FOLIA 289

Annales Universitatis Paedagogicae Cracoviensis

Studia de Securitate 9(3) (2019)
ISSN 2657-8549
DOI 10.24917/26578549.9.3.2

Olga Wasiuta
ORCID ID 0000-0003-0481-1567
Uniwersytet Pedagogiczny w Krakowie

Sergiusz Wasiuta
ORCID ID 0000-0003-3402-963X
Uniwersytet Pedagogiczny w Krakowie

Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość

Wprowadzenie

Wzrost znaczenia informacji na przełomie XX i XXI wieku doprowadził do tego, że bezpieczeństwo informacyjne wraz z innymi jego rodzajami stały się priorytetową dziedziną bezpieczeństwa narodowego. Konkurencja w zdobywaniu i posiadaniu informacji przyczyniła się do konieczności jej ochrony oraz doprowadziła do swoistej walki – zdobywania, zakłócania i obrony informacji. Dodatkowo wraz ze wzrostem przepływu informacji w cyberprzestrzeni problematyka bezpieczeństwa informacyjnego nabrała wymiaru globalnego. Agencje UE oraz poszczególne sektory organizacji międzynarodowych zajmują się różnymi aspektami bezpieczeństwa informacji. Państwa z kolei prowadzą własną politykę bezpieczeństwa informacyjnego, ukierunkowaną na ochronę istniejących systemów informacyjnych, zapewnienie bezpieczeństwa infrastruktury krytycznej państwa oraz podmiotów prywatnych narażonych na zagrożenia m.in. cyberatakiem i rozprzestrzenianiem szkodliwego oprogramowania.

Współczesny świat, zdominowany przez Internet i media społecznościowe jako wiarygodne źródła informacji o nim samym, to wyjątkowo podstępna przestrzeń[1]. Odbiorca mierzy się bowiem z natłokiem informacji i stale dokonuje selekcji, gdyż narażony jest na ryzyko odbioru nieprawdziwych wiadomości. Fake news jako względnie nowe zjawisko[2], rozprzestrzenia się w Internecie z zawrotną prędkością i co niezmiernie ważne ma już swojego następcę – deepfake wideo, które może stać się nową strategią wykorzystywaną w marketingu politycznym.

Manipulacja informacją nie jest zjawiskiem nowym, nabrała jednak zupełnie nowego wymiaru, ze względu na niespotykane dotąd możliwości Internetu i sieci

1 O. Wasiuta, *Sieci społecznościowe jako nowe narzędzia prowadzenia wojen informacyjnych we współczesnym świecie*, [w:] *Refleksje o przeszłości, spojrzenie na współczesność: monografia poświęcona Profesorowi Sergiuszowi Wasiucie z okazji 60-letniego Jubileuszu i 35-lecia pracy zawodowej,* red. O. Wasiuta. Drukarnia Styl Anna Dura, Kraków 2018, s. 184–209.

2 R. Klepka, *Fake news*, [w:] *Leksykon bezpieczeństwa,* red. O. Wasiuta, R. Klepka, R. Kopeć, Wydawnictwo Libron, Kraków 2018, s. 299–304.

społecznościowych w zakresie rozpowszechniania informacji i tworzenia z nich wirusów, a także kryzysu zaufania, jakiego obecnie doświadczają demokracje.

Ludzkość ma długą historię korzystania z obrazowych reprezentacji informacyjnych jako sposobu upowszechniania i przyswajania wiedzy, a wśród wszystkich obrazowych reprezentacji nic nie jest bardziej wiarygodne niż zdjęcia i filmy. Jako bezpośrednie wyrażenia rzeczywistości, przedstawiają one obraz świata, który ma wpływ na ludzką opinię w taki sam sposób, jak robią to bodźce ze świata rzeczywistego.

Zdjęcia i filmy są traktowane jako najdokładniejsze sposoby dokumentacji informacji wizualnych, zarówno obrazów ludzi i przedmiotów, jak i wydarzeń. Tworząc bezstronny i powszechnie dostępny zapis wydarzenia, są często wykorzystywane w dziedzinach takich jak dziennikarstwo, polityka, spory cywilne, procesy karne, planowanie obrony, nadzór i operacje wywiadowcze. Najlepszym dowodem na ich znaczenie jest fakt, iż w kryminalistyce były faktycznie używane niemal natychmiast po wynalezieniu fotografii.

Tworzenie filmów Deepfake

W drugiej dekadzie XXI wieku pojawiły się możliwości zakłamania rzeczywistości, chociażby poprzez nowe programydo obróbki wideo, które radykalnie zmieniły możliwości manipulacji obrazem. Jednym z takich programów do generowania komputerowo treści wideo jest Deepfakes – technika syntezy ludzkiego obrazu oparta na sztucznej inteligencji. Służy do łączenia i nakładania na wyjściowe istniejących obrazów i filmów. Deepfake to zespół algorytmów składający się na program komputerowy doskonale radzący sobie z wizualnymi przeróbkami. Technologia wykorzystuje sztuczną inteligencję używaną do tworzenia lub modyfikowania odwzorowanej twarzy, do tworzenia ultra-realistycznych fałszywych filmów, w których ludzie mówią i robią rzeczy, w rzeczywistości nie mające miejsca[3]. Oprogramowanie do edycji zdjęć, takie jak Photoshop, od dawna było używane do fałszowania obrazów, jednak do niedawna trudno było zmienić treści wideo w jakikolwiek znaczący sposób. Podważenie wiarygodności materiału wideo zmieniło postrzeganie informacji.

Jest to metoda tworzenia obrazów ludzkich w oparciu o sztuczną inteligencję (*Artificial Intelligence* – AI). W porównaniu ze starymi technikami Photoshopa używanymi w celu stworzenia fałszywych dowodów, kwalifikowałoby się to jako „videoshop 3.0". Problem wykrywania podróbek video prawdopodobnie nie zostanie rozwiązany technicznie, co w znaczącym stopniu może odmienić podejście do tego, co przekazują media. Konieczne w tej sytuacji staje się nawiązanie relacji zaufania z dziennikarzami, domami prasowymi, wydawcami lub innymi źródłami informacji[4]. Tym bardziej, że tylko w pierwszym roku pracy serwisu Google Foto

3 R. Heartfield, G. Loukas. *Protection Against Semantic Social Engineering Attacks*, *Protection Against Semantic Social Engineering Attacks,* „Versatile Cybersecurity. Advances in Information Security", Edition: 72, Chapter: 4, Publisher: Springer, Cham, 2018, p. 99–140.

4 T. Bezmalinovic, *Wenn Merkel plötzlich Trumps Gesicht trägt: die gefährliche Manipulation von Bildern und Videos*, 03.02.2018, https://www.aargauerzeitung.ch/leben/digital/

(2015–2016) pojawiło się tam 200 milionów użytkowników, a wśród zdjęć było 24 miliardy selfie[5].

Termin Deepfake związany jest z użytkownikiem o nazwie „DeepFakes", który w grudniu 2017 roku 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[6]. Filmy są tworzone przez załadowanie złożonego zestawu instrukcji do komputera wraz z dużą ilością 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 z tą osobą. Bardzo często oszuści tworzący materiały typu „deepfake" wykorzystują autentyczne nagrania, które łączą ze sztucznie wygenerowanym obrazem[7].

W ciągu kilku tygodni od pojawienia się sfałszowanych filmów Reddit usunął stronę z podróbkami i wszystkimi podobnymi treściami, powołując się na zasady dotyczące mimowolnej pornografii. Mniej więcej w tym samym czasie strony serwisów społecznościowych Twitter i Discord, a także strony pornograficzne Pornhub, ogłosiły podobne zakazy dotyczące treści z deepfakes[8].

Słowo „deepfake" oznacza wykorzystanie algorytmów uczenia maszynowego i technologii mapowania twarzy do cyfrowej manipulacji głosami, ciałami i twarzami ludzi. Technologia ta rozwija się w tak szybkim tempie, że coraz trudniej jest powiedzieć, co jest fałszywe. Z biegiem czasu, bez zastosowania odpowiedniego sprzętu, filmy Deepfake staną się nie do odróżnienia od prawdziwych zdjęć czy filmów i mogą być wykorzystywane również do tworzenia fałszywych wiadomości i złośliwych oszustw[9].

To, co odróżnia Deepfakes od innych technik służących do manipulacji materiałem wideo to potencjał umożliwiający uzyskanie realistycznych rezultatów. Z wystarczającą ilością obrazów danych aktorów i wystarczającą ilością czasu na

wenn-merkel-ploetzlich-trumps-gesicht-traegt-die-gefaehrliche-manipulation-von-bildern-und-videos-132155720 [dostęp: 16.02.2019].

5 Т. Ястремська, *Deepfake: як жити у світі, де підробку не відрізнити від реальності*. Частина І, 26 червня, 2018, https://kfund-media.com/deepfake-yak-zhyty-u-sviti-de-pidrobku-ne-vidriznyty-vid-realnosti-chastyna-i/ [dostęp: 18.02.2019].

6 M. Brundage et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Executive Summary February 2018, p. 49; *Scarlett Johansson na tropie deepfake'ów*, 13.01.2019, https://niezalezna.pl/253924-scarlett-johansson-na-tropie-deepfake39ow [dostęp: 21.02.2019].

7 *Fake News: Read All About It*, ed. The New York Times Editorial Staff. The Rosen Publishing Group, Inc, 2018, p.34-35; *A Reddit User Starts 'Deepfake'*. 27.10.2017, https://www.eyerys.com/articles/timeline/reddit-user-starts-deepfake?page=8 [dostęp: 16.02.2019].

8 J.Bailey, *The deepest fake: how new tech will test our belief in what we see*, 04.05.2018, https://www.smh.com.au/technology/the-deepest-fake-how-new-tech-will-test-our-belief-in-what-we-see-20180423-p4zb4w.html [dostęp: 16.02.2019].

9 M.Giles, *Five emerging cyber-threats to worry about in 2019*. January 4, 2019, https://www.technologyreview.com/s/612713/five-emerging-cyber-threats-2019/ [dostęp: 15.02.2019].

szkolenie komputerowe, powstające filmy mogą być niezwykle przekonujące, choć bardziej wnikliwy odbiorca, zwracając uwagę na brak sygnałów fizjologicznych właściwych człowiekowi: oddychania, mrugania oczu, pulsu, odróżni je od rzeczywistości.

Generowanie wideo

Nowa technologia pozwala każdemu stworzyć materiał wideo, w którym pojawiają się znane postaci, np. prezydent USA Donald Trump czy wysocy rangą dyplomaci, wypowiadający się na kontrowersyjne tematy w sposób podburzający opinię publiczną. Deepfakes zostały wykorzystane wielokrotnie do fałszywego przedstawienia znanych polityków na portalach wideo lub czatach. W eksperymencie badawczym z 2016 r. technika ta została użyta przy wykorzystaniu twarzy światowych liderów, takich jak George W. Bush, Władimir Putin i Barack Obama. W późniejszym czasie stosowano tę technikę właśnie przy wykorzystaniu wizerunku Obamy, co wzbudziło duże obawy ze względu na szybkie tempo rozprzestrzenienia się fałszywych wideo filmów. Na YouTube krąży Deepfake, w którym twarz argentyńskiego prezydenta Mauricio Macri została zastąpiona twarzą Adolfa Hitlera, a w drugim filmie Angela Merkel wygłosiła przemówienie ze ściśniętą twarzą Donalda Trumpa[10]. W lipcu 2017 roku świat obiegł filmik, w którym Barack Obama obrażał Donalda Trumpa. Okazało się, że byłego prezydenta USA wygenerowano w całości w aplikacji FakeApp, a głosu użyczył mu komik Jordan Peele. Akcja miała na celu zwrócenie uwagi na problem fake newsów[11]. W kwietniu 2018 r. Jordan Peele i Jonah Peretti stworzyli podróbkę, używając wizerunku Obamy dla zaalarmowania opinii publicznej o zagrożeniach związanych z podróbkami[12]. Powyższe przykłady jasno pokazują, że są to satyry polityczne. Ale co, jeśli manipulacja nie jest już rozpoznawalna jako taka?

Podróbki takie mogą być zastosowane w wojnie informacyjnej. Zagraniczny przeciwnik, chcący wpłynąć na wybory, może zamieścić w sieci film, oczerniający tego czy innego kandydata. Przekonująco zredagowane wideo może zmylić oficerów wojskowych w terenie. Niepewność ta może również zostać wykorzystana do podważenia dziennikarskiej wiarygodności; podróbka może być dzisiejszymi „fałszywymi wiadomościami".

Być może część tych „kreacji Frankensteina" jest łatwa do wykonania, zwłaszcza gdy oprogramowanie do określonej aplikacji – takie jak pornografia – jest publicznie dostępne. Można po prostu podłączyć wystarczającą ilość zdjęć lub nagrań

10 T. Bezmalinovic, *Wenn Merkel plötzlich*...

11 P.A. Kopciak, *Fake Algorithms: Your face in my video. The influence of elaborate fake videos on our perception and socjety. History, theories and current developments in the media landscape, regarding aesthetics and socjety,* University of Applied Sciences St. Pölten Master course „Digital Mediatechnologies", Vienna 2018, s. 10.

12 *Jordan Peele's simulated Obama PSA is a double-edged warning against fake news*, Apr 18, 2018, https://www.vox.com/2018/4/18/17252410/jordan-peele-obama-deepfake-buzzfeed; S. Suwajanakorn, S.M. Seitz, I. Kemelmacher-Shlizerman, *Synthesizing Obama: Learning Lip Sync from Audio,* ACM Transactions on Graphics (SIGGRAPH), 4, July 2017, https://doi.org/10.1145/3072959.3073640. [dostęp: 15.02.2019].

do wcześniej napisanego kodu i stworzyć realistyczne kłamstwo na temat wybranej osoby[13].

Deepfakes z natury są trudne do wykrycia. Technologia ich tworzenia używa tych samych algorytmów, które odróżniają fałszywe treści od rzeczywistych – więc jest dobrze przygotowana, aby uczynić te treści bardziej przekonującymi.

Deepfakes zostały potępione w australijskim parlamencie. W lutym 2018 roku Senat uchwalił projekt ustawy, nakładający kary w wysokości do 105 000 USD dla osób, które udostępniają intymne obrazy innych osób bez ich zgody, ze specjalnym przepisem obejmującym podróbki w ramach tej kategorii. Ustawa obejmuje także kary w wysokości do 525 000 USD dla firm winnych rozpowszechniania fałszywych zdjęć i filmów[14].

Masowa dostępność oprogramowania Deepfake to problem trudny do zignorowania. Dzięki tej technologii coraz bardziej trudniejsze staje się odfiltrowywanie prawdy od kłamstw. Technologia Deepfake jest już szeroko stosowana w fałszywych filmach pornograficznych i komediowych. A szybki postęp technologii za jakiś czas może prowadzić do poważnych konsekwencji. Realistyczne filmy Deepfake wykorzystywane są często przy próbach szantażu, linkach phishingowych[15] i oszustwach wymuszających. Filmy Deepfake dostarczają przestępcom narzędzia do tworzenia (przy minimalnym nakładzie pracy) realistycznych, trudnych do wykrycia (przynajmniej bez głębokiej analizy sądowej) nagrań wideo, które mogą podszywać się pod kogoś i oszukiwać każdego, w tym wymiar sprawiedliwości. Mogłyby one zostać wykorzystane do wymuszeń, wplątania niewinnych ludzi w zbrodnie, a w postępowaniu cywilnym te sfałszowane filmy mogą być wykorzystywane do przeprowadzania wszelkiego rodzaju roszczeń. Takie wideo otwiera drzwi do świata, w którym krążą nie tylko „fakty alternatywne", ale także cała alternatywna rzeczywistość[16].

O znaczeniu narzędzia deepfake świadczy wykorzystanie go w Chinach. W listopadzie 2018 r. chińska państwowa telewizja agencji informacyjnej Xinhua stworzyła wygenerowanego komputerowo prezentera, który poprowadził wieczorowe wiadomości (jego sylwetka wzorowana była na pracowniku agencji, Zhangu Zhao)[17]. *Artificial Intelligence Anchor* – taką nazwę nadała swojej kreacji agencja informacyjna. System sztucznej inteligencji odpowiada w nim za przeniesienia

13 W. Board, *A reason to despair about the digital future: Deepfakes*, 06.01.2019, https://www.washingtonpost.com/opinions/a-reason-to-despair-about-the-digital-future-deepfakes/2019/01/06/7c5e82ea-0ed2-11e9-831f-3aa2c2be4cbd_story.html?utm_term=.8f013c9f7e16 [dostęp 21.02.2019].

14 J. Bailey, *The deepest fake*…

15 Phishing to typ oszustwa internetowego, w którym podstępem wyłudza się od użytkownika jego osobiste dane. Phishing obejmuje kradzież haseł, numerów kart kredytowych, danych kont bankowych i innych poufnych informacji.

16 P. Gensin, *Deepfakes: Auf dem Weg in eine alternative Realität?* 22.02.2018, http://faktenfinder.tagesschau.de/hintergrund/deep-fakes-101.html [dostęp: 16.02.2019].

17 A. Cyganek, *Pomysłowość Chińczyków nie zna granic. Stworzyli prezentera za pośrednictwem DeepFake,* [dostęp: 09.11.2018], https://www.instalki.pl/aktualnosci/internet/33377-pomyslowosc-chinczykow-nie-zna-granic-stworzyli-prezentera-za-posrednictwem-deepfake.html; *Chiny: Serwisy informacyjne poprowadzą prezenterzy wygenerowani komputerowo*, [dostęp: 9.11.2018], https://www.cdaction.pl/news-55063/chiny-serwisy-informacyjne-poprowadza-prezenterzy-wygenerowani-komputerowo-wideo.html [dostęp: 20.02.2019]

zadanego tekstu na wygląd i zachowanie modelu prezentera – generowanie jego ruchów oraz synchronizacji syntezowanej mowy z ruchami ust[18]. Xinhua planuje stworzenie mediów, które będą prowadzone na okrągło przez komputerowych prezenterów, mogących jednocześnie nagrywać kilka różnych komunikatów i to w kilku językach. Treści, które ma przedstawić cyfrowy prezenter, wprowadzane są do pamięci Deepfake, a ruch jego warg synchronizowany jest ze słowami wypowiadanymi przez syntezator mowy. Stało się to, czego wielu się obawiało od dawna. Pierwszy raz w historii telewizji wiadomości ze świata zapowiada prezenter, który został sztucznie wygenerowany z pomocą technologii Deepfake[19].

Na początku stycznia 2019 r. na stronie *Technology review* pojawił się artykuł, analizujący największe cyberzagrożenia 2019 roku. Wśród nich na **pierwszym miejscu** – wykorzystywanie fałszywego wideo i audio generowanych przez sztuczną inteligencję[20]. Dzięki postępowi w rozwoju sztucznej inteligencji można teraz tworzyć fałszywe wiadomości, które są niezwykle trudne do odróżnienia od prawdziwych. Te „podróbki" mogą być dobrodziejstwem dla hakerów na kilka sposobów. Wiadomości e-mail generowane przez sztuczną inteligencję, mające na celu skłonienie ludzi do przekazywania haseł i innych poufnych danych, okazały się bardziej skuteczne, niż te generowane przez ludzi. Teraz hakerzy będą mogli dodawać bardzo realistyczne fałszywe wideo i audio, aby wzmocnić instrukcje w wiadomościach phishingowych[21] lub jako samodzielną taktykę[22].

Cyberprzestępcy mogą również wykorzystać tę technologię do manipulowania cenami akcji, publikując na przykład fałszywy film przedstawiający prezesa, który ogłosił, że firma boryka się z problemem finansowania lub innym kryzysem. Istnieje również niebezpieczeństwo, że podróbki takie mogą być wykorzystywane do rozpowszechniania fałszywych wiadomości w wyborach i podsycania geopolitycznych napięć[23].

---

18 M. Tomaszkiewicz, *Chińczycy stworzyli sztucznego prezentera wiadomości*, 09.11.2018, https://www.antyradio.pl/Technologia/Duperele/Chinczycy-stworzyli-sztucznego-prezentera-wiadomosci-27057 [dostęp: 20.02.2019].

19 A. Cyganek, *Pomysłowość Chińczyków nie zna granic*…; *Chiny: Serwisy informacyjne*…; *Prezenterzy, którzy nie istnieją? Agencja Xinhua wykorzystuje technologię DeepFake*, 10.11.2018, https://pl.aleteia.org/2018/11/10/prezenterzy-ktorzy-nie-istnieja-agencja-xinhua-wykorzystuje-technologie-deepfake/ [dostęp: 20.02.2019].

20 M.Giles, *Five emerging cyber-threats*...

21 Wiadomości phishingowe przybierają zazwyczaj formę fałszywych powiadomień z banków, komunikatów od dostawców systemów e-płatności i innych poważanych organizacji. Wiadomość taka zawsze próbuje zachęcić odbiorcę, z takiego lub innego powodu, aby w trybie pilnym wprowadził czy zaktualizował swoje poufne informacje, bo w przeciwnym razie dotknie go utrata krytycznych danych, awaria systemu lub inne nieszczęście (A. Suchorzewska, *Ochrona prawna systemów informatycznych wobec zagrożenia cyberterroryzmem,* Wolters Kluwer, Warszawa 2010, s. 89; E.M. Guzik-Makaruk, E.W. Pływaczewski, *Współczesne oblicza bezpieczeństwa*, Wydawnictwo Temida 2, Białystok 2015, s. 119).

22 M.Giles, *Five emerging cyber-threats*…

23 M.Giles, *Five emerging cyber-threats*…

Technologie do wykrywania Deepfake

Technologia tworzenia zmanipulowanych wideo nie jest jeszcze doskonała i wytrawne oko dostrzeże modyfikacje. Mechanizm konstruowania Deepfake wideo jednak cały czas się rozwija i już wkrótce odbiorca być może nie zauważy, że materiał filmowy jest fałszywy. Technologia tworzenia Deepfake wideo rozwinęła się bardzo szybko w ciągu ostatniego roku. By móc się przed nią obronić w USA aktywnie rozwija się technologie, które mogą wykrywać filmy Deepfakes. Na przykład Agencja Zaawansowanych Projektów Badawczych w Obszarze Obronności USA (*Defense Advanced Research Projects Agency* – DARPA)[24] w 2016 r. rozpoczęła projekt MediFor (ang. *Media Forensics*), którego celem jest opracowanie technologii do automatycznej oceny integralności zdjęć lub filmów i zintegrowanie jej w platformę wymiany materiałów pomiędzy użytkownikami końcowymi. Oprócz tego opracowuje inteligentne oprogramowanie wykrywające filmowe manipulacje, w tym analizujące m.in. opadanie włosów, ruch uszu, odbicie światła w oczach, a jeden z inżynierów opracował nawet test, który poszukuje pulsu na czole[25]. Niestety w najnowszych fałszywkach nie brakuje nawet naturalnych mrugnięć, których brak kiedyś pozwalał stosunkowo szybko wykryć takie materiały[26]. W niedalekiej przyszłości platforma MediFor automatycznie wykryje manipulacje, dostarczy szczegółowych informacji o tym, jak te manipulacje zostały wykonane, co ułatwi decyzje dotyczące użycia jakichkolwiek wątpliwych zdjęć lub filmów[27].

Podobnie z wideo manipulacjami walczy od 2017 roku amerykańska firma *AI Foundation*, opracowując oprogramowanie do weryfikacji autentyczności mediów. Pierwszy produkt *AI Foundation* o nazwie Reality Defender łączy w sobie moderację człowieka i uczenie maszynowe, aby zidentyfikować działania złośliwe, między innymi Deepfakes. Naukowcy zapraszają użytkowników do wysyłania im fałszywych materiałów, potrzebnych do tworzenia spersonalizowanej sztucznej inteligencji, z której mogą korzystać wszyscy ludzie. W tym celu Fundacja AI stworzyła własną Globalną Radę ds. Sztucznej Inteligencji, która stara się przewidywać i przeciwdziałać negatywnym skutkom rozwoju sztucznej inteligencji.

Nasze uzależnienie od cyfrowych treści multimedialnych wynika z możliwości dostarczania informacji z pierwszej ręki o zdarzeniu i naszej skłonności do wierzenia w to, co widzimy. Właśnie ze względu na tę zależność musimy upewnić się, że dana zawartość cyfrowa jest rzeczywista. Cyfrowe obrazy i wideo są niezwykle łatwe do manipulowania, a możliwość manipulacji jest szczególnie niepokojąca w scenariuszach, w których ta cyfrowa treść jest traktowana jako dowód i podstawa do podejmowania decyzji i osądów, mających długotrwałe następstwa. Zmusza to do

24 *Defense Advanced Research Projects Agency: Overview and Issues for Congress Updated,* July 24, 2018. Congressional Research Service, https://crsreports.congress.gov; *Creating Breakthrough Technologies And Capabilitiesfor National Security*, https://www.darpa.mil/ [dostęp: 16.02.2019].

25 K. Świtalski, *Deepfake: zabawa w kotka i myszkę z politycznymi szantażami, propagandą i celebryckim porno,* 31.12.2018, https://antyweb.pl/walka-z-deepfake/ [dostęp: 21.02.2019].

26 *Scarlett Johansson na tropie deepfake'ów*, 13.01.2019, https://niezalezna.pl/253924-scarlett-johansson-na-tropie-deepfake39ow [dostęp: 19.02.2019].

27 M. Turek, *Media Forensics (MediFor)*, https://www.darpa.mil/program/media-forensics [dostęp: 16.02.2019].

opracowania procedur dochodzeniowych, które są w stanie ustalić autentyczność i wiarygodność danej treści cyfrowej.

Technologia Deepfake staje się coraz bardziej zaawansowana i dostępna, co może stanowić zagrożenie dla dyskusji publicznej i bezpieczeństwa narodowego każdego państwa. Deepfake może być stosowany przez służby wywiadowcze obcych państw. Senator Marco Rubio, republikanin z Florydy, powiedział, że wierzy, że zmanipulowane wideo będą wykorzystywane w „kolejnej fali ataków przeciwko Ameryce i zachodnim demokracjom"[28]. Z uwagi na wyjątkową szybkość rozwoju Deepfakes wydaje się prawdopodobne, że w najbliższej przyszłości staniemy się świadkami oszustw przy użyciu tej metody. Być może w pewnym momencie do użytku zostanie wprowadzone specjalistyczne oprogramowanie przeciwdziałające oszustwom wideo, i zaczniemy z niego korzystać w taki sam sposób, w jaki przyzwyczailiśmy się korzystać z ochrony antyspamowej i ochrony przed złośliwym oprogramowaniem.

Deepfakes staje się narzędziem propagandy i politycznego szantażu. Wielu martwi się, że Deepfakes mogą w najbliższej przyszłości zniszczyć światowy klimat polityczny poprzez rozpowszechnianie realistycznie sfałszowanych filmów. Wszyscy wiemy, że wiadomości zyskują nową głębię, jeśli całość lub część danego elementu „news" jest realistycznym klipem wideo. Profesorowie prawa Robert Chesney[29] i Danielle Citron[30] rozważają szereg scenariuszy, w których technologia Deepfakes może okazać się katastrofalną, kiedy będzie wykorzystywana w wiadomościach jako np. fałszywy film ukazujący zbliżanie się pocisku rakietowego do Los Angeles lub nową pandemię w Nowym Jorku, doprowadzając tym samym do paniki. Takie zastosowania sfałszowanych filmów wideo mogą być niezwykle szkodliwe, kiedy będą rozpowszechniane w mediach społecznościowych. Humorystyczny efekt wkrótce ustąpi miejsce bardziej realistycznemu Deepfake z udziałem postaci politycznych i celebrytów mówiących humorystycznie, satyrycznie, fałszywie[31].

Wnioski

Podsumowując, chcemy podkreślić, że XXI wiek, wiek informacji, to czas pojawienia się nowego zjawiska w arsenale zagrożeń dla ludzkości, a mianowicie: bałaganu informacyjnego. Corocznie coraz więcej wydarzeń sygnalizuje, że przestrzeń informacyjna jest wypełniona dezinformacją, propagandą i sygnałami manipulacyjnymi. Każde państwo próbuje zdobyć jak najwięcej zasobów informacyjnych, podczas gdy

28 B. Stech, *Technologia Deepfake wciąż się rozwija. W USA bawi i budzi obawy,* 01.02.2019, https://www.purepc.pl/technologia/technologia_deepfake_wciaz_sie_rozwija_w_usa_bawi_i_budzi_obawy [dostęp: 18.02.2019].

29 Robert M. Chesney jest amerykańskim prawnikiem i profesorem prawa na Uniwersytecie Texas School of Law, gdzie pełni funkcję prodziekana ds. Nnaukowych i prowadzi kursy związane z bezpieczeństwem narodowym USA i prawem konstytucyjnym.

30 Danielle Keats Citron jest profesorem prawa na Uniwersytecie Maryland Francis King Carey School of Law.

31 J. Hayden, S.R. Stroud, *How Deep Does the Virtual Rabbit Hole Go? "Deepfakes" and the Ethics of Faked Video Content*, https://mediaethicsinitiativeorg.files.wordpress.com/2018/03/6-ethics-of-deepfakes-case-study1.pdf [dostęp: 16.02.2019].

niektóre kraje przekształciły informacje w swoją własną broń w walce o prymat. Informacje tracą swoją pierwotną funkcję, zaczyna się chaos. Pozwoliliśmy, by ilość informacji, jak i możliwość dzielenia się nimi, a przede wszystkim zdolność rozprzestrzeniania informacji tak szeroko, jak to możliwe, stały się synonimem prawdy. Musimy odpowiadać zdecydowaną reakcją na to nowe wyzwanie dla naszego demokratycznego życia.

O wiele większa swoboda informacji wynikająca z ery cyfrowej może być celem arbitralnej władzy politycznej, a także może stać się instrumentem stosowanym do manipulacji przez różne podmioty, w tym mocarstwa. W wyborach z ostatnich dwóch lat w różnych państwach Zachodu dostrzec można szerzące się fałszywe wiadomości i hacking, mające na celu zakłócenie porządku publicznego, zagrażając wiarygodności sondaży wyborczych, a tym samym siejąc zamieszanie, wątpliwości i niezgodę. Jest to atak na samą suwerenność wybranych państw, który wykorzystuje pasywne podejście do tego niedopuszczalnego zjawiska – bierności, która graniczy z nieodpowiedzialnością.

Sprawcy tych zachowń starają się odwrócić zasady, na których opierają się demokracje – otwartość, wolność informacji i komunikacji – aby uczynić je instrumentami interferencji i destabilizacji. To nowa era propagandy. Dezinformacja nie jest oczywiście nowym zjawiskiem, ale cyfrowa rewolucja i jej wpływ na to, w jaki sposób społeczeństwo, a zwłaszcza młodzi ludzie, przekazują swoje wiadomości, zapewnia jej niespotykany dotąd zakres, stając się poważnym zagrożeniem. Ta ingerencja musi zostać rozwiązana poprzez kooperacje działań władz publicznych, odpowiedzialności korporacyjnej i czujności ze strony społeczeństwa obywatelskiego i mediów.

Bibliografia

Bailey J., *The deepest fake: how new tech will test our belief in what we see,* 04.05.2018, https://www.smh.com.au/technology/the-deepest-fake-how-new-tech-will-test-our-belief-in-what-we-see-20180423-p4zb4w.html [dostęp: 16.02.2019].

Barni M., Bondi L., Bonettini N., Bestagini P., Costanzo A., Maggini M., Tondi B., and Tubaro S., *Aligned and nonaligned double jpeg detection using convolutional neural networks.* „Journal of Visual Communication and Image Representation" 2017, nr 49.

Bezmalinovic T., *Wenn Merkel plötzlich Trumps Gesicht trägt: die gefährliche Manipulation von Bildern und Videos*, 03.02.2018, https://www.aargauerzeitung.ch/leben/digital/wenn-merkel-ploetzlich-trumps-gesicht-traegt-die-gefaehrliche-manipulation-von-bildern-und-videos-132155720 [dostęp: 16.02.2019].

Board W., *A reason to despair about the digital future: Deepfakes,* 06.01.2019, https://www.washingtonpost.com/opinions/a-reason-to-despair-about-the-digital-future-deepfakes/2019/01/06/7c5e82ea-0ed2-11e9-831f-3aa2c2be4cbd_story.html?utm_term=.8f013c9f7e16 [dostęp: 21.02.2019].

Brundage M. et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation,* Executive Summary February 2018, p.49; *Scarlett Johansson na tropie deepfake'ów*, 13.01.2019, https://niezalezna.pl/253924-scarlett-johansson-na-tropie-deepfake39ow [dostęp: 21.02.2019].

Chiny: Serwisy informacyjne poprowadzą prezenterzy wygenerowani komputerowo, 9.11.2018, https://www.cdaction.pl/news-55063/chiny-serwisy-informacyjne-poprowadza-prezenterzy-wygenerowani-komputerowo-wideo.html [dostęp: 20.02.2019].

Cyganek A., *Pomysłowość Chińczyków nie zna granic. Stworzyli prezentera za pośrednictwem DeepFake.* 09.11.2018, https://www.instalki.pl/aktualnosci/internet/33377-pomyslowosc-chinczykow-nie-zna-granic-stworzyli-prezentera-za-posrednictwem-deepfake.html [dostęp: 20.02.2019].

Defense Advanced Research Projects Agency: Overview and Issues for Congress Updated, July 24, 2018. Congressional Research Service, https://crsreports.congress.gov; *Creating Breakthrough Technologies And Capabilitiesfor National Security*, https://www.darpa.mil/ [dostęp: 16.02.2019].

Dodge A., House L., Johnstone E., *Using Fake Video Technology To Perpetrate Intimate Partner Abuse Domestic Violence Advisory Ridder*, Costa & Johnstone LLP1, https://withoutmyconsent.org/sites/default/files/blog_post/2018-04-25_deepfake_domestic_violence_advisory.pdf [dostęp: 20.02.2019].

Fake News: Read All About It, ed. The New York Times Editorial Staff. The Rosen Publishing Group, Inc, 2018, p.34-35; *A Reddit User Starts 'Deepfake',* 27.10.2017, https://www.eyerys.com/articles/timeline/reddit-user-starts-deepfake?page=8 [dostęp: 16.02.2019].

FakeApp, https://www.malavida.com/en/soft/fakeapp/#gref [dostęp: 15.02.2019].

Gensin P., *Deepfakes: Auf dem Weg in eine alternative Realität?* 22.02.2018, http://faktenfinder.tagesschau.de/hintergrund/deep-fakes-101.html [dostęp: 16.02.2019].

Giles M., *Five emerging cyber-threats to worry about in 2019,* January 4, 2019.

Hayden J., Stroud S.R., *How Deep Does the Virtual Rabbit Hole Go? "Deepfakes" and the Ethics of Faked Video Content,* https://mediaethicsinitiativeorg.files.wordpress.com/2018/03/6-ethics-of-deepfakes-case-study1.pdf [dostęp: 16.02.2019].

Heartfield R., Loukas G., *Protection Against Semantic Social Engineering Attacks*. "Versatile Cybersecurity. Advances in Information Security", Edition: 72, Chapter: 4, Publisher: Springer, Cham, 2018.

Jordan Peele's simulated Obama PSA is a double-edged warning against fake news, 18.04.2018, https://www.vox.com/2018/4/18/17252410/jordan-peele-obama-deepfake-buzzfeed; [dostęp: 15.02.2019].

Just V., *AI could make dodgy lip sync dubbing a thing of the past,* 17.08.2018, https://techxplore.com/news/2018-08-ai-dodgy-lip-sync-dubbing.html [dostęp: 15.02.2019].

Kopciak P.A., *Fake Algorithms: Your face in my video. The influence of elaborate fake videos on our perception and socjety. History, theories and current developments in the media landscape, regarding aesthetics and socjety*. University of Applied Sciences St. Pölten Master course "Digital Mediatechnologies", Vienna, 2018.

Kruczkowski Ł., *FakeApp – czy powinniśmy obawiać się aplikacji, która z każdego może zrobić gwiazdę filmów dla dorosłych?* https://www.onet.pl/?utm_source=technologie_viasg&utm_medium=nitro&utm_campaign=allonet_nitro_new&srcc=ucs&pid=-554c238a-8523-51f1-8b35-dd5054acf3b9&sid=a6dee8f1-c258-4a9c-991e-af9040eb509b&utm_v=2 [dostęp: 19.02.2019].

Lindenberg G., *Głęboko fałszywa rzeczywistość. Jeśli nie wiadomo, co jest prawdą, to lepiej w nic nie wierzyć*, https://wiadomosci.onet.pl/tylko-w-onecie/deepfake-manipulacja-grozniejsza-niz-klasyczne-fake-newsy/c10cyzp [dostęp: 19.02.2019].

Prezenterzy, którzy nie istnieją? Agencja Xinhua wykorzystuje technologię DeepFake, 10.11.2018, https://pl.aleteia.org/2018/11/10/prezenterzy-ktorzy-nie-istnieja-agencja-xinhua-wykorzystuje-technologie-deepfake/ [dostęp: 20.02.2019].

Rivera D., et al. *Secure Communications and Protected Data for a Internet of Things Smart Toy Platform*. „IEEE Internet of Things Journal" 2019.

Scarlett Johansson na tropie deepfake'ów, 13.01.2019, https://niezalezna.pl/253924-scarlett-johansson-na-tropie-deepfake39ow [dostęp: 19.02.2019].

Stech B., *Technologia Deepfake wciąż się rozwija. W USA bawi i budzi obawy,* 01.02.2019, https://www.purepc.pl/technologia/technologia_deepfake_wciaz_sie_rozwija_w_usa_bawi_i_budzi_obawy [dostęp: 18.02.2019].

Stopka A., *Sztuczna inteligencja. Wielka nadzieja czy wielkie zagrożenie*? 13.09.2017, https://pl.aleteia.org/2017/09/13/sztuczna-inteligencja-wielka-nadzieja-czy-wielkie-zagrozenie/ [dostęp:16.02.2019].

Suwajanakorn S., Seitz S.M., Kemelmacher-Shlizerman I., *Synthesizing Obama: Learning Lip Sync from Audio,* ACM Transactions on Graphics (SIGGRAPH), 4 July 2017, https://doi.org/10.1145/3072959.3073640. [dostęp: 15.02.2019].

Świtalski K., *Deepfake: zabawa w kotka i myszkę z politycznymi szantażami, propagandą i celebryckim porno*. 31.12.2018, https://antyweb.pl/walka-z-deepfake/ [dostęp: 21.02.2019].

Tomaszkiewicz M., *Chińczycy stworzyli sztucznego prezentera wiadomości,* 09.11.2018, https://www.antyradio.pl/Technologia/Duperele/Chinczycy-stworzyli-sztucznego-prezentera-wiadomosci-27057 [dostęp: 20.02.2019].

Turek M., *Media Forensics (MediFor*), https://www.darpa.mil/program/media-forensics [dostęp: 16.02.2019].

Wasiuta O., *Sieci społecznościowe jako nowe narzędzia prowadzenia wojen informacyjnych we współczesnym świecie*, [w:] *Refleksje o przeszłości, spojrzenie na współczesność: monografia poświęcona Profesorowi Sergiuszowi Wasiucie z okazji 60-letniego Jubileuszu i 35-lecia pracy zawodowej*, red. O. Wasiuta, Drukarnia Styl Anna Dura, Kraków 2018.

What are deepfakes & why the future of porn is terrifying, https://www.highsnobiety.com/p/what-are-deepfakes-ai-porn/ [dostęp: 16.02.2019].

Ястремська Т., *Deepfake: як жити у світі, де підробку не відрізнити від реальності,* Частина І. 26 червня, 2018, https://kfund-media.com/deepfake-yak-zhyty-u-sviti-de-pidrobku-ne-vidriznyty-vid-realnosti-chastyna-i/ [dostęp 18.02.2019].

Deepfake as a complicated and deeply false reality

Abstract

Manipulation of information is not a new phenomenon, but has taken on a completely new dimension, due to the unprecedented possibilities of the Internet and social networks to disseminate information and create viruses out of them, as well as the crisis of trust currently experienced by democracies. However, the ability to distort reality has caused a rapid leap forward thanks to the technology of "deepfake" - a technique of synthesis of human image based on artificial intelligence, which enables creating real video with people speaking and doing things that they would never say or do. Machine learning techniques increase the level of technology advancement, making deep counterfeits more and more realistic and more resistant to detection.

Słowa kluczowe: DeepFakes, fałszywe wiadomości, media społecznościowe, manipulacja informacją, techniki uczenia maszynowego

Key words: DeepFakes, fake news, social media, information manipulation, machine learning techniques.

Olga Wasiuta
profesor zwyczajny, doktor habilitowany, Dyrektor Instytutu Nauk o Bezpieczeństwie, Kierownik Katedry Bezpieczeństwa Narodowego Uniwersytetu Pedagogicznego im. Komisji Edukacji Narodowej w Krakowie. Zajmuje się problematyką bezpieczeństwa regionalnego i europejskiego oraz wojną hybrydową. Jest autorką lub współautorką m.in. takich monografii, jak: *Wojna hybrydowa Rosji przeciwko Ukrainie* (Kraków 2017), *Państwo Islamskie ISIS. Nowa twarz ekstremizmu* (Warszawa 2018); współredaktor monografii wieloautorskich poświęconych bezpieczeństwu: *Współczesne wyzwania bezpieczeństwa europejskiego* (Kraków 2016), *Współczesne problemy bezpieczeństwa państwa* (Stalowa Wola 2017), *Wyzwania bezpieczeństwa międzynarodowego* (Stalowa Wola 2017)*; Vademecum bezpieczeństwa* (2018); jest również autorką licznych artykułów naukowych. Jest członkiem Polskiego Towarzystwa Geopolitycznego, członkiem Komitetu Redakcyjnego oraz przewodniczącą Rady Programowej czasopisma „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate", zastępca redaktora naczelnego czasopisma naukowego „Socjologia prawa" (Ukraina, Kijów). E-mail: olga.wasiuta@up.krakow.pl

Sergiusz Wasiuta
profesor zwyczajny, doktor habilitowany, pracownik Instytutu Politologii Uniwersytetu Pedagogicznego im. Komisji Edukacji Narodowej w Krakowie. Jest autorem ponad 250 prac naukowych opublikowanych w wydawnictwach krajowych i zagranicznych na temat historii stosunków polsko-ukraińskich na przełomie XX–XXI; zajmuje się politycznym i społeczno-ekonomicznym rozwojem Ukrainy w okresie niepodległości; w kręgu jego zainteresowań naukowych znajdują się również badania nad walka informacyjną, problemy bezpieczeństwa społeczno-informacyjnego i energetycznego w kontekście zagrożeń hybrydowych i cywilizacyjnych; przyczyny i skutki kryzysu ekologicznego oraz jego wpływ na bezpieczeństwo międzynarodowe. Jest autorem lub współautorem licznych publikacji naukowych, w tym ponad 10 monografii: *Wojna hybrydowa Rosji przeciwko Ukrainie* (Kraków 2017), *Państwo Islamskie ISIS. Nowa twarz ekstremizmu* (Warszawa 2018); *Екологічна політика*: *національні та глобальні реалії*. У 4-х томах. (Чернівці 2003–2004) і in.; jest autorem licznych artykułów naukowych. Dziś jest zastępcą redaktora naczelnego i współredaktorem kwartalnika „Przegląd Geopolityczny" (Geopolitical Review) (Polska, Kraków); zastępcą redaktora naczelnego czasopisma „Ekologiczne prawo Ukrainy" (Ukraina, Kijów); ekspertem Centrum Europy Wschodniej Uniwersytetu Marii Curie-Skłodowskiej w Lublinie w zakresie problematyki historycznej, politycznej, ekonomicznej, społecznej i kulturowej państw Europy Wschodniej. E-mail: sergiusz.wasiuta@up.krakow.pl.