

Jacek Wachowski

# Deepfejki – medialne falsyfikacje ciała. Cechy, funkcje, strategie komunikacyjne

## 1. Źródła

Deepfejki – *deepfakes*<sup>1</sup> – powstały pod koniec 2017 roku [Brandon]. Anonimowy autor podpisujący się nickiem *deepfakes* zamieścił – na stronie popularnego serwisu Reddit (prezentującego linki do informacji ukazujących się w sieci) – film pornograficzny z Gal Gadot (znaną izraelską aktorką i modelką). Montaż fikcyjnego materiału opierał się na stosunkowo prostej operacji. Polegała ona na przeniesieniu twarzy Gadot na postać nagiej aktorki porno o podobnej budowie ciała [Gillin]. Zainteresowanie filmem przerosło oczekiwania autora i pociągnęło za sobą podobne praktyki. Nie powstrzymały ich uzasadnione obawy, że fałszowanie danych publikowanych w sieci może być narzędziem służącym do ataku na osoby publiczne. Przeciwnie, akcentowano rozrywkowe i komercyjne walory *deepfejków* – oraz podkreślano, że są efektem liberalizacji społecznych obyczajów i nie łamią prawa, lecz umiejętnie wykorzystują jego niedostatki [“We Are Truly”].

---

1 W tekście będę używał spolszczonej wersji angielskiego określenia *deepfake*. Uproszczenie takie wynika zarówno ze względów praktycznych (odmiana obcego rzeczownika zakończonego na samogłoskę sprawia trudności polskiej fleksji), jak również z zakorzenienia pisowni *fejk* w polszczyźnie, zob. [“Fejk”].

W styczniu 2018 roku na rynku internetowym pojawiła się aplikacja FakeApp umożliwiająca nakładanie obrazów źródłowych (twarzy i całej postaci) na matryce tworzące nowe ekspresje ruchowe, gestyczne i werbalne [Schwartz; Dąbrowska 90]. Aplikacja wykorzystywała algorytm GAN – Generative Adversarial Networks (generatywne sieci przeciwstawne) – który współpracował z dwoma systemami nadzorowanymi przez sztuczną inteligencję: generatorem i dyskryminatorem [Young 14]. Podczas gdy pierwszy umożliwiał tworzenie fałszywego obrazu, drugi demaskował jego najsłabsze elementy. W efekcie wielokrotnych powtórzeń generator mógł „uczyć się” właściwego (i coraz trudniejszego do rozpoznania) sposobu fałszowania danych. Współpraca urządzeń kończyła się wtedy, gdy dyskryminator nie był w stanie wykryć nieprawidłowości (wad) pliku przesyłanego z generatora [Young 14; Dąbrowska 90].

Deepfejski nie powstały niespodziewanie. Miały ścisły związek z wykreowaniem środowiska technologiczno-komunikacyjnego, zainteresowanego upowszechnieniem nieprawdziwych informacji, plotek oraz teorii spiskowych [Klepka]. Nieprzypadkowa w tym kontekście wydaje się publikacja pierwszego deepfejska na platformie Reddit, znanej z rozpowszechniania anonimowych materiałów (niezweryfikowanych przez ekspertów), oferującej możliwość wypowiedzenia dowolnej opinii na dowolny temat oraz wspierającej tworzenie ryzykownych sposobów przetwarzania i dystrybucji danych [Brandon]. W wypadku Reddita o wiarygodności informacji decydowała liczba komentarzy i wyświetleń. W ten sposób ich ocena merytoryczna została zastąpiona przez kryterium popularności (zasięg) [Soroush et al.].

Równie ważną okolicznością sprzyjającą powstaniu deepfejsków była dostępność narzędzi służących do ich konstruowania. Większość aplikacji opierała się na prostych (intuicyjnych) zasadach. Należały do nich zarówno narzędzia „domowe” (takie jak: Snapseed, Snapchat czy Photoshop Express), jak i programy przygotowane do specjalistycznej obróbki danych (np. mGIMP, Photopea, Polarr Photo Editor, Fotor, Photoscape X, Pixar, Photoshop). Zakres działania tych ostatnich dobrze ilustruje algorytm Few-Shot Adversarial Learning of Realistic Neural Talking Head Models. Na podstawie materiału wyjściowego – który stanowiła rejestracja danych pochodzących z ośmiu niezależnych kamer – program tworzył wstępną matrycę, wybierając najbardziej charakterystyczne cechy postaci (związane z uśmiechem, wyrazem oczu, powtórzeniami napięć mięśni twarzy, charakterystyczną gestykulacją i ruchem), a następnie uczył się ich naśladowania. Na podstawie zapamiętanych ekspresji powstawały autonomiczne sekwencje (ruchowe, gestykulacyjne i werbalne) – schematy działań oparte na powtórzeniach zakodowanych informacji (a więc tych, których system się wcześniej „nauczył”) [Zakharov].

## 2. Przykłady

Najczęściej cytowanymi przykładami deepfejków są wypowiedzi polityków, sprawujących urzędy i osób publicznych. Słowa włożone w ich usta – których nigdy nie wypowiedzieli – i gesty oraz działania – których nigdy nie wykonali – wywołują trudne do oszacowania skutki społeczne i prawne. Listę dobrze znanych nadużyć otwiera kanał BuzzFeedVideo, w którego wideo Barack Obama wypowiada słowa na temat zagrożeń deepfejkami [BuzzFeedVideo]. Kłopot polega jednak na tym, że były prezydent Stanów Zjednoczonych nigdy takich słów nie wypowiedział, a materiał został skompilowany z fragmentów jego ogólnie dostępnych wywiadów. Na tej samej liście znajdują się: spot, w którym Mark Zuckerberg ostrzega przed zagrożeniami wynikającymi z powszechnego stosowania nowych technologii [TimesLIVE Video]; przemówienie koreańskiego dyktatora Kim-Jong Una na temat zalet demokracji [RepresentUs] oraz życzenia bożonarodzeniowe złożone przez królową Elżbietę II [Channel 4]. W ostatnim miesiącu na listę słynnych deepfejków trafił materiał przygotowany przez kremłowską telewizję, w którym widać, jak ręka Władimira Putina przenika przez mikrofon i jak w czajnikach stojących na stole nie odbija się to, co na ich szklanej powierzchni powinno być widoczne [śląski].

Deepfejki mogą jednak naruszać dobra osób nie tylko publicznych. Świadczy o tym aplikacja DeepNude, usuwająca odzież ze zdjęć kobiet zamieszczanych w sieci (dodać trzeba, że program nie był przystosowany do obróbki fotografii mężczyzn). Algorytm powstał na bazie popularnego programu pix2pix, opracowanego w 2017 roku na Uniwersytecie w Berkley dla zupełnie innych celów (najlepsze efekty przynosił wtedy, gdy przekształceniom poddawane były podobizny osób występujących w kostiumach kąpielowych – eksponujących odsłonięte fragmenty ciała) [Rymsza]. DeepNude – korzystając z otwartych zasobów Windowsa i Linuksa, otworzył drogę do podobnych praktyk [“This Horrifying App”]. Ich celem były nie tylko łatwe do zdemaskowania „zabawy twarzami” – takie jak nakładanie mimiki białostockiego wideoblogera Krzysztofa Kononowicza na postać znanej dziennikarki Doroty Wellman [Ten Deepfake] lub podkładanie twarzy i głosu Mariana Paździocha (postaci wykreowanej przez Ryszarda Kotysa w znanym sitcomie) do postaci Wiedźmina [Vinnie The Kielbasa King] – ale także symulacje znacznie trudniejsze do rozpoznania: komunikaty o wysokim poziomie prawdopodobieństwa, o których moglibyśmy pomyśleć, że miały szansę się wydarzyć.

Deepfejki – niezależnie od tego, że większość z nich nie cieszy się dobrą sławą – mogą też pełnić funkcje społecznie akceptowalne. Do takich należy Living Portraits, aplikacja animująca działania Mony Lisy [The Telegraph]. Nie chodzi tu wyłącznie o zabawę znanym obrazem, którą przed stu laty rozpoczął Marcel Duchamp, ale raczej o to, że Gioconda – która mówi, uśmiecha się, mruga i gestykuluje – wprowadza widzów w alternatywną rzeczywistość, ignorującą systemy społeczne, aksjologiczne i prawne. Pokazuje przy tym, jak proste w obsłudze

oprogramowanie może prowadzić do powstawania symulaków rewitalizujących obieg sztuki, łamiąc przy okazji szereg interpretacyjnych przyzwyczajęń.

Z podobnych inspiracji zrodził się pomysł kustoszy Muzeum Salvadora Dalego w Saint Petersburgu na Florydzie, którzy – wraz z agencją Goodby, Silverstein & Partners – zebrali ponad sześć tysięcy próbek graficznych i dźwiękowych utrwalających ekspresje Dalego z różnych okresów jego życia. Efektem projektu było stworzenie awatara artysty, który nie tylko rozmawiał z widzami – wchodził z nimi w interakcje, animował ich działania i zarządzał muzealnym ruchem – ale też zapraszał do „wyprodukowania” wspólnego deepfejk [Borzechowski].

Living Portraits i praktyki muzealników z Florydy można traktować jak zapowiedź niewinnych zabaw – takich jak kręcenie filmów z Marilyn Monroe czy nagrywanie wirtualnych koncertów ABBY – a także spopularyzowanych w ostatniej dekadzie „zabaw własnymi twarzami”, do których zapraszani są mniej sławni użytkownicy. Warto przypomnieć, że Snapchat – stworzony przez Evana Spiegla i Bobby’ego Murphy’ego w 2011 roku – ma obecnie około 330 milionów użytkowników (a ich liczba podwoiła się w okresie pandemii) [Statista; tw].

### 3. Strategie komunikacyjne – metody weryfikacji

Przykłady znanych deepfejków zwracają uwagę na chaos występujący zarówno na poziomie ich opisu, jak i konceptualizacji (dokonywanej najczęściej w obrębie dyskursów: medioznawczego, filozoficznego, kulturoznawczego i związanego ze studiami nad wizualnością). Uświadamiają, że samo pojęcie (deepfejk) ma charakter tyleż ogólny, co ambiwalentny. Obejmuje zróżnicowane praktyki, ale niewiele mówi o metodach, na jakich się opierają. Nie przybliży też do opisanego różnic, które występują między poszczególnymi ich odmianami.

Uporządkowaniu refleksji poświęconej deepfejkom może służyć przypomnienie, że ich materią są medialne wizualizacje ciała, zarejestrowane w plikach filmowych (fabularnych i dokumentalnych), w spotach reklamowych i promocyjnych oraz prezentacjach multimedialnych. Deepfejki powstają dwuetapowo. Najpierw wprowadza się pliki cyfrowe do przestrzeni sieciowej – dzięki czemu stają się one własnością publiczną – a następnie tworzy się ich nowe (przekształcone) wizualizacje, publikowane jako formaty zweryfikowane i autoryzowane (a więc niewywołujące sprzeciwu dysponentów danych, które zostały przetworzone). Ciało poddane takim zabiegom staje się obiektem cyfrowej kolektywizacji – jego właścicielami okazują się wszyscy użytkownicy sieci, którzy mają narzędzia i kompetencje do jego przekształceń. Z jednej strony podlega dowolnym i bezkarnym transformacjom, a z drugiej pozostaje emanacją społecznych pragnień, potrzeb i wyobraźni – przedmiotem kreacji i falsyfikacji. Jest ciałem zmediatyzowanym, technologicznie wprowadzonym – *fake body* – hybrydą zacierającą granice między

wizualizacją i reprezentacją. Obrazem danych, możliwych do nieustannego przeprogramowywania i aktualizowanych w czasie rzeczywistym. Dzieje się tak dlatego, że medialne wizualizacje ciała nie wchodzą w relacje ze światem zewnętrznym (z analogowymi reprezentacjami, które mają doniosłe znaczenie społeczne i polityczne), ale nawiązują kontakt z innymi bazami danych, tworząc wzajemnie synchronizujące się układy *data-to-data*, zawierające dodatkowo autonomiczne kody operacyjne, co oznacza, że same stają się programami [Hoelzl and Remi]. W ten sposób binarne podziały na prawdę i fikcję, a także reprezentację i wizualizację, zostają zastąpione przez hybrydowe kontaminacje prowokujące do tego, aby myśleć o nich poprzez pryzmat „rzeczywistości platform cyfrowych” [Hinterwaldner; Hoelzl and Remi; Gunkel].

Przypomnienie, że materia deepfejków są medialne wizualizacje ciała, prowadzi do scharakteryzowania zasad, na jakich się opierają – trwałych i nieskomplikowanych schematów konstrukcyjnych. Przybierają one postać „działań przez ciało”, „działań na ciało” i „działań wokół ciała”. W pierwszym wypadku chodzi o połączenie wizerunków znanych postaci z treściami, które nie zostały przez nie autoryzowane. Włożenie w usta osób publicznych słów, których nigdy nie wypowiedziały, stanowi próbę naruszenia ich społecznych wizerunków i wykreowania nowych (zaskakujących, prowokacyjnych i niewiarygodnych). Przykładem mogą być spoty z wypowiedziami Baracka Obamy, Marka Zuckerberga, królowej Elżbiety i koreańskiego dyktatora, a także Doroty Wellman i Ryszarda Kotysa. Przekształceniu nie ulegają tu wizerunki znanych postaci, ale treści, które zostały dołączone do owych wizerunków – idee, które zostały z nimi powiązane.

W drugim przypadku – „działania na ciało” – mamy do czynienia z dodaniem do obrazu ciała (lub odjęciem od niego) takich atrybutów, które wpływają na jego wygląd i zmieniają sposób jego postrzegania. Ich celem jest remediacja, naruszająca konstytutywne własności ciała. Przykładem takich aplikacji może być Nude (oraz seria eksperymentów zamieszczanych na platformach podobnych do Reddit), których celem jest zmiana wyglądu ciała, wpływająca na jego odbiór, społeczny status.

W trzecim wypadku – „działań wokół ciała” – chodzi o zmianę środowiska, w którym ciało zostało umieszczone. Przykładem mogą być zabawy muzealników z Moną Lisą (Living Portraits) i Salvadorem Dalim. Opierają się one na przekształceniu obrazu ciała za pomocą kontekstu – społecznego otoczenia. Podobne narzędzia wykorzystywane są w dobrze znanych aplikacjach edukacyjnych (takich jak Teams czy Zoom), umożliwiających tworzenie spersonalizowanych środowisk (najprostszą operacją jest zmiana tła), w których mogą występować zarówno gospodarz spotkania, jak i jego uczestnicy.

Trzy strategie komunikacyjne – kierunki, w jakich zmierzają medialne przekształcenia ciała – mają też dalek idące konsekwencje. Wyznaczają przestrzeń, w której możliwa jest analiza danych źródłowych, i pozwalają na określenie

stopnia ich przekształcenia. Do ich rozpoznawania służą autoenkodery (*autoencoders*) – sieci neuronowe składające się z elementów określanych mianem sztucznych neuronów, których matematyczny model został opracowany przez Warrena McCullocha i Waltera H. Pittsa w 1943 roku [Kosiński; Gately; Tadeusiewicz]. Ich celem jest analiza danych wyjściowych dotyczących postaci, jak też kontekstu, w którym zostały umieszczone. Składają się one z co najmniej trzech warstw: wejściowej (kodera – *encoder*), ukrytej (przekształcającej dane wejściowe i tworzącej ich zakodowaną reprezentację) oraz dekodera (*decoder*) przywracającego zakodowane dane do poziomu wyjściowego. System kodowania i dekodowania polega na tym, że poszczególnym połączeniom informacji nadaje się określone parametry (zwane wagami), a następnie porównuje się ich wartość występującą między poziomem wejściowym i wyjściowym [Aggarwal; Foster]. Porównanie wybranych połączeń (wag) pozwala na ustalenie, czy w materiale przesyłanym między enkoderem i dekoderelem zostały dokonane zmiany [Xie et al.; Młynarczyk]. Autoenkodery mają też dodatkowe właściwości – potrafią się uczyć – zapamiętywać zakodowane dane wyjściowe oraz anomalie i powiększać bazy danych, gromadzące wiedzę na temat technik manipulacyjnych, a także opracowywać na nie odpowiedzi [Aggarwal; Foster].

Jednym z pierwszych systemów zaprojektowanych do ochrony przed deepfejkami był MediFor opracowany przez DARPA (Defense Advanced Research Projects Agency) w 2016 roku, umożliwiający automatyczną ocenę integralności cyfrowych zdjęć i filmów. Okazał się on bardziej sprawny niż wcześniejsze algorytmy, które analizowały tylko jedną z własności obrazu – na przykład mrugnięcia powiek [Wasiuta and Wasiuta 25]. MediFor umożliwiał badanie złożonych połączeń występujących między rozproszonymi elementami obrazu i wykrywanie manipulacji na podstawie: układu włosów (sposobu ich opadania), profilu uszu, światła odbijającego się w źrenicach, a także – co może najważniejsze – poprzez uchwycenie detektora pulsu na czole (cyfrowego identyfikatora) pozwalającego na weryfikację autentyczności zdjęcia [Świtalski].

Podobna metoda – polegająca na oznaczaniu zdjęć za pomocą unikatowego znaku wodnego (niewidocznego dla oka) wygenerowanego przez sieć neuronową w trakcie wykonywania fotografii – została opracowana w Tandon School of Engineering na New York University. System pozwalał na wychwycenie zmian w formacie zdjęcia (kontrastu, światła czy nieznacznych różnic w kompozycji kadru), a następnie weryfikował dane przed przesłaniem ich na portal społecznościowy. Weryfikacja okazała się bardzo skuteczna – pozwalała na zidentyfikowanie około 90% deepfejków [Kołtunowicz] – i dawała możliwość automatycznego blokowania plików budzących wątpliwości oraz umożliwiała tworzenie repozytoriów katalogujących najczęściej występujące manipulacje.

W 2017 roku (rok po zainicjowaniu projektu DARPA) amerykańska AI Foundation opracowała oprogramowanie Reality Defender pozwalające na ocenę stopnia

ingerencji dokonywanych w przesyłanych na serwer plikach. Głównym filarem stworzonego systemu była systematycznie aktualizowana baza danych – zarządzana przez Globalną Radę ds. Sztucznej Inteligencji – której celem było analizowanie danych i przewidywanie kierunku, w jakim mogą zmierzać nowe rozwiązania technologiczne [Wasiuta and Wasiuta 25].

Podobne systemy wykorzystuje się powszechnie do analizy mowy. Technologie takie jak ASR (Automatic Speech Recognition) albo CSR (Computer Speech Recognition) lub też STT (Speech to Text) wykorzystywane są w licznych aplikacjach, na przykład do wybierania kontaktu telefonicznego czy sterowania urządzeniami domowymi. Interfejsami głosowymi posługuje się obecnie wiele systemów operacyjnych (w tym również bankowych), które za pomocą analizy głosu są w stanie identyfikować klienta [Tomar].

Systemy zaprojektowane do wykrywania deepfejków – zarówno te, które są szeroko wykorzystywane, jak i te pozostające w fazie prototypowej – uświadamiają, że jesteśmy uczestnikami gry między nadawcami informacji, działającymi głównie w mediach społecznościowych, oraz wspieranymi przez wyspecjalizowane agencje i serwisy informacyjne odbiorcami, którzy próbują je weryfikować. Z jednej strony powstają więc algorytmy umożliwiające manipulowanie danymi, z drugiej systemy pozwalające na ich demaskowanie. Efektem tej rywalizacji jest zarówno systematyczne obniżanie wymagań sprzętowych i kompetencji związanych z ich obsługą, jak i dalsza technologizacja, która łączy się z koniecznością korzystania z coraz większej liczby aplikacji – a więc z przekazywaniem uprawnień do zarządzania algorytmami innym algorytmom.

#### 4. Cechy deepfejków

Analiza podstawowych kierunków, w jakich zmierzają medialne wizualizacje ciała, umożliwia opisanie ich najważniejszych cech. Pozwala na stworzenie siatki pojęć charakteryzujących ich właściwości. Prowadzi do wniosku, że deepfejki odznaczają się nieskomplikowaną budową, czytelnymi intencjami i nastawione są na realizowanie łatwych do rozpoznania celów. Co zatem potrafimy o nich powiedzieć?

Deepfejki są konstrukcjami z założenia nieprawdziwymi, przekraczającymi granice ustalonych społecznie norm aksjologicznych oraz zasad dotyczących dystrybucji danych. Są nie tylko intencjonalnie fałszywe – mają na celu promowanie informacji nieprawdziwych, prowokacyjnych i skandalizujących – ale tworzą też wrażenie, że są społecznie akceptowane, a więc że rama komunikacyjna, którą wyznaczają głównie media społecznościowe, usprawiedliwia charakter i skalę dokonywanych w ich obrębie transformacji. Zwracają też uwagę na to, że sfalsyfikowane informacje mają szeroki zasięg i mogą zostać wykorzystane jako: środek perswazji

politycznej, metoda rewitalizacji przeszłości (unieśmiertelniająca historyczne postaci i przywracająca je do społecznego obiegu), metoda komercjalizacji danych oraz narzędzie występujące na rynku gier, nastawione na kreowanie scenariuszy przekraczających ograniczenia czasoprzestrzenne i logiczne.

Ważną cechą deepfejków jest to, że redefiniują granice między reprezentacją i wizualizacją. Sprawiają wrażenie, że stworzone przez nie wyobrażenia są nową odmianą rzeczywistości dostosowaną do społecznych potrzeb i akcentującą ukryte w nich aspiracje. Medialne hybrydy nie odsyłają jednak do reprezentacji opierających się na tradycyjnych wizualnych modelach, ale do baz danych, które podlegają dalszym aktualizacjom. W ten sposób zawarte w nich informacje stają się przedmiotem sieciowych gier, zmieniających ich wcześniejsze ontologiczne charakterystyki. Opierają się na wyobrażeniach i aktualizacjach zacierających pamięć o pierwowzorach – artefaktach stanowiących ich podstawę. Nie zmagają się z tworzeniem zamkniętych wyobrażeń – końcowych produktów, lecz skłaniają do nieustannych ich modyfikacji. Sprzyja temu umiejętne wykorzystanie symulacji: dwu- i trójwymiarowych. Pierwsze wykorzystują osie płaskie i pionowe, których połączenie umożliwia rysowanie, przedstawienie ruchu, ustawienie czasu, a także import dźwięków. Drugie oprócz osi płaskich i pionowych wykorzystują osie kątowe, umożliwiające przedstawianie złożonych relacji między obiektami oraz środowiskiem i charakteryzujące się wysokim stopniem podobieństwa do oryginałów<sup>2</sup>.

O deepfejkach wypada też powiedzieć, że mogą stanowić groźny instrument kontroli i nadzoru. Mogą brać udział w tworzeniu informacji na temat logujących się telefonów i transakcji dokonywanych za pośrednictwem kart płatniczych, mogą być wykorzystywane do monitoringu ulicznego, tworzenia danych biometrycznych umieszczanych w kodach kreskowych na dokumentach identyfikacyjnych i algorytmów konsumenckich śledzących preferencje dotyczące zakupów i usług, mogą być wykorzystywane w systemach cyfrowych zabezpieczeń, składających się z pączkującej liczby haseł, kodów, a także numerów i liter, identyfikujących usługi bankowe, telefoniczne, medyczne, a nawet biblioteczne [Jelonek 47; Repesgaard 25]. Łatwo sobie wyobrazić, że deepfejki przeniesione do społecznych praktyk mogą być nie tylko źródłem zagrożeń, ale również prawdziwych nieszcześć.

---

2 Ilustracją symulacji dwuwymiarowych może być dobrze znany Adobe Flash (produkt firmy Adobe Systems). Do najpopularniejszych programów umożliwiających symulację trójwymiarową należą: Blender – oprogramowanie typu *open source*, przeznaczone do produkcji multimedialnych i projektowania symulacji wizualnych, 3D Studio Max – oprogramowanie stworzone przez Autodesk i przeznaczone do modelowania, renderowania i animacji 3D, umożliwiające tworzenie rozbudowanych światów, a także do tworzenia filmów i animacji, czy Autodesk Maya – stworzone przez Autodesk, umożliwiające tworzenie realistycznych postaci i efektów specjalnych oraz animacji VFX, oferujących tworzenie znikających obiektów, efektów specjalnych i sterowanie środowiskiem.



Podsumowując zatem, można powiedzieć, że deepfejki to: konstrukcje z założenia nieprawdziwe (intencjonalnie fałszywe), których osiłą są przekształcenia wizualizacji związanych z ciałem, zamazujących granicę między reprezentacją a wizualizacją, charakteryzujące się otwartością i skłaniające do nieustannych aktualizacji danych, podejmujące grę z publicznością, która polega na tworzeniu nowych kanałów komunikacyjnych, a także kreowaniu środowisk promujących ich obieg.

## 5. Konsekwencje poznawcze

Deepfejki nie są wyłącznie problemem medialnym. Stanowią ważny głos w dyskusji na temat ciała. Przede wszystkim dlatego, że zmieniają relacje między *body image* – wyobrażeniem ciała, które powstaje w oczach podmiotu, i *body schema* – schematem ciała, czyli społecznie ukształtowanym jego modelem. Pojęcia te – jak mówi Shaun Gallagher – niezależnie od skłonności do ich utożsamiania (co dobrze widać w pracach Jeana Piageta, Ulrica Neissera, a nawet Maurice'a Merleau-Ponty'ego) opisują odmienne strategie pozwalające podmiotowi na nawiązanie relacji ze światem i z sobą samym [Gallagher 225-244].

Warto przypomnieć, że istnieją co najmniej trzy powody świadczące o ich odrębności. Po pierwsze, wyobrażenie ciała ma charakter wewnętrzny – powstaje w efekcie sprzężenia doznań percepcyjnych, konceptualnych oraz emocjonalnych – i ma unikatową postać, wypływającą z cech oraz predyspozycji podmiotu. Schemat zachowań (*body schema*) ma natomiast charakter zewnętrzny i jest zbiorem arbitralnie narzuconych ciału zobowiązań [Gallagher 226]. Po drugie, wyobrażenia na temat ciała mają status czysto intencjonalny. Niezależnie od tego, czy są świadome (czy też nieświadome lub podświadome), pozostają doznaniem, a więc statycznym obrazem – stanem niewymagającym działania. Inaczej jest ze schematem zachowań, który przenosi ciało w sferę czynu, wymusza niezbędną aktywność i konieczność dostosowania się do rozmaitych (zmiennych) okoliczności. Po trzecie, wyobrażenie ciała (*body image*) łączy się najczęściej z fragmentem ciała lub wybraną jego funkcją albo czynnością. Nie mówi o całym ciele, ale o jego części – opisuje ciało w określonej sytuacji, roli lub położeniu. Schemat przeciwnie. Nie rozkłada ciała na fragmenty i nie mówi o konieczności napinania określonych partii mięśni w celu uzyskania właściwego (pożądanego) efektu – odnosi się do całości [Gallagher 230-244].

Deepfejki zaburzają ten podział. Pokazują, że ciało przekształcone cyfrowo i wprowadzone do sieciowego obiegu pozostaje w znacznie większym stopniu emanacją potrzeb grupy niż podmiotu. Tracąc swoje indywidualne cechy, należy do wszystkich, którzy mają dostęp do baz danych utrwalających jego wizerunek i są zainteresowani kształtowaniem jego nowego obrazu [Piaget; Merleau-Ponty].

W ten sposób ciało wyznacza przestrzeń, w której powstają nowe kontrakty społeczne – umowy dotyczące sposobu i zakresu wykorzystania jego zmediatyzowanych wersji. To nie podmiot jest ich animatorem i gospodarzem i to nie on identyfikuje ciało zbiorowe jako część „wspólnoty genealogicznej” [Roach 124-136]. W jego imieniu czyni to grupa społeczna – a jej status nie wynika z kryteriów etnicznych, kulturowych czy religijnych [Bauman 70], lecz technologicznych. Ciało staje się w ten sposób bohaterem zbiorowego scenariusza i obiektem nieograniczonych kreacji, realizującym najbardziej zaskakujące społeczne fantazje. Współtworzy technokulturowe środowiska, które posługują się własnymi regułami komunikacyjnymi i systemami wartości.

Zmiana ta wywołuje również dalej idące skutki. Przede wszystkim wpływa na refleksję dotyczącą związków ciała z myśleniem (*body-mind*). Napięcia występujące między nimi stały się ważnym składnikiem zachodniej filozofii i doprowadziły – jak wiadomo – do powstania dwóch przeciwstawnych stanowisk: dualistycznego (wedle którego umysł i ciało należą do osobnych i pod wieloma względami zantagonizowanych bytów) oraz monistycznego (uznającego, że są one aspektami tej samej rzeczy). Zbiorowość – która przejmując prawo do dysponowania ciałem i tworzy jego sfalsyfikowaną wersję [Logika 68; *Wiedza a zagadnienie*] – kwestionuje obie tradycje i marginalizuje zarówno pytania dotyczące tego: „w jaki sposób umysł połączony jest z ciałem?”, jak też dociekania metafizyczne, które wyznaczyły najważniejsze sposoby myślenia o bycie [Chirpaz 94-97]. Związki ciała z myśleniem skłaniają raczej do namysłu nad biotechnologicznym wymiarem bytu – zarówno nad jego jakością, jak i poznawczymi ograniczeniami. Prowokują do pytań: w jakim kierunku będą zmierzać aplikacje tworzące nowe techno-ontologie? Jaki udział w dystrybucji nieprawdziwych danych będzie miała kultura nadmiaru informacji? Jakie skutki społeczne może wywołać długotrwałe wykorzystywanie deepfejków oraz w jaki sposób mogą one wpłynąć na kształtowanie się przyszłych standardów komunikacyjnych?

Pytania te skłaniają również do ponownego przemyślenia takich pojęć jak zwrot cyfrowy [Krzysztofek 38; Kluszczyński 7], który umożliwił skonstruowanie nowoczesnej infrastruktury technologicznej [Bustamante], czy koncepcja cyberprzestrzeni [Slouka; Grau; Koepsell], której literacką interpretację zaprezentował William Gibson w *Neuromancerze* [Gibson]. Nawet jeśli wizja sieci – wypełnionej komputerami podłączonymi do telekomunikacyjnych zasobów, umożliwiających transfer między rzeczywistością fizykalną i wirtualną – wydaje się przestarzała, to nie oznacza, że jest bezużyteczna. Pozwala na opisanie relacji, w jakie użytkownik (odbiorca) może wchodzić z urządzeniami technicznymi oraz innymi uczestnikami działań [Murray 72-85], a także dostarcza wiedzy na temat cielesnego przemieszczania w przestrzeni, przybierającego postać symulakry, określonej przez Gibsona mianem *consensual hallucination* – intensywnego złudzenia wprowadzającego do serii psychosomatycznych doznań [Gibson; Baudrillard]. Pozwala

wreszcie zrozumieć, że cyfrowe interfejsy, które stały się metajęzykiem służącym do zapisu danych oraz ich optymalizacji, a także okazały się odzwierciedleniem rozwoju języka matematycznego, pojmowanego jako specyficzny sposób rozumienia świata [Maciąg 100; Mirzoeff], mogą zostać wykorzystane do analizy deepfejków, w szczególności do rozpoznania problemów związanych z ich weryfikacją.

## 6. Wnioski

Deepfejki nie są jednorazową prowokacją. Wiele wskazuje na to, że wraz z rozwojem technologii ich zakres będzie się poszerzał. Nie można wykluczyć, że wpłynie to na redefinicję podstawowych pojęć znajdujących się w społecznym obiegu – także tych, które wcześniej wydawały się nienaruszalne. Trudne do oszacowania są również długofalowe skutki, wynikające z dystrybucji nieprawdziwych informacji. Już dzisiaj można jednak powiedzieć, że najważniejsze wyzwania związane z ich obiegiem ujawniają się w pięciu planach [Bąkowicz, Niklewicz; Soroush et al.].

Po pierwsze, dotyczą refleksji na temat prawdy i fałszu. Nie chodzi tu bynajmniej o wydłużanie listy definicji obu pojęć [Antas 131-162], ale raczej o pytanie: czy można je stosować do praktyk posttechnologicznych? I dalej: w jaki sposób odbiorcy medialnych komunikatów mieliby rozpoznawać sfałszowane konstrukcje? W jakie narzędzia należałoby ich wyposażyć, aby taka weryfikacja była możliwa? Pytania te uświadamiają, że jedną z najpilniejszych potrzeb społecznych są inwestycje w nowe technologie i systemy edukacyjne. Działania podejmowane w tym zakresie przez takie agencje jak DFDC (Deepfake Detection Challenge) czy DARPA (Defense Advanced Research Projects Agency) – które zajmują się tworzeniem narzędzi umożliwiających demaskowanie deepfejków – a także duże serwisy informacyjne oraz media społecznościowe wymagają wsparcia. W przeciwnym razie możemy być świadkami niewyobrażalnej zapaści poznawczej, degradującej porządek aksjologiczny, prawny, system wiedzy, nauczania, nie wyłączając autorytetów oraz instytucji, na których zostały one zbudowane.

Po drugie, deepfejki prowokują do refleksji na temat społeczeństwa opartego na informacji („społeczeństwa informacyjnego”). Chodzi tu przede wszystkim o przemyślenie statusu informacji, jak też ich roli w kształtowaniu więzi społecznych [Castells; Roszak] – w tym również o uwzględnienie tego, że informacje mają coraz częściej postać interaktywną i kwestionują podział na twórców (tzn. tych, którzy mają dostęp do baz danych i na ich podstawie konstruują przekaz) oraz odbiorców (którzy nie mają większego wpływu na ich treść, formę oraz dystrybucję). Konieczne w tym kontekście wydaje się postawienie pytania: na czym miałby polegać nowoczesny obieg informacji? Dotyczy to w szczególności dyskursów – na których opierają się założenia liberalnej demokracji – wyrażających przekonanie,

że każdy może powiedzieć, co chce, pod warunkiem, że nie narusza zasad zbiorowego porządku i prawa (cokolwiek ów „porządek” i „prawo” oznaczają).

Po trzecie, deepfejkki skłaniają do podjęcia badań kontekstualnych, poświęconych dystrybucji sfalsyfikowanych informacji, w tym badań poświęconych udostępniającym je platformom. Ważnym elementem takiej refleksji powinny być również studia poświęcone wykorzystywaniu danych w obrębie robotyki, cyfrowej pornografii i kradzieży tożsamości, a także analizy związane z *fake newsami* oraz szeroko rozumianą kulturą falsyfikacji [Jelonek; Reppesgaard].

Po czwarte, deepfejkki uzasadniają krytyczne opinie formułowane pod adresem podstawowych pojęć wypracowanych w latach 90. ubiegłego wieku (dotyczących zarówno cyberprzestrzeni, jak też AR – *artificial intelligence* i VR – *virtual reality*). Pojęcia te w dużym stopniu zostały zastąpione strategiami opartymi na współdziałaniu platform cyfrowych tworzących warunki do kreowania poglądów społeczno-politycznych na podstawie sieciowej struktury danych. Nie oznacza to jednak, że stare pojęcia są bezużyteczne. Deepfejkki prowokują raczej do przyjrzenia się ich znaczeniom w nowych poznawczych kontekstach. Skłaniają do namysłu nad przemianami technologicznymi, jakie zaszły w ostatnich dekadach.

Po piąte, deepfejkki prowokują do refleksji na temat przyszłości sfalsyfikowanych komunikatów. Wiele wskazuje na to, że mogą one zmierzać w dwóch kierunkach: „od ontologii do nowej sensualnej partycypacji” oraz „od sensualnej partycypacji do nowej ontologii”. Pierwszy kierunek polega na tworzeniu prototypowych rozwiązań, testujących i rozszerzających repertuar doświadczeń somatycznych, które mogłyby stać się częścią interfejsu pozwalającego odbiorcom na uczestnictwo w doznaniach zmysłowych sztucznie wykreowanego ciała. Drugi z kolei na tworzeniu nowej ontologii ciała – zdarzeń, które rewitalizują przeszłość i wprowadzają do niej fikcyjne informacje. W pierwszym wypadku technologia mogłaby umieścić widzów na tylnym siedzeniu prezydenckiej limuzyny jadącej ulicami Dallas i spowodować, że poczują zapach wody toaletowej Jackie Kennedy, a następnie będą uczestniczyć w sekwencji dobrze znanych wydarzeń. W drugim wypadku ta sama technologia mogłaby sprawić, że przejazd prezydenckiej limuzyny będzie miał inny finał. Oto siedzący obok Jackie widz mógłby w ostatniej chwili popchnąć jej męża, ratując mu życie, a następnie wyprowadzić prezydencką parę z limuzyny i zaaranżować serię zdarzeń, które przyjdą mu do głowy. Podobny scenariusz (lub scenariusze) – niezależnie od tego, że wydaje się dziełem nieograniczonej wyobraźni – musiałby (rzecz jasna) zostać poprzedzony badaniami rynku informującymi o oczekiwaniach widzów i o tym, w jaki sposób należałoby przekształcić rzeczywistość, aby zaspokoić ich potrzeby.

## Lista prac cytowanych

- Aggarwal, Charu. *Neural Networks and Deep Learning*. Springer, 2018.
- Antas, Jolanta. *O kłamstwie i kłamaniu. Studium semantyczno-pragmatyczne*. Universitas, 2000.
- Baudrillard, Jean. *Symulakry i symulacja*. Translated by Sławomir Królak, Wydawnictwo Sic!, 2005.
- Bauman, Zygmunt. *Ciało i przemoc w obliczu ponowoczesności*. Wydawnictwo Uniwersytetu im. Mikołaja Kopernika, 1995.
- Bąkowicz, Katarzyna. "Cel uświęca środki". *Fake news. Produkt medialny czasów postprawdy*, PhD dissertation, 2020, pp. 20-21, [https://depotuw.ceon.pl/bitstream/handle/item/3623/tekst\\_dr\\_calość\\_final\\_druk.pdf?sequence=3](https://depotuw.ceon.pl/bitstream/handle/item/3623/tekst_dr_calość_final_druk.pdf?sequence=3).
- Borzechowski, Emil. "W amerykańskim muzeum zrobisz sobie selfie z Dalim". *Na Ekranie*, 13 maja 2019, <https://naekranie.pl/aktualnosci/w-amerykan-skim-muzeum-zrobisz-sobie-selfie-z-dalim-1557743898>.
- Brandon, John. "Terryfying High-tech Porn. Creepy »Deepfake« Videos Are on the Rise". *Fox News*, 16 Feb. 2018, <https://www.foxnews.com/tech/terryfying-high-tech-porn-creepy-deepfake-videos-are-on-the-rise>.
- Bustamante, Enrique. "Cultural Industries in the Digital Age. Some Provisional Conclusions". *Media, Culture & Society*, vol. 26, 2004, pp. 802-820.
- BuzzFeedVideo. "You Won't Believe What Obama Says in This Video!". *YouTube*, 17 kwietnia 2018, <https://www.youtube.com/watch?v=p1b5aiTrGzY>.
- Castells, Manuel. *Spoleczeństwo sieci*. Translated by Janusz Stawiński, et al., Wydawnictwo Naukowe PWN, 2008.
- Channel 4. "Deepfake Queen: 2020 Alternative Christmas Message". *YouTube*, 25 grudnia 2020, <https://www.youtube.com/watch?v=IvY-Abd2FfM>.
- Chirpaz, François. *Ciało*. Translated by Jacek Migasiński, Wydawnictwo Instytutu Filozofii i Socjologii PAN, 1988.
- Cole, Samantha. "This Horrifying App Undresses a Photo of Any Woman With a Single Click". *Vice*, 26 Jun. 2019, <https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman>.
- . "We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now". *Vice*, 24 Jan. 2018, <https://www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley>.
- Dąbrowska, Ilona. "Deepfake – nowy wymiar internetowej manipulacji". *Zarządzanie mediami*, vol. 8, 2020, pp. 89-101.
- "Fejk". *Wikisłownik*, <https://pl.wiktionary.org/wiki/fejk>.
- Foster, David. *Generative Deep Learning*. O'Reilly, 2019.
- Gallagher, Shaun. "Body Schema and Intentionality". *The Body and the Self*, edited by Jose Luis Bermudes, et al., Routledge 1998, pp. 225-244.
- Gately, Ed. *Sieci neuronowe*. Translated by Adam Olejniczak, Wydawnictwo Wig-Press, 1999.
- Gibson, William, *Neuromance*. Ace Science Fiction Books, 1984.

- Gillin, Joshua. "Fact-checking Fake News Reveals How Hard It Is to Kill Pervasive 'Nasty Weed' Online". *PolitiFact*, 27 Jan. 2017, <https://www.politifact.com/article/2017/jan/27/fact-checking-fake-news-reveals-how-hard-it-kill-p/>.
- Grau, Oliver. *Virtual Art. From Illusion to Immersion*. MIT Press, 2003.
- Gunkel, David. *An Introduction to Communication and Artificial Intelligence*. MIT Press, 2020.
- Hinterwaldner, Inge. *The Systemic Image. A New Theory of Interactive Real-Time Simulations*. MIT, 2017.
- Hoelzl, Ingrid, and Marie Remi. *Softimage. Towards a New Theory of the Digital Image*. Intellect, 2015.
- Jelonek, Dorota. "Problem przeładowania informacyjnego w społeczeństwie informacyjnym". *Ekonomiczne Problemy Usług*, no. 67, 2011, pp. 45-52, [https://bazhum.muzhp.pl/media/files/Ekonomiczne\\_Problemy\\_Uslug/Ekonomiczne\\_Problemy\\_Uslug-r2011-t-n67/Ekonomiczne\\_Problemy\\_Uslug-r2011-t-n67-s45-52/Ekonomiczne\\_Problemy\\_Uslug-r2011-t-n67-s45-52.pdf](https://bazhum.muzhp.pl/media/files/Ekonomiczne_Problemy_Uslug/Ekonomiczne_Problemy_Uslug-r2011-t-n67/Ekonomiczne_Problemy_Uslug-r2011-t-n67-s45-52/Ekonomiczne_Problemy_Uslug-r2011-t-n67-s45-52.pdf).
- Klepka, Rafał. "Fake news". *Leksykon bezpieczeństwa*, edited by Olga Wasiuta, et al., Wydawnictwo Libron, 2018, pp. 299-304.
- Kluszczyński, Ryszard. *Społeczeństwo informacyjne. Cyberkultura. Sztuka multimedialności*. Rabid, 2001.
- Koepsell, David R. *The Ontology of Cyberspace. Philosophy, Law, and the Future of Intellectual Property*. University of Michigan Press, 2000.
- Kołtunowicz, Jędrzej. "Prosty sposób na deepfakes. Skuteczny?". *Sztuczna Inteligencja*, 22 stycznia 2020, <https://www.sztuczna-inteligencja.org.pl/prosty-sposob-na-deepfakes-skuteczny/>.
- Kosiński, Robert A. *Sztuczne sieci neuronowe. Dynamika nieliniowa i chaos*. Wydawnictwo Naukowo-Techniczne, 2014.
- Krzysztofek, Kazimierz. "Zwrot cyfrowy: ku pracy rozproszonej". *Zwrot cyfrowy w humanistyce. Internet. Nowe media. Kultura 2.0*, edited by Andrzej Radomski, and Radosław Bomba, E-naukowiec, 2013, pp. 37-56.
- Maciąg, Rafał. *Deus ex machina. Nowe media i ich projekt poznawczy*. Universitas, 2012.
- Merleau-Ponty, Maurice. *Fenomenologia percepcji*. Translated by Małgorzata Kowalska, Jacek Migasiński, forwarded by Jacek Migasiński, Fundacja Aletheia, 2001.
- Mirzoeff, Nicholas. *An Introduction to Visual Culture*. Routledge, 2003.
- Młynarczyk, Albert. "Autoencoder: Deep Learning Swiss Army Knife". *Digital Fingerprint*, 23 May 2021, <https://fingerprints.digital.pl/autoencoder-deep-learning-swiss-army-knife-2/>.
- Murray, Janet H. *Hamlet on the Holodeck. The Future of Narrative in Cyberspace*. MIT Press, 1998.
- Niklewicz, Konrad. *Weeding out Fake News: An Approach to Social Media Regulation*. Wilfried Martens Centre for European Studies, 2017.
- Piaget, Jean. *Play, Dreams, and Imitation in Childhood*. W.W. Norton, 1962.
- Popper, Karl R. *Logika odkrycia naukowego*. Translated by Urszula Niklas, Wydawnictwo Naukowe PWN, 2002.

- . *Wiedza a zagadnienie ciała i umysłu*. Translated by Tadeusz Baszniak, Książka i Wiedza, 1998.
- Reppesgaard, Lars. *Imperium Google*. Translated by Paulina Sadurska, Wydawnictwo BC.edu, 2009.
- RepresentUs. "Dictators – Kim Jong-Un". *YouTube*, 29 września 2020, [https://www.youtube.com/watch?v=ERQlaj\\_czHU](https://www.youtube.com/watch?v=ERQlaj_czHU).
- Roach, Joseph. "Culture and Performance in the Circum-Atlantic World". *Performativity and Performance*, Routledge, 1976, pp. 124-136.
- Roszak, Theodore. "The Cult of Information". *The Information Society Reader*, edited by Frank Webster, et al., Routledge, 2004, pp. 55-61.
- Rymsza, Anna. "DeepNude. Aplikacja, która rozbiera kobiety. Robi to przerażająco dobrze". *WP.pl*, 27 czerwca 2019, <https://tech.wp.pl/deepnude-aplikacja-ktora-rozbiera-kobiety-robi-to-przera-zajaco-dobrze,6396461084944513a>.
- Schwartz, Oscar. "You Thought Fake News Was Bad? Deep Fakes Are Where Truth Goes to Die". *The Guardian*, 12 Nov. 2018, <https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth>.
- Slouka, Mark. *War of the World. Cyberspace and the High-tech Assault of Reality*. Basic Books, 1995.
- Soroush, Vasoughi, et al. "The Spread of True and False News Online". *Science*, vol. 359, no. 6380, pp. 1146-1151.
- Statista. "Leading Countries Based on Snapchat Audience Size as of April 2022". *Statista*, <https://www.statista.com/statistics/315405/snapchat-user-region-distribution/>.
- śląski. "Przenikające palce Putina przez mikrofon Nexta vs Realty Deep Fake?". *YouTube*, 18 marca 2022, <https://www.youtube.com/watch?v=XjZ9SYAOVB4>.
- Świtalski, Kamil. "Deepfake: zabawa w kotka i myszkę z politycznymi szantażami, propagandą i celebryckim porno". *Antyweb*, 31 grudnia 2018, <https://antyweb.pl/walka-z-deepfake/>.
- Tadeusiewicz, Ryszard. "Zastosowanie sieci neuronowych do rozpoznawania mowy". *Analiza, synteza i rozpoznawanie sygnału mowy dla celów automatyki, informatyki, lingwistyki i medycyny*, edited by Łukasz Richter, Polska Akademia Nauk, Instytut Podstawowych Problemów Techniki, 1994, pp. 137-151.
- Ten Deepfake. "Deepfake: Dorota Weltmajster (Konon-Wellman)". *YouTube*, 2 lutego 2020, <https://www.youtube.com/watch?v=8b9ru-oFzLg>.
- The Telegraph. "AI Brings Mona Lisa to Life". *YouTube*, 24 maja 2019, <https://www.youtube.com/watch?v=thQ7QjqNPIY>.
- TimesLIVE Video, "Artists Create Zuckerberg 'Deepfake' Video". *YouTube*, 13 czerwca 2019, <https://www.youtube.com/watch?v=cnUdoTpuoXI>.
- Tomar, Vikrant. *Discriminant Feature Space Transformations for Automatic Speech Recognition*, Department of Electrical & Computer Engineering McGill University, 2012; [https://www.tomar.cc/pubs/reports/ML\\_report.pdf](https://www.tomar.cc/pubs/reports/ML_report.pdf).
- tw. "Snapchat z pierwszym zyskiem w historii, kurs akcji wystrzelił o 50 procent". *Wirtualne Media*, 27 lutego 2022, <https://www.wirtualnemedial.pl/artykul/>

- snapchat-z-pierwszym-zyskiem-w-historii-kurs-akcji-wystrzelil-o-50-proc.
- Vinnie The Kielbasa King. "Paździuch z Rivii. Trailer. [Wiedzmin PL Deep Fake]". *YouTube*, 28 stycznia 2020, <https://www.youtube.com/watch?v=sTR1vMxlcJ4&t=29s>.
- Wasiuta, Olga, and Sergiusz Wasiuta. "Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość". *Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate*, vol. 9, no. 3, 2019, pp. 19-30.
- Xie, Junyuan, et al. "Unsupervised Deep Embedding for Clustering Analysis". *Proceedings of the 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*; Cornell University, 24 Jun. 2022, <https://arxiv.org/pdf/1511.06335.pdf>.
- Young, Norbert. *DeepFake Technology. Complete Guide to Deepfakes, Politics and Social Media*. Independently published, 2019.
- Zakharov, Egor. "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models". *YouTube*, 21 maja 2019, <https://www.youtube.com/watch?v=p1b5aiTrGzY>.

#### Abstract

Jacek Wachowski

### **Deepfakes – Media Falsifications of the Body. Features, Functions, and Communication Strategies**

Deepfakes are based on media visualizations of the body, transforming its image by adding or subtracting certain features. In doing so, they falsify the body, changing its appearance and casting it in roles in which it did not previously appear. It is achieved via specialized applications and collaborative online platforms. They require careful analysis, as well as characterization of the mechanisms behind their manipulative strategies. This article attempts to organize reflection on deepfakes, paying attention to their technological and social dimensions. At the same time, it shows that deepfakes are based on simple construction schemes (they are easy to construct but also to deconstruct). The knowledge resulting from the analyses conducted in this area can be helpful both for "disarming" deepfakes and raising social competence to recognize them, and for creating systemic technological solutions – responses to their expansion.

**keywords:** deepfakes, digital media, manipulation techniques