Zatrzymajmy się przy pojęciu dezinformowania. M. Mazur doszedł do wniosku, że proces dezinformacji ma na celu oddzielenie wszystkich łańcuchów kodowych. Niestety, daje się zauważyć, że niektóre z informacji nie są pełne. Idąc drogą wskazaną przez polskiego uczonego, należy podkreślić, że pod pojęciem dezinformowania kryje się czynność,

[1] I. Dąbrowska, *Deepfake – nowy wymiar internetowej manipulacji*, „Zarządzanie Mediami" 2020, nr 2, s. 89–101; O. Wasiuta, S. Wasiuta, *Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate" 2019, nr 9, s. 19–31; *eidem*, *FakeApp jako nowe zagrożenie bezpieczeństwa politycznego i informacyjnego*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate" 2019, nr 9, s. 129–139; P. Muniak, W. Kulesza, *Sztuka dezinformacji*, „Newsweek" 2022, nr 30, s. 90–92; W. Sokała, *Jak przegrać wygraną wojnę*, „Dziennik Gazeta Prawna" 2022, nr 156, s. 8–9.

która tworzy łańcuchy zawierające kody (dane) jako niepełne struktury[2]. Dezinformacja to rodzaj informacji, w której brakuje obrazów w zbiorze, jaki go tworzą. Tego rodzaju informacja może przybrać postać: symulacyjną, dysymulacyjną i konfuzyjną. Symulacyjny wymiar przekazywania informacji polega na tym, że niektóre łańcuchy kodowe nie zawierają oryginałów. Informacja jest skutkiem transformacji, wskutek czego powstaje nowy obraz, który nie ma nic wspólnego z oryginałem[3]. Dysymulacyjny wymiar wskazuje na proces, w którym wybrane łańcuchy kodowe w ogóle nie zawierają obrazów. W ten sposób przerywa się przesyłanie danych i tworzenie informacji[4]. Konfuzyjny wymiar dokonuje się, gdy informacja w łańcuchach danych – kodach nie zawiera oryginałów oraz obrazów. Powstaje na skutek połączenia dezinformacji symulacyjnej z dysymulacyjną[5].

W przypadku dezinformacji należy podkreślić, że jej celem jest przekazanie informacji fałszywej odbiorcy, który będzie przekonany, że jest ona prawdziwa. Brak odniesienia do oryginałów przez odbiorcę sprawia, że bardzo trudne lub niekiedy niemożliwe staje się zweryfikowanie prawdziwości informacji lub obrazu.

Drugim ważnym sposobem przekazywania informacji jest parainformowanie. To proces, który tworzy informację z łańcucha danych lub obrazów nienależących do żadnego łańcucha kodowego. Parainformowanie nie występuje samodzielnie, lecz stanowi uzupełnienie procesu informowania, który polega na transformowaniu informacji zawartych w łańcuchu oryginałów w informacje zawarte w łańcuchu obrazów[6]. Skutkiem parainformowania jest parainformacja. W przypadku gdy parain-

---

[2] Por. M. Mazur, *Jakościowa teoria informacji*, Warszawa 1968, s. 140–141.

[3] Por. B. Piasecki, *Kontrwywiad atak i obrona*, Łomianki 2021, s. 244; A. Kowalski, *Kontra. Sztuka walki z wywiadem przeciwnika*, Łomianki 2021, s. 57–97; H. Münkler, *Wojny naszych czasów*, Kraków 2004, s. 97–110. Przykłady dezinformowania symulacyjnego to: stworzenie fałszywych dokumentów, wykorzystanie cudzych podpisów, tworzenie tzw. legendy dla oficera wywiadu, sporządzanie fałszywych pokwitowań, tworzenie nieprawdziwych meldunków, alarmowanie o zagrożeniu, mimo że w rzeczywistości ono nie występuje oraz szeroko rozumiana propaganda.

[4] M. Mazur, *op. cit*., s. 143: Przykładem dezinformacji dysymulacyjnej jest przypadek, który może pojawić jako niezamierzone działanie, np. cennik jakiegoś towaru, który został pominięty, a w rzeczywistości jest dostępny w sprzedaży, rozkład jazdy autobusów, pociągów i samolotów, gdzie brak wzmianki o dacie i godzinie odjazdu, a w rzeczywistości środek transportu wykonuje zaplanowany kurs, oraz w obszarze propagandy, gdy przemilcza się kompromitujące fakty i okoliczności o przeciwnikach politycznych, wrogach społecznych, grupach etnicznych, jednostkach, aferach oraz o faktach, które dobrze świadczą o przeciwnikach i wrogach.

[5] *Ibidem*, s. 146–147: Ten rodzaj dezinformowania jest związany z ludzką słabością. Człowiek wielokrotnie popełnia pomyłki, jest niedokładny i nonszalancki. Występuje najczęściej na skutek ludzkiego błędu. Można się z nią spotkać, gdy rozkład jazdy autobusów, samolotów, pociągów i innych środków transportu zawiera na skutek ludzkiego błędu inną godzinę odjazdu lub przyjazdu. W sklepie, gdy ktoś umieścił w cenniku mylną cenę jakiegoś towaru oraz gdy ktoś poda komuś zły numer telefonu lub zły adres zamieszkania. Zalicza się do tej kategorii również przeinaczenia, m.in. przerobienie cyfr na czeku lub pokwitowaniu, gdy ktoś występuje pod fałszywym nazwiskiem lub gdy w zestawieniu lub bilansie widnieją różne kwoty.

[6] *Ibidem*, s. 156–158: Przykładem tego typu informowania są środki językowe, mimika i gestykulacja, którą człowiek wykorzystuje podczas komunikowania się z drugim człowiekiem. Innymi przykładami parainformowania są aluzja, przenośnia, odgadnięcie ludzkich intencji, przeniknięcie cudzych zamiarów, zrozumienie zaleceń zawartych w przypowieściach.

formacje w zbiorze obrazów różnią się od parainformacji zwartych w oryginałach, to parainformowanie będzie pełniło rolę dezinformowania. Informacja domniemana w takim przypadku przybiera postać informacji fałszywej, a proces ten nazwa się paradezinformowaniem. Na skutek tego działania mamy do czynienia z określonymi konsekwencjami, mianowicie przekazywanie tego rodzaju informacji będzie niczym innym jak paradezinformowaniem, polegającym na procesie parainformacji, czyli asocjacji obrazów lub braku parainformacji, która może powstać w wyniku paradezinformowania.

Kontynuując ten wątek, możemy wyróżnić trzy obszary paradezinformowania. Pierwszy z nich ma charakter symulacyjny. Jest to proces, w którym parainformacja występuje w zbiorze obrazów, natomiast nie występuje w zbiorze oryginałów. Najczęściej przybiera to postać domniemania ukrytego sensu w zdaniach wypowiedzianych tylko w dosłownym znaczeniu, dopatrywania się aluzji, której nie było, a także przypisywania innym intencji i zamiarów, których nie mieli, albo fałszywe wyobrażenia o nieistniejących uczuciach innej osoby[7]. Kolejny obszar ma charakter dysymulacyjny. Cechuje go to, że parainformacje występują w zbiorze oryginałów, natomiast brak ich w zbiorze obrazów. Ten rodzaj występuje najczęściej wtedy, gdy chodzi o niedomyślanie, się o co chodzi rozmówcy, nierozumienie aluzji, nieprzeniknięcie cudzych zamiarów, nieodgadnięcie cudzych intencji oraz nierozpoznanie cudzych uczuć[8]. Ostatni wymiar paradezinformowania stanowi jej konfuzyjny charakter. Przejawia się tym, że parainformacje występujące w zbiorze obrazów różnią się od parainformacji występujących w zbiorze oryginałów. Konfuzyjny charakter może przybrać postać pojedynczą lub podwójną. Z pojedynczą mamy do czynienia, gdy osoba domyśla się czegoś innego, niż miał na myśli rozmówca, dopatruje się nie tej aluzji, o którą chodziło rozmówcy, przypisuje komuś intencje odmienne od rzeczywistych, posądza o pewne zamiary kogoś mającego inne, bierze oznaki pewnych uczuć za oznaki innych, pomniejsza lub wyolbrzymia sensu cudzych wypowiedzi[9]. Podwójna postać konfuzyjnego wymiaru będzie polegać na tym, że osoba będzie opacznie interpretować cudze zachowanie, dopatrywać się wrogości w tym, w czym jej nie było, uznawać za zachętę wypowiedzi zniechęcające i przeciwnie. Ta postać najczęściej występuje w grach karcianych, w sztuce, podczas opowiadania dowcipów, anegdot, a także w trakcie spotkań towarzyskich z dużą dozą humoru[10].

GENEZA ZJAWISKA *DEEPFAKE*

Powyższe rozważania znalazły zastosowanie w świecie cyfrowym. W 2017 r. dzięki technologii wirtualnej powstał pierwszy fałszywy film pornograficzny, który został

[7] *Ibidem*, s. 160–161.
[8] *Ibidem*, s. 161.
[9] *Ibidem*, s. 163.
[10] *Ibidem*, s. 165–166.

umieszczony w serwisie internetowym Reddice. Przedstawiał aktorkę Gal Gadot jako główną bohaterkę seksualnych igraszek – w ten sposób anonimowy twórca naruszył jej dobra osobiste. Anonimowy twórca określił siebie za pomocą nicku *deepfake* i to od niego pochodzi używany współcześnie termin[11]. Przywołany twórca w grudniu 2017 r. ponownie wykorzystał technologię, dodając twarze celebrytów aktorom grającym w filmach pornograficznych. Od tego momentu w sieci internetowej można było natknąć się na filmy z udziałem polityków i celebrytów o różnym zabarwieniu gatunkowym. Termin *deepfake* dotyczy zarówno technologii, jak i tworzenia obrazów oraz filmów. Technologia związana z tworzeniem fałszywych informacji i obrazów przejawia się w retuszu zdjęć oraz ich modyfikowaniu. Rozwój tego zjawiska był uzależniony od rynku z aplikacjami mobilnymi (m.in. *FaceApp*), które umożliwiały dokonywanie zmian w obrębie fotografii cyfrowych przez posiadacza aplikacji, wedle jego upodobań[12]. Aplikacje te posiadały praktyczny interfejs oraz błyskawicznie pokazywały efekty w postaci nowych obrazów. Stały się idealnym narzędziem dla osób, które brały udział w zbieraniu internetowych polubień – lajków. Niektóre aplikacje otwierały nowe perspektywy dla celebrytów, osób prywatnych poprzez to, że zawierały filtry i nakładki do obróbki zdjęć. Dzięki temu fotografie wstawiane w *social mediach* były piękniejsze i niepowtarzalne.

Zdaniem I. Dąbrowskiej tego typu praktyki najczęściej przybierały niewinne nazwy, m.in. korygowanie, poprawianie, upiększanie. Według autorki właściwym terminem na określenie tego typu działań jest fałszerstwo. Aplikacje, gdyby jest nazwać oszukiwaczami fotograficznymi, zapewne nie cieszyłyby sie zainteresowaniem wśród potencjalnych użytkowników[13]. Warto zauważyć, że m.in. w aplikacji FaceApp można się natknąć na różne filtry, które też zostały nazwane bardzo niewinnie. W związku z funkcjonowaniem aplikacji narodziła się globalna moda na najlepsze selfie. Większość ludzi i obserwatorów *social mediów* zwraca uwagę na pozytywny wymiar tego zjawiska – osoby będące na zdjęciu są uśmiechnięte, a w tle pojawiają się interesujące krajobrazy, wydarzenia i zjawiska.

Tymczasem moda na najlepsze selfie zabrała już setki istnień ludzkich. Potwierdza to zawierający porażające dane raport przygotowany przez naukowców z uczelni

[11] O. Wasiuta, S. Wasiuta, *Deepfake jako skomplikowana...*, s. 21: Termin *deepfake* związany jest z użytkownikiem o nazwie „DeepFakes", który w grudniu 2017 r. opublikował na portalu Reddit kilka internetowych filmów pornograficznych, wykorzystując sztuczną inteligencję do podmieniania twarzy aktorek na twarze m.in.: Daisy Ridley, Emmy Watson, Gal Gadot czy Scarlett Johansson. Materiały pornograficzne oczywiście były fałszywe, ale wykonano je w sposób bardzo realistyczny. Filmy są tworzone przez załadowanie złożonego zestawu instrukcji do komputera wraz z dużą liczbą zdjęć i nagrań dźwiękowych. Następnie program komputerowy uczy się, jak naśladować i odtwarzać mimikę danej osoby, jej głos, ruchy, indywidualne maniery, intonację oraz rodzaj używanego słownictwa. Wystarczająca liczba filmów i zapisów dźwiękowych danej osoby umożliwia systemowi stworzenie nagrania; M. Brundage, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Executive Summary February 2018, s. 49; https://niezalezna.pl/253924-scarlett-johansson-na-tropie-deepfakeow [dostęp: 12.09.2022].

[12] I. Dąbrowska, *op. cit.*, s. 91–92; O. Wasiuta, S. Wasiuta, *FakeApp jako nowe zagrożenie...* s. 129–130.

[13] I. Dąbrowska, *op. cit.*, s. 92–93.

medycznych w New Delhi: *Selfie: dobrodziejstwo, czy zmora?*[14]. Przeanalizowała go I. Dąbrowska. W latach 2011–2017 na całym świecie z powodu robienia zdjęcia selfie zginęło 259 osób. Najczęstszą przyczyną ich śmierci były utonięcie oraz wypadek komunikacyjny. Ponadto inne powody zgonu to: upadek z wysokości, kontakt ze zwierzęciem, bronią lub porażenie prądem. Najwięcej zgonów odnotowano w Indiach, Rosji, Stanach Zjednoczonych oraz Pakistanie. 85% ofiar stanowiły osoby bardzo młode lub młode (od 10 do 30 r.ż.)[15]. Reasumując powyższe dane, można pokusić się o postawienie tezy, że skoro ludzie potrafią zaryzykować, niekiedy stracić swoje życie dla zdjęcia, którym będą mogli pochwalić się w sieci, to tym bardziej zjawisko *deepfake* znajdzie amatorów i użytkowników. *Deepfake* zyskuje popularność z powodu niefrasobliwości ludzi umieszczających różne zdjęcia i obrazy w sieci. W następnym podrozdziale warto zdefiniować, na czym polega istota *deepfake*.

NATURA *DEEPFAKE*

Zjawisko *deepfake* zdefiniował N. Young. Według niego to technologia informatyczna, która wykorzystuje sztuczną inteligencję w celu tworzenia lub edytowania treści wideo albo obrazu, by pokazać coś, co nigdy się nie wydarzyło[16]. W przypadku wideo zjawisko *deepfake* polega na użyciu dwóch konkurencyjnych systemów, w którym pierwszy z nich to generator, zaś drugi jest określany jako dyskryminator. Generator tworzy fałszywy obraz, a rolą dyskryminatora jest ustalenie prawdziwości lub fałszywości nowego obrazu. Zdaniem M. Labbba, J. Burke i R. Priesta zjawisko *deepfake* łączy w sobie dwa zagadnienia: uczenia się obsługi sztucznej inteligencji w różnych obszarach życia, m.in. bezpieczeństwa, finansów, informatycznym, edukacyjnym oraz fałszerstwa. Wspomniani badacze podkreślają, że w przypadku obrazów *deepfake* stanowi ich syntezę dokonaną przy użyciu sztucznej inteligencji[17] i polega na łączeniu i nakładaniu istniejących obrazów i filmów na obrazy i wideo źródłowe za pomocą specjalnej technologii uczenia maszynowego. O. Schwartz zauważył, że technika uczenia się maszynowego była ograniczona aż do 2017 r. Wtedy właśnie przez społeczność badawczą AI został stworzony generator, następnie zbudowano

[14] A. Bansal, Ch. Garg, A. Pakhare, S. Gupta, *Selfies: A born or bane?*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6131996/pdf/JFMPC-7-828.pdf [dostęp: 12.09.2022].

[15] I. Dąbrowska, *op. cit*., s. 93.

[16] N. Young, *Deepfake Technology: Complete Guide to Deepfakes, Politics and Social Media*, New York 2019, s. 13–14.

[17] M. Labbe, *Microsoft deepfake software comabts election propaganda*, w: https://www.techtarget.com/searchenterpriseai/news/https://www.techtarget.com/searchenterpriseai/news/252488582/Microsoft-deepfake-software-combats-election-propaganda252488582/; J. Burke, *Why Itleaders need to be aware of deepfake security risks*, https://www.techtarget.com/search/query?q=deepfake&type=article&pageNo=1&sortField=; R. Priest, *Deepfake it until you make it*, w: https://www.computerweekly.com/blog/Downtime/Deepfake-it-until-you-makeit?_ga=2.245290114.623414774.1663061863852034460.1663061863&_gl=1*1b2n5ic*_ga*ODUyMDM0NDYwLjE2NjMwNjE4NjM.*_ga_TQKE4GS5P9*MTY2MzA2MTg2Mi4xLjEuMTY2MzA2MjU5OC4wLjAuMA [dostęp: 13.09.2022]; I. Dąbrowska, *op. cit*., s. 90.

sieć GAN (*generative adversarial networks*) przy użyciu darmowego oprogramowania Tensor Flow, który nakładał twarze celebrytów na ciała kobiet w filmach pornograficznych[18]. Sieć GAN stanowi algorytm, który umożliwia nakładanie innych zdjęć lub obrazów na obiekty źródłowe. Zdaniem Younga ten proces w przestrzeni graficznej nazywa się nakładaniem[19]. W 2018 r. powstała wykorzystująca go aplikacja *FakeApp*. Zdaniem I. Dąbrowskiej oraz O. i S. Wasiutów daje ona możliwość użytkownikom tworzenia obrazów lub filmów dzięki zmienieniu zdjęcia lub jego części. Nowy obraz można umieścić w sieci oraz *social mediach*[20]. W podobnym duchu wypowiedzieli się na temat zjawiska *deepfake* W. Kulesza oraz P. Muniak. Obaj eksperci definiują tego typu technologię jako hiperrealistyczną cyfrową podmianę twarzy i głosu. Dokonuje się to w taki sposób, że ciężko rozszyfrować montaż. Podkreślają, że zjawisko *deepfake* bazuje na zawansowanej technologii, która wykorzystuje cechy fizyczne, fizjologiczne i behawioralne człowieka. Cechą wyróżniającą jest, że nowe obrazy powstałe dzięki *deepfake* mają charakter wirusowy, czyli są nieprzewidywalne oraz bez kontroli rozpowszechniane w Internecie[21].

W ramach prowadzonych rozważań należy wyjaśnić różnicę pomiędzy *fake news* a *deepfake*. Termin *fake news* odnosi się do wiadomości medialnej, która nie niesie ze sobą prawdy ani fałszu. Opiera się na dezinformacji i często zawiera prawdziwe fragmenty. Słownik języka polskiego definiuje *fake news* jako zabieg manipulowania faktami, chętnie stosowany przez dziennikarzy, których celem podczas przygotowania publikacji jest jak największe zainteresowania tematem, a nie jego zgodność z rzeczywistością[22]. M. Drzazga określa *fake news* jako działanie, które ma na celu wprowadzić w błąd słuchacza, po to, by osiągnąć korzyści finansowe, polityczne, społeczne i propagandowe[23]. *Deepfake* natomiast to nowy wymiar internetowej manipulacji w postaci fałszerstw obrazów i wideo. Zdaniem I. Dąbrowskiej ustawodawstwo nie nadąża w kwestii *deepfake* , ponieważ jest to zjawisko stosunkowo świeże, ulegające rozpowszechnianiu. Większość użytkowników w świecie wirtualnym upowszechnia nagrania lub obraz, nie mając świadomości, w jaki sposób powstały[24]. *Deepfake* ma na celu stworzenie fałszywego obrazu lub wideo, które mogą być różnorodnie wykorzystane. W dalszej części narracji zostaną przedstawione różne typy *deepfake*, które występują w przestrzeni wirtualnej.

---

[18] O. Schwartz, *You Thought fake news was bed? Deefakes are where truth goes to die*, https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth [dostęp: 13.09.2022].

[19] N. Young, *op. cit*., s. 15–16.

[20] I. Dąbrowska, *op. cit.,* s. 90; O. Wasiuta, S. Wasiuta, *Deepfake - nowy wymiar internetowej...*, s. 132–133.

[21] P. Muniak, W. Kulesza, *op. cit*., s. 90.

[22] https://sjp.pwn.pl/ [dostęp: 14.09.2022].

[23] M. Drzazga, *Cała prawda o fake news, czyli jak rozpoznać fałszywe wiadomości*, w: https://www.legalniewsieci.pl/aktualnosci/cala-prawda-o-fake-news-czyli-jak-rozpoznac-falszywe-wiadomosci [dostęp: 14.09.2022].

[24] I. Dąbrowska, *op. cit*., s. 91.

TYPY *DEEPFAKE*

Od 2017 r. zjawisko *deepfake* występuje w wielu obszarach przestrzeni wirtualnej. Twórcy tej technologii posłużyli się nią na potrzeby m.in. polityki, władzy, biznesu i kinematografii. Wspomniana technologia miała zastosowanie w licznych produkcjach amerykańskich, m.in. *Gra o tron*, *Gwiezdne wojny*, *Szybcy i wściekli*, *Ludzie honoru* i wiele innych. W tym przypadku wykorzystywana była najczęściej do stworzenia cyfrowej repliki. Technologię *deepfake* stosuje się, aby nieżyjącego aktora umieścić w kolejnym wątku serii, która cieszyła się popularnością wśród widzów. Jednak wykorzystanie omawianej technologii może służyć dezinformacji i mieć charakter przestępczy. W przypadku fałszowania obrazów z ludzką twarzą cel tego zabiegu jest bardzo prosty. Chodzi o zniszczenie reputacji konkretnej osoby oraz naruszenie dóbr osobistych. W 2018 r. powstał film z prezydentem USA Barackiem Obamą wypowiadającym słowa, które nigdy w rzeczywistości nie padły w przestrzeni publicznej. Wszystko po to, żeby zdestabilizować relacje dyplomatyczne USA z partnerami. Kolejnym przykładem są działania wymierzone przeciwko aktorowi Nicolasowi Cage'owi, które miały na celu zmniejszenie popularności aktora oraz rentowności produkcji filmowej, w której występował. Na tej podstawie można zaobserwować, że omawiana technologia może naruszać dobra osobiste człowieka i prawa podmiotowe, a z drugiej strony być pomocna podczas tworzenia kolejnych odcinków seriali z nieżyjącymi już aktorami. Użycie jest uzależnione od tego, jaki cel ma zostać zrealizowany.

W 2019 r. użytkownicy mogli zainstalować na telefonach komórkowych program *Deep Nude*. Powstał on w celu tworzenia realistycznych obrazów nagich kobiet za pomocą przesłanych zdjęć prawdziwych osób. Mimo że wycofano tę aplikację, nadal jest dostępna na niektórych stronach internetowych. W tym samym roku, wykorzystując *deepfake*, zhakowano obraz *Mona Lisa*. Za pomocą sztucznej inteligencji stworzono krótkie animacje na bazie obrazu Leonarda da Vinci. Algorytmy kopiowały mimikę bohaterki płótna, a następnie dzięki specjalnym znacznikom były w stanie animować nieruchome obrazy i zdjęcia twarzy. Ożywienie Mona Lisy stało się hitem w internecie, a jej „nowe" wersje były rozpowszechniane przez nieświadomych użytkowników internetu. W tym samym roku za pomocą *deepfake* stworzono film, który miał na celu zwiększenie świadomości na temat procederu porywania dzieci w Pakistanie[25]. 24 lutego 2022 r. rozpoczęła się wojna pomiędzy Federacją Rosyjską a Ukrainą, a już 16 marca 2022 r. za pomocą technologii *deepfake* zaatakowano kanał telewizyjny Ukraine 24. Na tym kanale transmitowano wystąpienie prezydenta Ukrainy Wołodymyra Zełenskiego. Podczas przemówienia rzekomo wzywał swoich rodaków do porzucenia broni w obliczu rosyjskiej agresji. Przemówienie błyskawicznie umieszczono na YouTube, Facebooku, Twitterze. W tym przypadku chciano rozsiać fałszywą informację. Większość ludzi podeszła do tego bardzo krytycznie

[25] P. Muniak, W. Kulesza, *op. cit*., s. 91.

i nie dała się nabrać na te rewelacje. Powyższe przykłady pokazują zróżnicowanie typów *deepfake*. Zdaniem I. Dąbrowskiej można wyróżnić cztery ich grupy[26]. Pierwszy typ to obszar rozrywki. *Deepfake* ma charakter ludyczny. W jego produkcji twórcy nawiązują do popkultury, głównymi bohaterami są celebryci, aktorzy, osoby animowane lub fikcyjne. Drugi typ dotyczy edukacji. Najczęściej technologia wykorzystuje wizerunki znanych osób, które nie żyją, aby przybliżyć widzom historię ich życia. Trzeci obszar jest związany z dezinformacją. Użycie *deepfake* ma wywołać szum medialny oraz niepokój społeczny. W tym kontekście w głównej mierze wykorzystuje się wizerunki osób publicznych lub prywatnych, aby zniszczyć ich reputację w przestrzeni społecznej. Ostatni typ ma charakter dyskredytacyjny. Wykorzystuje się go w przestrzeni politycznej. Celem w tym przypadku jest osłabienie pozycji danej osoby, organizacji, niekiedy firm lub znanych marek samochodowych, odzieżowych albo biznesowych. Cyberprzestępcy w tym obszarze mogą manipulować cenami akcji, publikując na przykład fałszywy film przedstawiający prezesa, który ogłasza, że firma boryka się z problemami finansowymi lub innym kryzysem.

DEEPFAKE A WYBRANE PRZESTĘPSTWA W POLSKIM PRAWIE KARNYM

Wykorzystanie technologii *deepfake* może mieć znamiona przestępstwa.

W art. 190 a § 2 k.k. ustawodawca penalizuje określone działania, które zmierzają do przestępstwa: „Kto, podszywając się pod inną osobę, wykorzystuje jej wizerunek, inne jej dane osobowe lub inne dane, za pomocą których jest ona publicznie identyfikowana, w celu wyrządzenia jej szkody majątkowej lub osobistej. Podlega karze pozbawienia wolności od 6 miesięcy do 8 lat"[27]. Jest ono przestępstwem formalnym – bezskutkowym, które, jak słusznie zauważa się w doktrynie, zostaje dokonane w chwili, gdy sprawca przystąpił już do „robienia użytku" z danych osobowych lub wizerunku, nawet gdy szkody jeszcze nie wyrządził[28]. Przestępstwo to może zostać popełnione jedynie w zamiarze bezpośrednim kierunkowym, gdzie działanie sprawcy musi zostać podjęte w celu wyrządzenia pokrzywdzonemu konkretnej szkody materialnej lub osobistej. Tym samym do realizacji przedmiotowego przestępstwa w żadnym zakresie nie jest wystarczające działanie podjęte przez sprawcę w zamiarze ewentualnym – a więc aby sprawca jedynie godził się na wyrządzenie swoim działaniem szkody osobie, pod którą się podszywa, wykorzystując jej wizerunek lub dane osobowe[29]. Podobne stanowisko wyraził także Sąd Najwyższy w wyroku z 27.01.2017 r., wskazując w tezie przedmiotowego orzeczenia, iż „przestępstwo

[26] *Ibidem*, s. 92.

[27] Ustawa z dnia 6 czerwca 1997 roku – Kodeks karny (t.j. Dz. U. z 2022 r. poz. 1138 ze zm., dalej jako: k.k.).

[28] M. Filar, *Komentarz art. 190a §2*, [w:] *Kodeks Karny, Komentarz*, red. M. Filar, Warszawa 2016, s. 1175.

[29] M. Królikowski, A. Sakowicz, *Komentarz do art. 190a § 2*, [w:] *Kodeks karny. Część szczególna*, red. M. Królikowski, R. Zawłocki, t. 1, Warszawa 2017, s. 592.

określone w art. 190a § 2 k.k. może być popełnione wyłącznie w zamiarze bezpośrednim, tak więc dla realizacji jego znamion nie jest wystarczające, aby sprawca jedynie godził się na wyrządzenie swoim działaniem szkody osobie, pod którą się podszywa, wykorzystując jej wizerunek lub dane osobowe"[30]. Przestępstwo z art. 190a § 2 k.k. nie będzie możliwe do popełnienia przez zaniechanie, ponieważ użycie w nim znamienia czasownikowego „podszywać się" wymaga od sprawcy podjęcia określonego działania[31]. Stroną przedmiotową przestępstwa z art. 190a § 2 k.k. jest przy tym podszywanie się pod inną osobę poprzez wykorzystanie jej wizerunku lub innych jej danych osobowych w celu wyrządzenia jej szkody majątkowej lub osobistej[32].

Pojęcie podszywania się nie jest zdefiniowane w Kodeksie karnym. Należy więc przyjąć, że oznaczać ono będzie każdą formę wykorzystania cudzych danych osobowych lub wizerunku, która będzie stwarzała wrażenie, że użycie tych danych dokonane zostało przez ich faktycznego dysponenta, nie zaś przez osobę fikcyjnie podającą się za pokrzywdzonego. Jak słusznie zauważa przy tym A. Lach, zachowanie sprawcy może być bezpośrednio nakierowane na inną osobę – np. przez podanie jej danych osobowych lub też może ono oddziaływać na określone urządzenie informatyczne, weryfikujące dostęp do niego na podstawie podawanych danych[33].

Dobrem chronionym jest wizerunek osoby. Według definicji słownikowej pod pojęciem wizerunku kryją się następujące zwroty: czyjaś podobizna lub wyobrażenie czegoś[34]. Natomiast według E. Wojnickiej wizerunek to „dostrzegalne fizyczne cechy człowieka, tworzące jego wygląd i pozwalające na identyfikację osoby wśród innych ludzi"[35]. Zdaniem J. Sieńczyło-Chlabicz wizerunek jest to nie tylko obraz fizyczny, ale również głos, który jest wizerunkiem dźwięcznym. Wizerunek tworzą wszystkie elementy identyfikujące daną jednostkę jako konkretną osobę fizyczną"[36]. A. Ziobroń podkreśla, że czynność podmieniania twarzy aktorki z filmu pornograficznego na twarz innej osoby wypełnia znamiona przestępstwa. Argumentuje to tym, że twarz stanowi cechę człowieka, która jest dostrzegalna, fizyczna i tworzy jego wygląd[37]. Pogląd ten jest zbieżny z tym, który zaprezentowała E. Wojnicka[38]. Zazwyczaj

[30] Wyrok SN z 27.01.2017 r., V KK347/16, LEX.

[31] K. Sowirka, *Przestępstwo kradzieży tożsamości w polskim prawie karnym*, „Ius Novum" 2013, nr 1, s. 64–80.

[32] A. Grześkowiak, K. Wiak, *Kodeks karny. Komentarz*, Warszawa 2012, s. 865; M. Bojarski, *Prawo karne materialne. Część ogólna i szczególna*, Warszawa 2017, s. 608–609; A. Zoll, *Komentarz do art. 190a*, [w:] *Kodeks karny. Część szczególna*, t. 2, red. A. Zoll, W. Wróbel, Warszawa 2017. Lex.

[33] A. Lach, *Kradzież tożsamości*, „Prokuratura i Prawo" 2012, nr 3, s. 29–39.

[34] A. Grzegrzółka-Maciejewska, *Wizerunek*, [w:] *Uniwersalny słownik języka polskiego PWN*, t. T-Ż, red. S. Dubisz, Warszawa 2008, s. 457–458.

[35] E. Wojnicka, *Prawo do wizerunku w ustawodawstwie polskim*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace z Wynalazczości i Ochrony Własności Intelektualnej" 1990, nr 56, s. 107–108.

[36] J. Sieńczyło-Chlabicz, *Przedmiot, podmiot i charakter prawa do wizerunku*, „Przegląd Ustawodawstwa Gospodarczego" 2003, nr 8, s. 20.

[37] A. Ziobroń, *Deepfake a prawo karne. Uwagi* de lege lata *i* de lege ferenda *dotyczące fałszywej pornografii*, „Studenckie Prace Prawnicze, Administratywistyczne i Ekonomiczne" 2021, nr 57, s. 228.

[38] E. Wojnicka, *op. cit.*, s. 107.

rozpowszechnianie nagrania z wizerunkiem przyczynia się do wyrządzenia szkody osobistej, która przejawia się w uszczerbku na dobrach o charakterze niematerialnym, m.in. poczuciu dyskomfortu oraz cierpieniu psychicznym, niekiedy poniżeniu. Podsumowując, do tego typu przestępstwa dochodzi, gdy sprawca podszywa się pod cudzy wizerunek i przejmuje go oraz upublicznia w sieci.

Kolejnym przestępstwem, które można popełnić, wykorzystując technologię *deepfake*, jest rozpowszechnianie wizerunku nagiej osoby. W art. 191a § 1–2 k.k. ustawodawca stwierdza, że „ kto utrwala wizerunek nagiej osoby lub osoby w trakcie czynności seksualnej, używając w tym celu wobec niej przemocy, groźby bezprawnej lub podstępu, albo wizerunek nagiej osoby lub osoby w trakcie czynności seksualnej bez jej zgody rozpowszechnia, podlega karze pozbawienia wolności od 3 miesięcy do lat 5. Ściganie następuje na wniosek pokrzywdzonego”[39]. Przedmiotem ochrony jest intymność i prywatność człowieka. J. Kosonoga stwierdza, że przedmiot ochrony należy interpretować szeroko. Chodzi o swobodę jednostki w zakresie dysponowania własnym intymnym wizerunkiem, czyli wizerunkiem przedstawiającym osobę nago lub podczas czynności seksualnej. Chodzi zarówno o wolność od zastraszania lub stosowania podstępu w celu utrwalania takiego wizerunku, jak i o wolność do decydowania o jego rozpowszechnianiu. Jest to więc wolność decyzyjna człowieka, rozumiana jako wolność od niedozwolonego oddziaływania innych ludzi na swobodę i integralność procesu podejmowania decyzji, a następnie realizowania przez jednostkę określonego postępowania[40]. W podobnym duchu wypowiada się S. Hypś, który podkreśla, że przedmiot ochrony powinien być interpretowany ściśle i dotyczy dwóch sfer ludzkiego życia: nagości oraz wykonywania czynności seksualnej[41]. T. Bojarski podkreśla, że przedmiot ochrony powinien uwzględniać również wolność jednostki, w której zawarte są prawo do czynności seksualnej, upowszechniania przeżyć fizycznych i emocjonalnych[42]. J. Lachowski podobnie jak pozostali komentatorzy stwierdza, że do przestępstwa może dojść tylko w przypadku, gdy pokrzywdzony uczestniczy w czynnościach seksualnych, a upowszechnienie musi dokonać się bez jego zgody. M. Mozgawa zauważa, że aktorki występujące w filmach pornograficznych co do zasady wyrażają zgodę na utrwalenie wizerunków i ciał. Precyzuje, że czynność seksualna to obcowanie płciowe oraz inna czynność

[39] Art. 191a k.k.: „§ 1. Kto utrwala wizerunek nagiej osoby lub osoby w trakcie czynności seksualnej, używając w tym celu wobec niej przemocy, groźby bezprawnej lub podstępu, albo wizerunek nagiej osoby lub osoby w trakcie czynności seksualnej bez jej zgody rozpowszechnia, podlega karze pozbawienia wolności od 3 miesięcy do lat 5. § 2. Ściganie następuje na wniosek pokrzywdzonego”.

[40] J. Kosonoga, *Komentarz do art. 191 § 1*, [w:] *Kodeks karny. Komentarz*, red. R. Stefański, Warszawa 2021, Legalis.

[41] S. Hypś, *Komentarz do art. 191a*, [w:] *Kodeks karny. Komentarz*, red. A. Grześkowiak, K. Wiak, Warszawa 2012, s. 865.

[42] T. Bojarski, *Komentarz do art. 191 § 1*, [w:] *Kodeks karny. Komentarz*, red. T. Bojarski, Warszawa 2013, s. 508–509.

seksualna[43]. A. Zoll nagość rozumie jako prezentacja pośladków i piersi kobiecych oraz narządów płciowych męskich i damskich[44]. A. Ziobroń konstatuje powyższe poglądy i zauważa, że przy użyciu technologii *deepfake* nie można stwierdzić, że uczestnik czynności seksualnej jest osobą, której jedynie została doklejona twarz, ponieważ nie jest ona podmiotem czynności seksualnej ani nie jest naga. Podkreśla, że w tym kontekście komentowany przepis jest nieprzydatny[45].

Dzięki technologii *deepfake* może dojść również do dwóch niezależnych od siebie przestępstw – zniesławienia[46] (art. 212 k.k.) oraz znieważenia[47] (art. 216 k.k.). Dyspozycja zawarta w art. 212 chroni osoby, grupy osób, instytucję przed pomówieniem. Treść pomówienia musi się odnosić do postępowania lub okoliczności, które mogą poniżyć osobę w opinii publicznej lub narazić na utratę zaufania. Przedmiotem ochrony w tym przypadku jest cześć. Jej składowymi w wymiarze zewnętrznym są wartość jednostki w społeczeństwie, natomiast w wymiarze wewnętrznym godność osoby. O przestępstwie zniesławienia decyduje treść pomówienia. Natomiast wielu autorów zwraca uwagę na formę pomówienia. Według M. Mozgawy może się ono dokonać ustnie, przybrać formę pisemną, z wykorzystaniem druku, rysunku, jak również przy użyciu technicznych środków przekazu[48]. J. Sobczak podkreśla, że tego typu przestępstwa można dopuścić się za pomocą środków masowego przekazu[49]. Orzeczenie Sądu Najwyższego z 20.11.1933 r. dopuszcza popełnienie tego czynu

[43] M. Mozgawa, *Komentarz do art. 191 § 1*, [w:] *Kodeks karny, Komentarz aktualizowany*, red. M. Budyn-Kulik, Lex 2021.

[44] A. Zoll, *Komentarz do art. 191 § 1*, [w:] *Kodeks karny. Część szczególna*, t. 2, red. A. Zoll, W. Wróbel, Warszawa 2017. LEX.

[45] A. Ziobroń, *op. cit*., s. 230.

[46] Art. 212 k.k.: „§ 1. Kto pomawia inną osobę, grupę osób, instytucję, osobę prawną lub jednostkę organizacyjną niemającą 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. § 2. Jeżeli sprawca dopuszcza się czynu określonego w § 1 za pomocą środków masowego komunikowania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do roku. § 3. W razie skazania za przestępstwo określone w § 1 lub 2 sąd może orzec nawiązkę na rzecz pokrzywdzonego, Polskiego Czerwonego Krzyża albo na inny cel społeczny wskazany przez pokrzywdzonego. § 4. Ściganie przestępstwa określonego w § 1 lub 2 odbywa się z oskarżenia prywatnego".

[47] Art. 216 k.k.: „§ 1. 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. § 2. Kto znieważa inną osobę za pomocą środków masowego komunikowania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do roku. § 3. Jeżeli zniewagę wywołało wyzywające zachowanie się pokrzywdzonego albo jeżeli pokrzywdzony odpowiedział naruszeniem nietykalności cielesnej lub zniewagą wzajemną, sąd może odstąpić od wymierzenia kary. § 4. W razie skazania za przestępstwo określone w § 2 sąd może orzec nawiązkę na rzecz pokrzywdzonego, Polskiego Czerwonego Krzyża albo na inny cel społeczny wskazany przez pokrzywdzonego. § 5. Ściganie odbywa się z oskarżenia prywatnego".

[48] M. Mozgawa, *Komentarz do art. 212*, [w:] *Kodeks karny. Komentarz aktualizowany*, red. M. Budyn-Kulik, Lex 2021 [dostęp: 23.09.2022].

[49] J. Sobczak, *Komentarz do art. 212*, [w:] *Kodeks karny. Część szczególna*, t. 2, red. A. Zoll, W. Wróbel, Warszawa 2017. LEX (dostęp: 23.09.2022 r.)

w formie wizerunku lub karykatury[50]. J. Raglewski, jeden z ekspertów z zakresu prawa karnego, zaznacza, że zniesławienie można popełnić w każdej formie umożliwiającej realizację przekazu informacyjnego innej osobie. Jego pogląd rozwija A. Ziobroń. Oboje stwierdzają, że zniesławienie może przybrać postać przetworzonego filmu lub fałszywego zdjęcia. Na podstawie takiego materiału można zniesławić, sugerując na przykład prowadzenie niemoralnego trybu życia, braku przyzwoitości, czy zboczenia seksualne[51]. Rozpowszechnienie powyższych materiałów będzie typowym kwalifikowanym zniesławieniem za pomocą środków masowego przekazu, do którego zalicza się internet.

Zupełnie inny charakter ma czynność znieważenia, o którym mowa w art. 216 k.k. Polega na okazywaniu pogardy, uwłaczaniu jej czci lub inne obraźliwe zachowanie w stosunku do pokrzywdzonego. Zakres przedmiotowy przestępstwa określił Sąd Najwyższy w wyroku z 05.06.2012 r., który zauważył, że „o tym czy, zachowanie ma charakter znieważający decydują w społeczeństwie oceny i normy obyczajowe, a nie subiektywne przekonanie osoby znieważanej"[52]. J. Raglewski podkreśla, że podobnie jak w przypadku zniesławienia, znieważenie może dokonywać się nie tylko w formie ustnej, lecz może przybrać formę karykatury, filmu, obrazu, listu, opinii, wpisu w *social mediach*, w którym zostanie umieszczona pogarda dla innego człowieka. W przypadku znieważenia wystarczająca jest forma ekspresji, która będzie stanowiła formę wyrażenia pogardy. W tym przypadku nie jest wymagane, aby dokonywało się to za pomocą środków masowego przekazu. Zdaniem A. Ziobroń w przypadku przerobionego zdjęcia lub fałszywego filmu umieszczonego w sieci kluczowa przy znieważeniu staje się publiczność zniewagi - odbiorcy danego materiału[53]. Osoba, która za pomocą technologii *deepfake* stworzyła film lub obraz pornograficzny, powinna zostać pociągnięta do odpowiedzialności karnej z powyższych artykułów 212 i 216 k.k. Na podstawie fałszywych materiałów można odnieść wrażenie, że pokrzywdzony wystąpił w filmie pornograficznym, co ma charakter poniżający i może wiązać się to z faktem utraty zaufania. Zwłaszcza gdy materiał sugerowałby, że osoba przez swoją aktywność seksualną łamie normy obyczajowe lub promuje dewiacyjne postawy[54].

Warto podkreślić, że technologia *deepfake* może zostać użyta do popełnienia innych przestępstw, m.in. oszustwa (art. 286 k.k.), zmuszania (art. 191 § 1 k.k.) oraz stalkingu (art. 190a k.k.). Analiza wskazanych przestępstw wymaga stworzenia oddzielnego opracowania, w którym zostanie omówiony zakres przedmiotowy i podmiotowy wskazanych czynów zabronionych.

---

[50] Wyrok SN z 20.11.1933 r., III K 1037/33, „Orzeczenia Sądu Najwyższego" 1934, nr 4, poz. 5.

[51] A. Ziobroń, *op. cit*., s. 231.

[52] Wyrok SN z 05.06.2012 r., OSN 2012, nr 26, Lex nr 1231618.

[53] J. Raglewski, *Komentarz do art. 216*, [w:] *Kodeks karny. Część szczególna*, t. 2, red. A. Zoll, W. Wróbel, Warszawa 2017, s. 668-669; A. Ziobroń, *op. cit*., s. 232.

[54] A. Ziobroń, *op. cit*., s. 233-234.

NARZĘDZIA DO WYKRYWANIA *DEEPFAKE*

Technologia *deepfake* nie jest narzędziem idealnym. Za jej pomocą powstają fałszywe zdjęcia oraz filmy, ale eksperci z zakresu informatyki oraz cyfrowej fotografii potrafią dostrzec różnorakie modyfikacje. W Stanach Zjednoczonych Ameryki Północnej powstały technologie, które skutecznie zwalczają użytkowników korzystających z *deepfake*. Agencja Zaawansowanych Projektów Badawczych z obszaru obronności stworzyła technologię MediFor. Miała on na celu dokonać automatycznej oceny integralności zdjęć oraz wideo. Materiały przeznaczone do sprawdzenia były umieszczane na platformie wymiany pomiędzy użytkownikami końcowymi. Ponadto stworzono oprogramowania, które wykrywały różnorakie manipulacje. Zdaniem K. Świtalskiego twórcy oprogramowania analizowali opadanie włosów, ruch uszu, a nawet poszukiwali pulsu na czole. J. McGregor podkreśla, że walka z technologią *deepfake* nie jest łatwa. Niekiedy pojawiają się naturalne mrugnięcie okiem, które utrudniają wykrycie manipulacji. Z każdym rokiem powstają nowoczesne narzędzie do wykrywania fałszerstwa i manipulacji, a tymczasem użytkownicy technologii *deepfake* modyfikują zdjęcia twarzy coraz lepiej[55].

W 2017 r. amerykańska firma Al Foundation opracowała aplikację do weryfikacji autentyczności mediów – Reality Defender. Oprogramowanie łączyło dwie płaszczyzny: moderację człowieka i uczenie maszynowe. Chciano w ten sposób zidentyfikować złośliwe działania modyfikacji oraz fałszerstwa za pomocą technologii *deepfake*. Według O. i S. Wasiutów amerykańscy inżynierowie zachęcali do wysyłania fałszywych materiałów, potrzebnych do tworzenia spersonalizowanej sztucznej inteligencji, z której mogą korzystać wszyscy ludzie. W związku z tym stworzono Globalną Radę ds. Sztucznej Inteligencji, która stara się przewidywać negatywne skutki rozwoju sztucznej inteligencji i przeciwdziałać im[56].

W Australii postawiono na wypróbowaną metodę – zaczęto karać finansowo użytkowników technologii *deepfake*. W 2018 r. australijski parlament potępił technologię *deepfake*. Ustalono karę w wysokości 105 000 USD dla osób, które udostępniają intymne obrazy innych osób bez ich zgody, ze specjalnym przepisem obejmującym podróbki w ramach tej kategorii. Karę 525 000 USD ustalono dla firm i osób prywatnych, które rozpowszechniają fałszywe zdjęcia i filmy[57]. Wracając do głównego wątku tej części, należy podkreślić, że cyfrowe obrazy oraz filmy są łatwe do podrobienia. Zmusza to do opracowania kolejnych nowych programów oraz aplikacji, które będą skutecznie walczyć z tego typu zjawiskiem.

---

[55] K. Świtalski, *Deepfake. Zabawa w kotka i myszkę z politycznymi szantażami, propagandą i celebryckim porno*, https://antyweb.pl/walka-z-deepfake [dostęp: 21.09.2022].

[56] O. Wasiuta, S. Wasiuta, *Deepfake jako skomplikowana...*, s. 25.

[57] *Ibidem*, s. 23.

ZAKOŃCZENIE

Zjawisko *deepfake* jest coraz bardziej zaawansowane i dostępne, co może stanowić zagrożenie dla wielu obszarów społecznych. Wykorzystanie tego typu technologii może przybrać różne formy. Mogą z niej skorzystać za równo służby specjalne, jak i zwykły obywatel. *Deepfake* może stać się narzędziem propagandy i politycznego szantażu, jak zarówno niewinną zabawą, która ma na celu zrobić innej osobie żart. Bezapelacyjnie zjawisko *deepfake* jest nośnikiem przekazywania informacji fałszywej lub domniemanej. Szczegółowy sposób powstania tego informacji pokazuje jakościowa teoria informacji, która stanowi część cybernetyki. Należy stwierdzić, że *deepfake* jest nowym sposobem manipulacji, który zatacza nowe obszary w świecie online. Tego typu zjawisko znalazło podatny grunt, ponieważ informacje tracą swoją pierwotną funkcję. Wielu ekspertów zauważa, że w przestrzeni wirtualnej powstał nie tylko szum informacyjny, ale chaos. Doprowadzono do stanu, że duża liczba informacji i możliwość dzielenia się nimi stały się synonimem prawdy. Geneza, natura oraz typy *deepfake* pokazują, że żyjemy w nowej erze propagandy, która stała się skutecznym narzędziem destabilizacji i interferencji. Niektóre przestępstwa dzięki technologii stały się łatwiejsze do popełnienia. Bez doświadczonego eksperta niemożliwe jest odróżnienie fałszywego zdjęcia lub wideo, które narusza cześć, godność oraz wolność człowieka. W związku z tym wiele instytucji powinno połączyć siły i stworzyć skuteczne narzędzia do walki z technologią *deepfake*.

BIBLIOGRAFIA

AKTY NORMATYWNE:

Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny (Dz. U. z 2020 r. poz. 1444 ze zm.).

LITERATURA:

Bojarski M., *Prawo karne materialne. Część ogólna i szczególna*, Warszawa 2017.
Dąbrowska I., *Deepfake – nowy wymiar internetowej manipulacji*, „Zarządzanie Mediami" 2020, nr 2.
Grzegrzółka-Maciejewska A., *Wizerunek*, [w:] *Uniwersalny słownik języka polskiego PWN*, t. T-Ż, red. Dubisz Stanisław, Warszawa 2008.
Kodeks karny część szczególna, red. M. Królikowski, R. Zawłocki, t. 1, Warszawa 2017.
Kodeks karny, Komentarz aktualizowany, red. M. Budyn-Kulik, Lex 2021.
Kodeks karny. Część szczególna, t. 2, red. A. Zoll, W. Wróbel, Warszawa 2017.
Kodeks karny. Komentarz, red. T. Bojarski, Warszawa 2013.
Kodeks Karny. Komentarz, red. M. Filar, Warszawa 2016.
Kodeks karny. Komentarz, red. A. Grześkowiak, K. Wiak, Warszawa 2012.
Kodeks karny. Komentarz, red. R. Stefański, Warszawa 2021, Legalis.
Kowalski A., *Kontra. Sztuka walki z wywiadem przeciwnika*, Łomianki 2021.
Lach A., *Kradzież tożsamości*, „Prokuratura i Prawo" 2012, nr 3.
Mazur M., *Jakościowa teoria informacji*, Warszawa 1968.
Muniak P., Kulesza W., *Sztuka dezinformacji*, „Newsweek" 2022, nr 30.
Münkler H., *Wojny naszych czasów*, Kraków 2004.
Piasecki B., *Kontrwywiad atak i obrona*, Łomianki 2021.

Sieńczyło-Chlabicz J., *Przedmiot, podmiot i charakter prawa do wizerunku*, „Przegląd Ustawodawstwa Gospodarczego" 2003, nr 8.
Sokała W., *Jak przegrać wygraną wojnę*, „Dziennik Gazeta Prawna" 2022, nr 156.
Sowirka K., *Przestępstwo kradzieży tożsamości w polskim prawie karnym*, „Ius Novum" 2013, nr 1.
Wasiuta O., Wasiuta S., *Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate" 2019, nr 9.
Wasiuta O., Wasiuta S., *FakeApp jako nowe zagrożenie bezpieczeństwa politycznego i informacyjnego*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate" 2019, nr 9.
Wojnicka E., *Prawo do wizerunku w ustawodawstwie polskim*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace z Wynalazczości i Ochrony Własności Intelektualnej" 1990, nr 56.
Young N., *Deepfake Technology: Complete Guide to Deepfakes, Politics and Social Media*, New York 2019.
Ziobroń A., *Deepfake a prawo karne. Uwagi de lege lata i de lege ferenda dotyczące fałszywej pornografii*, „Studenckie Prace Prawnicze, Administratywistyczne i Ekonomiczne" 2021, nr 57.

ORZECZNICTWO:

Wyrok SN z 27.01.2017 r., V KK347/16, LEX.
Wyrok SN z 05.06.2012 r., OSN 2012, nr 26, Lex nr 1231618.
Wyrok SN z 20.11.1933 r., III K 1037/33, OSN 1934, nr 4, poz. 5.

INNE ŹRÓDŁA:

https://www.legalniewsieci.pl/aktualnosci/cala-prawda-o-fake-news-czyli-jak-rozpoznac-falszywe--wiadomosci
https://sjp.pwn.pl/
https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth
https://antyweb.pl/walka-z-deepfake
https://www.computerweekly.com/blog/Downtime/Deepfake-it-until-you-makeit?_ga=2.245290114.623414774.1663061863852034460.1663061863&_gl=1*1b2n5ic*_ga*ODUyMDM0NDYwLjE2NjMwNjE4NjM.*_ga_TQKE4GS5P9*MTY2MzA2MTg2Mi4xLjEuMTY2MzA2MjU5OC4wLjAuMA
https://www.techtarget.com/searchenterpriseai/news/https://www.techtarget.com/searchenterpriseai/news/252488582/Microsoft-deepfake-software-combats-election-propaganda252488582/; J. Burke, *Why Itleaders need to be aware of deepfake security risks*, w: https://www.techtarget.com/search/query?q=deepfake&type=article&pageNo=1&sortField=https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6131996/pdf/JFMPC-7-828.pdf

Deepfake as a Tool of Disseminating False and Implied Information. Legal, Criminal and Cyber Analysis

SUMMARY

Since 2017, the phenomenon of deepfake has emerged online. It has very quickly been applied to various areas of human life. The developing digital world, social media, entertainment and other areas have brought the exchange of information between people online. There is a lack of monographs and articles in the Polish and foreign-language literature that would reliably describe this type of threat in the virtual world. The essence of the deepfake phenomenon was to transmit false or assumed information, which is done by means of photographs or videos. In most cases,

it is not possible to verify the authenticity of the material. Deepfake technology has specific types that make users of the web and many applications unaware of this type of threat. The article describes the deepfake phenomenon as a tool for conveying false and implicit information. The analysis of the phenomenon was carried out in two dimensions: legal-criminal and cyber-criminal. A number of research methods have been used, including dogmatic-legal, historical, comparative, philological and historical. Amongst numerous conclusions, the most important one indicates that deepfake is a new form of manipulation and disinformation. This type of technology makes it possible to commit many crimes, which are difficult to prove to the perpetrator during criminal proceedings. The cooperation of many state institutions, non-governmental organisations, software or application manufacturers will help to win the fight against this type of phenomenon.

Keywords: deepfake, manipulation, information, disinformation, crime, cybernetics