

DOI 10.51558/2490-3647.2021.6.3.561

UDK 323.266:316.774

Primljeno: 10. 05. 2021.

Pregledni rad

Review paper

Amina Vatreš

DEEPPFAKE FENOMEN: NAPREDNA FORMA LAŽNIH VIJESTI I NJENE IMPLIKACIJE NA KREDIBILNO NOVINARSTVO

U epohi novomedijski obilježene društvene stvarnosti proces diseminacije dezinformacija postaje sve obuhvatniji i intenzivniji. Kao jedan od tehnoloških novuma na tom polju izdvaja se i deepfake fenomen – hiperrealistična digitalna falsifikacija video i audio formata. Shodno činjenici da za svoju osnovu ima najsofisticiraniju tehnologiju, podržanu dostignućima u domenu umjetne inteligencije, deepfake nedvosmisleno predstavlja svojevrsnu prekretnicu u kontekstu produciranja i distribucije lažnih audio-vizuelnih sadržaja. U težnji da poliperspektivno ukaže na mnogostruke reperkusije navedenog fenomena, rad pruža uvid u definiciju i specifične faktore koji doprinose prihvaćenosti i širenju deepfakea, s posebnim akcentom na njegove implikacije na kredibilno novinarstvo i njegovu društvenu funkciju. U tekstu se artikulira uloga deepfakea u procesu kreiranja i recepcije medijskih sadržaja, kao i problem pojave erozije povjerenja javnosti unutar hiperrealnog medijskog okruženja.

Ključne riječi: deepfake; dezinformacije; lažne vijesti; erozija povjerenja; kredibilno novinarstvo

UVOD

U vremenima obmane, istina postaje revolucionarni akt.
(George Orwell)

Uprkos činjenici da manipulacija diseminacijom poruka unutar javnog diskursa ne predstavlja nov fenomen već svojevrsnu medijsko-društveno-historijsku kategoriju,

čini se da proces širenja dezinformacija nikada nije uzeo toliko maha kao danas. To je očito do te mjere da neki od savremenih teoretičara masovnih komunikacija poput Livingston i Bennett (2020) i Keyes (2004), recentno doba ne označavaju više kao informacijsko, već *ex adverso* kao doba dezinformacija ili doba post-istine. To se s jedne strane može pripisati bržem i jednostavnijem načinu distribucije informacija bez prethodne temeljite verifikacije sadržaja, dominaciji društvenih mreža, ali i sve značajnijem medijsko-tehnološkom progresu podržanom dostignućima na polju umjetne inteligencije. Jedan od najeklatantnijih primjera sinteze tehnoloških dostignuća u domenu umjetne inteligencije i kompjuterskih algoritama, generiranih iz velike baze internetski dostupnih biometrijskih podataka, jeste hiperrealistična forma videa nazvana deepfake. Zasnovan na najsofisticiranijoj tehnologiji, deepfake zasigurno predstavlja svojevrsnu prekretnicu u produciranju lažnih audio-vizuelnih sadržaja. Samim tim i izazov njegove identifikacije i dekonstrukcije postaje sve veći, kako za recipijente, tako i za same novinare.

Ne obilježavajući *a priori* deepfake fenomen kao malicioznu pojavu i ne umanjujući njegov zabavno-edukacijski potencijal, primarni cilj i naglasak ovog rada bit će propitati i poliperspektivno ukazati na njegove negativne reperkusije, a poglavito dugoročne implikacije na profesionalno novinarstvo, njegove postulate i njegovu društvenu funkciju.

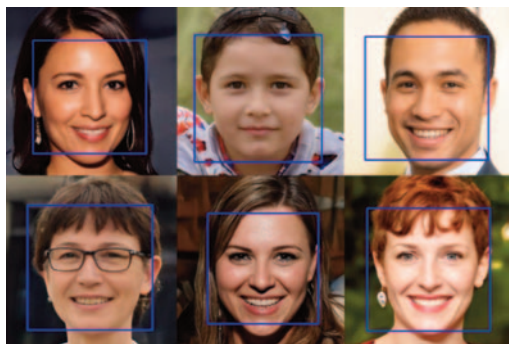
FIKCIJA ILI FAKCIJA: GRANICA IZMEĐU LAŽNOG I ISTINITOG U DANAŠNJEM (DEZ)INFORMACIJSKOM SISTEMU

Forma hiperrealističnog videa – deepfake – terminološki predstavlja složenicu nastalu spajanjem sintagme „deep learning“¹ i lekseme „fake“, a naziv potiče od anonimnog korisnika internetske platforme Reddit. Prema Calderi (2020) fenomen deepfake se može definirati kao hiperrealistična digitalna falsifikacija foto, audio i video formata. Jednostavno rečeno, posrijedi je vještački konstruiran video koji izrazito realistično prikazuje nešto što se nikada nije dogodilo manipuliranjem prethodno postojećih snimaka ili fotografija. Već uvodna kontekstualizacija problema deepfakea iziskuje da se barem ukratko ukaže na pozadinu nastanka i njegovog kvantitativno sve zastupljenijeg produciranja. Teoretičari poput Chesney i Citron definiraju ga kao „naprednu, visoko realističnu formu lažnih vijesti, omogućenu ponajprije rapidnim napretkom na polju umjetne inteligencije“ (2019: 1757). Umjetna inteligencija, kao

¹ Koncept „deep learning“ odnosi se na granu umjetne inteligencije u okviru koje softver uči kako prepoznati obrasce iz podataka. Softver počiva na realističnom modelu oponašanja ljudskog mozga i načina na koji on koristi neurone za razmišljanje. Vidjeti Hof (2013).

koncept čiji se društveni utjecaj u posljednjih nekoliko decenija sve više problematizira, u značajnoj mjeri počiva na velikoj bazi dostupnih podataka. Međutim, ono što se nerijetko zanemaruje jeste činjenica da je većina konzumenata Interneta, neposredno ili posredno, svjesno ili nesvjesno, involvirana u ovaj kompleksni proces. Naime, mnogo je besplatnih aplikacija koje omogućavaju plasiranje vlastitih lica kako u tijela drugih ljudi, tako i u različite ambijente, scene iz stotina filmova i slično. S tim u vezi, esencijalni problem je u tome da prosječni konzumenti ovakvih sadržaja i brojnih aplikacija nerijetko nisu ni svjesni činjenice da svoje privatne biometrijske podatke, svoja lica, a i glasove, javno plasiraju pristajući time i na njihovu zloupotrebu u različite svrhe. Deepfake upravo djeluje na principu prethodno prikupljenih podataka i umjetne inteligencije i to posredstvom dostupne baze podataka.

Do koje mjere je napredovala umjetna inteligencija pokazuju sljedeće fotografije, konstruirane kompjuterskom integracijom dostupnih biometrijskih podataka. Na njima su lica ljudi koji uopće ne postoje.



Slika 1. Fotografije fiktivnih ljudskih figura, nastale kao rezultat integracije umjetne inteligencije i kompjuterskih algoritama (izvor: <https://thispersondoesnotexist.com/>, pristupljeno 15. 3. 2021.)

Potrebno je ukazati i na distinkciju između spomenutih aplikacija i sofisticiranog deepfakea, a koja deepfake čini tako moćnim sredstvom manipulacije i diseminacije dezinformacija. Ona se, naime, sastoji u tome da je deepfake skoro nemoguće detektirati ljudskim okom. Deepfake se „odnosi na algoritamsku metodu koja se koristi za zamjenu stvarne osobe u videozapisu drugom osobom. Uprkos ovom procesu, videozapis izgleda poput originala“ (Temir 2020: 1012). Gotovo je aksiomatski očito da upravo odatle proizlazi i najveća prijetnja odnosno rizik glede ovog tehnološkog

fenomena u smislu utjecaja na individue, pa i društvo u cijelosti. Riječ je zapravo o nemogućnosti ljudske opservacije i dekonstrukcije ovog naprednog tipa lažnog sadržaja kao sprega vizuelne i auditivne komponente.

Tragajući za odgovorom na pitanje zašto deepfake postaje široko prihvaćen kod konzumenata Interneta i internetskih socijalnih mreža i platformi i iz čega proizlazi njegova najveća opasnost, identificirali smo tri faktora: 1) *uvjerljiva vizualizacija i auditivna podrška*, 2) *sveopća dostupnost*, 3) *nemogućnost povlačenja jasne granice između istinitog i lažnog*. „U rasponu od oko godinu dana, deepfake je napredovao do te mjere da se gotovo ne razlikuje se od autentičnih videozapisa. Koristeći mješavinu umjetne inteligencije i mašinskog učenja, tehnologija koja stoji iza njega samo će nastaviti napredovati“ (Caldera 2020: 178).

Vizuelno-kognitivni sistem ljudskoga mozga veoma jednostavno može postati predmetom manipulacije budući da čovjek u pravilu vjeruje onome što vidi, a primjer za to su optičke iluzije, slike s više figura, te mađioničarski trikovi koji nam pokazuju u kojoj to mjeri zapravo „vjerujemo svojim očima“. Ne treba zanemariti ni podatak da je posljednjih nekoliko decenija uočen pad povjerenja recipijenata u kredibilnost fotografije, shodno brojnim naprednim mogućnostima efektivne montaže (Westling 2019). S druge strane, prema Kietzmann et al. (2020), poznati glasovi kojima ljudi vjeruju, te sinteza vizuelnog i audio segmenta predstavljaju najznačajniju dimenziju koja deepfake diferencira od predašnjih dezinformacijskih formata u smislu uvjerljivosti i intenziteta utjecaja na publiku. Drugim riječima, deepfake naročito opasnim čini to što na implicitan način iskorištava prirodnu ljudsku tendenciju oslanjanja na opažanje posredstvom čula vida i sluha. Sveopća dostupnost, kao i mogućnost jednostavnog i uglavnom besplatnog kreiranja ovakvih sadržaja, također su bitan činilac široke rasprostranjenosti, ali i prihvaćenosti deepfakea prilikom recepcije. Sve prethodno rečeno vodi nas ka središnjem pitanju: *Šta se dešava u doba kada više ne možemo vjerovati ni onome što vidimo i čujemo? Šta se dešava kada se fikcija i faktografija sve snažnije isprepliću podrivajući uporedo s tim procesom i mogućnost kritičke valorizacije i empirijskog razlikovanja virtuelnog od stvarnog, laži od istine?*

Chesney i Citron (2019) zastupaju tezu da deepfake može biti posebno destruktivan budući da dolazi u trenutku kada je već teško razdvojiti činjenice od fikcije. Čini se da tehnološki progres pobuđuje i dobro i loše u ljudima, tjerajući nas naprijed, ali istovremeno i vraćajući nas unazad. U tom smislu deepfake nije izuzetak, baš kao i ostali tehnološki novumi imat će kako svijetlu, tako i tamnu stranu svoje pojavnosti (v. Baccarella et al. 2018). Kada govorimo o pozitivnoj dimenziji složit

ćemo se sa Chesney i Citron (2019) da deepfake može biti izuzetno koristan u domenu edukacije, umjetnosti, kao i zabavne industrije. Uistinu, on može poslužiti kao alat u pripremi edukacijskog materijala, a njegova upotreba u polju umjetnosti već je pokazala djelotvornost. Prema Chandleru (2020), koji pozitivnu stranu ovog novuma ističe u prvi plan, zahvaljujući deepfakeu ljudima će biti pružena nesvakidašnja mogućnost da vide stvari koje više ne postoje ili nikada nisu postojale. No, nipošto se ne smije zanemariti proučavanje i anticipiranje njegovih negativnih društvenih konsekvenci, posebice uzimajući u obzir to da je uz ovu naprednu tehnologiju iz malo bolje opremljenog studija moguće inicirati probleme i nesporazume svjetskih razmjera. Sposobnost stvaranja uvjerljivih deepfake formata zasigurno neće ostati u rukama tehnološki sofisticiranih ili odgovornih aktera (Vidjeti Dunaway 2018). Opasnost se u prvom redu ogleda i u činjenici da se ovakve forme dezinformacija i lažnih sadržaja u okviru internetskog i generalno novomedijskog diskursa uglavnom predstavljaju, pa shodno tome i percipiraju, kao zabavni formati, što je upravo prostor pogodan za manipulaciju i maskiranje potencijalno negativnog utjecaja, naročito kada se ima u vidu da je riječ o iznimno uvjerljivim audio-vizuelnim predstavama kakvi su deepfake sadržaji. U novomedijskom cyber-okruženju, obilježenom sveopćom dostupnošću i prezasićenošću informacijama koje zadobijaju viralnu dimenziju, deepfake, takav kakav jeste, može, van svake sumnje, izazvati višestruke negativne, potencijalno i katastrofalne učinke.

DIHOTOMNI PRISTUP DRUŠTVENIM REPERKUSIJAMA DEEPPFAKE FENOMENA

„Deepfake nova je dimenzija problema povezanih s lažnim vijestima, a kao takva mogla bi predstavljati ozbiljan problem za nacionalnu sigurnost. Pogotovo sa širenjem društvenih mreža, broj i učinak lažnih vijesti značajno se povećao“ (Beridze, Butcher 2019: 332). O društvenim medijima govorimo kao o najčešćem kanalu i najosjetljivijem okruženju za cirkulaciju i distribuciju deepfake sadržaja. Razlog tome je prevažno dereguliranost ovog područja, kao i brzina diseminacije informacija bez prethodne verifikacije i valorizacije određene vijesti. Brown (2020) zastupa tezu prema kojoj je i sam potencijal brze viralne distribucije ovakve forme lažnih vijesti zabrinjavajući koliko i deepfake *eo ipso*. Razlog tome vidi prvenstveno u činjenici da ne postoji mogućnost promptne tehnološke dekonstrukcije deepfakea. Posljedično, čak i u situacijama kada njegova autentičnost nakon izvjesnog vremena biva opovrgnuta, šteta koja je već učinjena njegovom diseminacijom može biti

ireverzibilna. Ovo se svakako odnosi i na starije oblike diseminacije dezinformacija, poput klevete i manipulacije fotografijom. Bez obzira na regulaciju putem demantija, šteta je učinjena, no situacija s deepfakeom još je složenija uzmemo li u obzir uvjerljivu audio-vizuelnu predstavu posredstvom koje je određena dezinformacija distribuirana. Mjereći utjecaj deepfakea na političke stavove, u okviru eksperimenta koji su proveli Dobber et al. (2021) na uzorku od 278 sudionika, samo 12 njih je prepoznalo da je riječ o vještački produciranom videu, a skoro 75% učesnika nikada nije čulo za deepfake.

Osim uvjerljive vizuelizacije, nemogućnosti razlikovanja istinitog od lažnog i sveopće dostupnosti, o čemu je prethodno bilo riječi, prihvatanje deepfakea se usložnjava i u kontekstu interakcije s kognitivnom pristrasnošću recipijenata. Iz psihološkog rakursa promatrano, kao individue smo predisponirani da bez propitivanja prihvatamo nove informacije koje koreliraju s našim prethodnim uvjerenjima, predrasudama ili stavovima. „Pojedini gledaoci deepfakea mogu poduzeti akciju kao odgovor na lažnu informaciju prije nego ona bude dekonstruirana; drugi bi čak mogli odbiti dokaze da je snimak lažan, osobito ako njen sadržaj afirmira već postojeće predrasude ili pretpostavke“ (Brown 2020: 10).² K tome, ljudi su skloniji prihvatiti informacije kojima su konzistentno izloženi. Hambrick i Marquardt (2018) razmatrajući kognitivnu prijemčivost za lažne vijesti, navode da će recipijenti vjerovatnije prihvatiti lažnu informaciju kao istinitu pri ponavljajućem susretu s njom, čak i onda kada su inicijalno te informacije odbacivali kao lažne. Nedvojbeno je da će navedena psihološka konstitucija doprinositi kako povjerenju konzumenata prema takvim sadržajima, tako i procesu *peer-to-peer* širenja lažnih vijesti, čak i u slučaju da one nakon izvjesnog vremena budu dekonstruirane i demantirane. „Problem nije samo u tome što ova tehnologija postoji i neprestano se poboljšava, već upravo u tome što je jednom stvoren deepfake lako distribuirati, a teško iskorijeniti“ (Brown 2020: 16).

Potencijal digitalnih medija, podržan rapidnim tehnološkim dostignućima, da transformiraju i afektiraju društvenu stvarnost, značajno se odmaknuvši od tradicionalne uloge medija, predmetom je teorijskog promišljanja mnogih savremenih mislilaca. Karnouskos primjerice navodi: „Takve tehnološke mogućnosti imaju potencijal da preoblikuju digitalne medije, dok bi društvene implikacije također mogle biti ozbiljne jer potkopavaju povjerenje javnosti u ono što se vidi, čuje i na kraju vjeruje da je istina“ (2020: 138). Razmatrajući potencijalne posljedice deepfa-

² Sintagma „poduzeti akciju“ odnosi se na potencijalno djelovanje, odnosno radnju koju pojedinac poduzima u skladu sa sadržajem dezinformacije diseminirane putem deepfake videa. Primjerice, ukoliko je izvjesni politički akter predstavljen korumpiranim, to može utjecati na glasačko ponašanje individua izloženih navedenom deepfakeu.

kea, uočavamo svojevrstu dihotomiju reflektiranu kroz dvije implikacije dugoročnog utjecaja ovog tehnološko-medijskog novuma na medijski diskurs i društveni realitet³:

1. Reperkusije koje prirodno proizlaze iz koncepta deepfakea *per se*, odnosno mogućnost fabrikovanja potpuno imaginarnih događaja i situacija, zloupotrebljavajući prethodno prikupljene biometrijske podatke, pogotovo svjetski poznatih i utjecajnih osoba. Jasno je da implikacije široke prihvaćenosti i povjerenja u ovakve forme mogu rezultirati teškim posljedicama, kako unutar individualnog (narušavanje privatnog digniteta, krađa identiteta i dr.), tako i društvenog života. Propitujući potencijal deepfakea da afektira društvenu zajednicu, Chesney i Citron (2018) navode da bi dobro tempiran i osmišljen deepfake mogao odlučiti o rezultatima izbora, prouzročiti štetu za demokratski diskurs, izazvati nasilje u gradovima u kojima tinjaju građanski nemiri te pojačati postojeće ili čak stvoriti nove antagonizme i političke podjele. U tom smislu, jedna od posljedica može biti i povratak čovječanstva u stanje *bellum omnium contra omnes*, koje će sada biti podržano virtuelnom, bodrijarovski kazano, hiperrealističnom stvarnosti, unutar novomedijske paradigme.

2. Suprotno navedenom, oprečna stručna mišljenja zagovaraju stajalište prema kojem fenomen deepfakea ne predstavlja društveni rizik *per se*. Naime, sama upoznatost recipijenata s postojanjem fenomena deepfake ostavlja prostor za rezervu, u mjeri i na način da se čak i empirijska utemeljenost nekog događaja/izjave pripiše tehnološkoj konstrukciji ili zloupotrebi. Ali, u tom smo slučaju suočeni sa naličjem istog fenomena – stvarni događaji, afere, rizici i slične društvene pojave negativnog predznaka mogu biti negirane unutar javnog diskursa, što je podjednako opasan modalitet odnosno ishod manipulacije javnošću. Deepfake samo dodatno ohrabruje ljude da negiraju stvarne događaje stvarajući vlastite subjektivne realnosti koje su po mjeri njihove kognitivne pristrasnosti. U najkraćem, nerazlikovanje autentičnih od fabrikovanih videozapisa rezultat će mogućnošću da bilo koji oblik videozapisa bude klasificiran kao „lažna vijest“ (Caldera 2020: 178). Deepfake će se tako pokazati korisnim instrumentom u izbjegavanju istine i odgovornosti na jednako poguban način. Paradoksalno, akteri kojima je cilj izbjeći odgovornost za svoje stvarne riječi i djela postajat će vjerodostojniji kako javnost bude u većoj mjeri educirana o prijetnjama koje deepfake predstavlja sam po sebi.

³ Prema istraživanju provedenom u okviru izrade master teze *Uloga medija u konstrukciji društva rizika* (Vatreš 2020), društvene implikacije deepfake fenomena moguće je promatrati u okviru dvije distinktivne dimenzije.

Sve navedeno predstavlja i jedan od najvećih izazova za profesionalno novinarstvo, koje mora pronaći djelotvorne odgovore ireverzibilnom podripanju njegove društvene funkcije. „Uz mogućnost da lažno predstavi stvarnim, deepfake čini lakšim tvrditi da je i ono stvarno zapravo lažno. To može uzrokovati beskrajne probleme u politici, zakonu i društvenom životu“ (Temir 2020: 1015).

DEEFAKE KAO PRIJETNJA KREDIBILNOM NOVINARSTVU

Fenomen deepfake neki autori jednostavno definiraju kao „novi instrument za stari problem“ – dezinformacije (Lewis 2019). Shodno realnoj mogućnosti da se deepfake masovno producira i diseminira, prema Calderi (2020) krajnji rezultat mogao bi se očitovati i u svojevrsnoj „dezinformacijskoj oluji“. Kao što smo prethodno istakli, *differentia specifica* ove forme lažnih vijesti jeste sinteza audio i vizuelne komponente, čime se na efektivan način manipulira prirodnom tendencijom čovjeka da vjeruje onome što vidi i čuje.

U cilju ilustracije snage ove forme fabrikovanih videa, poslužiti ćemo se jednim od primjera deepfakea koji je izazvao pažnju medija i javnosti na globalnom nivou. Riječ je o izrazito realističnom deepfakeu koji manipulira figurom bivšeg predsjednika SAD-a Baraka Obame. Sjedeći ispred američke zastave, Obama u hiperrealističnom videu izgovara: „Ulazimo u eru kada naši neprijatelji mogu učiniti da bilo ko govori bilo šta u bilo kojem trenutku, iako nikada ne bi rekli takve stvari... Primjerice, ja koji govorim da je predsjednik Tramp totalni idiot“.⁴ U trideset i šestoj sekundi videa, zaslona se dijeli i postaje očigledno da zapravo nije riječ o Obami. Ovaj deepfake konstruirao je američki oskarovac i komičar Jordan Peele koji je želio ukazati na prijetnju i nesaglediv društveni rizik koju digitalna falsifikacija video formata uistinu može prouzrokovati. Koristeći umjetnu inteligenciju, manipulacijom prethodno dostupnih Obaminih izjava, stvorio je posve novi video u kome on izgovara ono što u stvarnosti nikada nije, odnosno ne bi rekao. Fabrikovana izjava utjecajne osobe, poput predsjednika SAD-a, iskorištena u maliciozne svrhe, u određenim uslovima socio-emocionalne klime može dovesti i do tragičnih, katastrofalnih ishoda. Rizici su u tom kontekstu svakako mnogostruki i, onkraj svake sklonosti za paranoične scenarije, sežu od difamacije određenih individua pa sve do stvaranja antagonizama i uzrokovanja sukoba koji mogu poprimiti globalne razmjere.

⁴ Izvor: BuzzFeedVideo, „You Won't Believe What Obama Says in This Video!“, YouTube (april 17, 2018) <https://www.youtube.com/watch?v=cQ54GDm1eL0>, pristupljeno 25. 3. 2021. U trenutku pisanja rada ovaj deepfake pregledan je više od 8 miliona puta, a emitirali su ga i medijski giganti poput BBC-a.



Slika 2. Uvjerljivi deepfake Baraka Obame iz 2018. godine (izvor: youtube.com, pristupljeno 25. 3. 2021.)

Međutim, ukoliko deepfake postane popularan unutar mainstreama, snaga i kredibilnost stvarnih videa neosporno će devalvirati. Valja pritom ukazati da je video format, sve do sada, bio relativno kredibilan izvor informacija, kako za recipijente, tako i za same novinare (Foer 2018). Foer (ibid.) ističe kako je videozapis kao forma vijesti prethodno stekao veliki autoritet te shodno tome i najveće povjerenje recipijenata unutar medijskog diskursa. Nemogućnost oslanjanja na video format vijesti nedvosmisleno bi prouzrokovala veliku štetu medijskom sistemu, a ništa manje značajno i novinarstvu kao profesiji. Valja podsjetiti i na to da su se različite mogućnosti manipulacije audio-vizuelnim tehnikama primjenjivale i prije opisanih novih fenomena, međutim njihova sofisticiranost i uvjerljivost nije bila ni blizu današnjoj. S deepfakeom i širenjem njegove popularnosti, te upoznatosti publike s mogućnošću realistične digitalne falsifikacije video formata, vrijednost bilo kojeg videozapisa, stvarnog ili lažnog, srazmjerno će opadati. Sve to implicira povećanje nesigurnosti i slabljenja povjerenja javnosti u različite sadržaje i vijesti. Erozija povjerenja može poprimiti takve razmjere da vidjeti više neće značiti i vjerovati.

„Kako deepfake postaje rasprostranjen, javnost može imati poteškoće povjerovati u ono što im govore njihove oči ili uši – čak i kad su informacije stvarne. Zauzvrat, širenje deepfakea prijeti padom povjerenja potrebnog za učinkovito funkcioniranje demokratije“ (Chesney i Citron 2019: 1786). Pored svih navedenih negativnih društvenih implikacija, koje su detaljnije razmatrane u prethodnom poglavlju, u ovom segmentu rada posebno ćemo se osvrnuti upravo na pojavu erozije povjerenja recipijenata. „Najveća prijetnja pojavljuje se kada građani počinju razumijevati da tehnologija omogućava stvaranje realističnog videa i njegovo korištenje u nezakonite svrhe. Povjerenje javnosti može pasti kada ne možemo vjerovati onome što vidimo“ (Brown 2020: 11). U kompleksnoj istraživačkoj studiji koju su sprovedli Vaccari i

Chadwick (2020) na uzorku od 2.005 ispitanika uočeno je da nesigurnost u vijesti na društvenim mrežama predstavlja kvantitativno zastupljeniju kategoriju, negoli sama manipulacija i obmana recipijenata dezinformacijama prezentiranim posredstvom politički motiviranog deepfakea. Naime, nakon emitirane četiri sekunde deepfakea, ispitanici su trebali dati odgovor na pitanje da li je riječ o stvarnom ili fabrikovanom videu, a čak 35,1% njih odgovorilo je da „nisu sigurni“. Nakon dvadeset i šest sekundi istog videa, taj procenat porastao je na 36,9%. Evaluirajući nalaze navedenog istraživanja, ističu: „Politički deepfake ne mora nužno prevariti individue, ali može povećati nesigurnost, što zauzvrat smanjuje povjerenje u vijesti na društvenim mrežama“ (Vaccari i Chadwick 2020). Na tom tragu, Zimmermann i Kohring (2020) su došli do zaključka da što manje ljudi vjeruju vijestima medija, veća je vjerovatnoća da će biti podložni dezinformacijama, što zauzvrat može utjecati na njihovo glasačko ponašanje.

Prema Brown (2020), erozija povjerenja recipijenata u vijesti, ali i medijske organizacije u cjelini, manifestirat će se na dva distinktivna, ali podjednako opasna, načina. Najprije, ohrabrit će ljude da osporavaju stvarne događaje zabilježene videozapisom kako bi sve intenzivnije producirali subjektivne realnosti. Sukladno tome, fabrikovanje vijesti ubitnome će doprinijeti snaženju već postojećih iskrivljenih uvjerenja, kao i različitih teorija zavjera uvećavajući masu njihovih sljedbenika. Kao argument u prilog ovome Chesney i Citron (2019) navode da u okruženju gdje snažni narativi o nepovjerenju javnosti u medije već postoje, provokativni deepfake nalazi dobar fundament i pripremljenu publiku. Drugi modus manifestacije erozije povjerenja javnosti očituje se u tome da građani u potrazi za istinom neće moći prepoznati koji su video dokazi kredibilni, smatra Brown. Ukoliko recipijenti ne mogu samostalno razlikovati autentične videozapise od intencionalno konstruiranih, oni postaju skloni nevjerovanju bilo kakvim oblicima video dokaza, makar da su prezentirani kao vijesti ili čak kao dokaz u sudnici. Kombinacija pada povjerenja i propadanja vrijednosti objektivne istine u značajnoj mjeri devalvira kako novinarstvo, tako i samu demokratiju u njenom fundamentalnom značenju. U tom kontekstu, Chesney i Citron ističu sljedeće:

„Kako se širi kapacitet proizvodnje deepfakea, novinari će se sve češće susretati s dilemom: kada neko pruži video ili audio dokaz o značajnom događaju, može li se vjerovati u njegovu autentičnost? To nije novo pitanje, ali će odgovor na njega biti sve teže naći kako se deepfake više širi. Medijske organizacije mogu odustati od brzog izvještavanja o stvarnim, uznemirujućim događajima iz straha da će se dokazi o njima ispostaviti lažnim.“ (2019: 1784)

Neosporno je da medijsko izvještavanje u realnom vremenu ostvaruje veliki utjecaj na društvo. Međutim, uzimajući u obzir činjenicu da danas i ova vrsta sadržaja može potpasti pod kategoriju deepfake, medijske organizacije i novinari nalaze se pred velikim izazovom. U opisanom informacijskom okruženju medijski subjekti mogu odustati od preuzimanja takve vrste društvenog rizika. S tim u vezi, u nedostatku validnih i brzih načina provjere autentičnosti audio i video izvora, novinarstvo će sve teže ispunjavati etičku i moralnu obavezu istinitog i društveno odgovornog izvještavanja, koja je još i prije pojave društvenih mreža i naprednih formi dezinformacija dijelom narušena kao rezultat djelovanja različitih političkih, ideoloških, ekonomskih i drugih faktora. Shodno rastu nepovjerenja u vijesti, pa u krajnoj liniji i u medije u cjelini, čak i stvarne značajne priče i društveno relevantni događaji ostvarivat će sve manji utjecaj na recipijente. Deepfake predstavlja „glavnu prijetnju našem društvu, političkom sistemu i biznisu, te vrši pritisak na novinare koji se trude filtrirati istinite informacije od lažnih“ (Westerlund 2019: 42). Konačno, prema Temiru (2020), navedeni izazovi dovode problem povjerenja i brzine do tačke koja nagovještava i sam kraj kredibilnog novinarstva. U tom smislu, Chesney i Citron (2018) izražavaju strah da ćemo iz doba lažnih vijesti veoma brzo preći u doba dubokih lažnih vijesti (*deep fake news*).

Prije nego je pojava i ekspanzija Interneta rezultirala demokratizacijom širenja informacija, ljudi su se u velikoj mjeri oslanjali na tradicionalne oblike elektronskih i štampanih medija pri recepciji vijesti i informacija. Tradicionalni masovni mediji u demokratskim liberalnim društvima primjenjivali su mehanizme verifikacije informacija, što u novomedijskom diskursu nerijetko nije slučaj. „Kada korisnici konzumiraju informacije iz netradicionalnih izvora, kao što su stranice na društvenim mrežama, teret verifikacije sadržaja prenosi se s novinara na njih“ (Brown 2020: 16). Napredni cyber obrasci širenja dezinformacija neuporedivo su efektivniji i dodatno usložnjavaju situaciju iz temelja mijenjajući konvencionalnu društvenu ulogu novinara kao „čuvara kapije“. Dodatni problem sadržan je u činjenici da korištenje drugih online medija kao kredibilnih izvora informacija, postaje dominantna praksa u okviru internetske medijske paradigme. Prema Brautović et al. (2020) novinari na taj način preuzimaju i doprinose širenju grešaka u izvještavanju drugih medija. Rezultati njihove analize sadržaja online medija pokazuju da je čak 69% analiziranih tekstova imalo određenu grešku, a temeljni uzrok tome pronađen je u korištenju sekundarnih umjesto primarnih izvora. U skladu s navedenim procesom tačnost i preciznost u izvještavanju nedvojbeno opada. Dodamo li tome nemogućnost brze i vjerodostojne detekcije i dekonstrukcije deepfakea, kao najnaprednije forme lažnih vijesti, situacija

u novomedijskom sistemu biva još kompleksnijom, a novinarska načela tačnosti, objektivnosti i preciznosti sve su teže ostvariva.

Obilježavajući koncept lažnih vijesti, deepfake i njima srodne kategorije kao destruktivne socijalne fenomene i najopasnije informacijsko oružje našeg vremena, Chudinov et al. (2019) navode kako deepfake narušava čak i krucijalni princip novinarstva – „nemoguće je prikazati ono što ne postoji“ (2019: 1851).

ZAKLJUČNA RAZMATRANJA

Iako je riječ o relativno novom fenomenu u epohi izrazito novomedijski obilježene stvarnosti, u posljednjih nekoliko godina evidentan je porast interdisciplinarnog istraživačkog interesa za izučavanje deepfakea i njegovih društvenih reperkusija. Pozivajući se na Westerlund (2019), konstatujemo da dosadašnji naučni doprinosi identifikacije i dekonstrukcije sugeriraju da je borba protiv ovog tehnološkog novuma izuzetno kompleksna, ali hipotetički moguća u okviru sljedećih domena i područja djelovanja: 1) zakonodavstvo i regulacija; 2) korporativna politika i dobrovoljno djelovanje; 3) edukacija i trening; 4) anti-deepfake tehnologija koja uključuje detekciju deepfakea, autentikaciju sadržaja, kao i sprečavanje produciranja deepfakea.

U nedostatku dugoročnih pragmatičnih rješenja kada je riječ o regulatornom pristupu, kao svojevrsni imperativ u ovom trenutku nameće se poliperspektivno razumijevanje pozicije i reperkusija deepfakea. Ovim radom nastojali smo pružiti teorijski uvid u suštinu deepfakea, agense njegove prihvaćenosti, kao i njegovu dihotomnu socijalnu percepciju, uz poseban naglasak na implikacije po tradicionalnu društvenu ulogu novinarstva. Prvi korak u borbi protiv maliciozno motiviranog deepfakea je upravo edukacijsko ukazivanje na činjenicu da problem nije isključivo tehnološkog, već općedruštvenog karaktera. Kao „posljednje oružje u borbi protiv istine“ (Brown 2020) deepfake, orvelovski, potvrđuje da u vremenima obmane istina i njeno prepoznavanje uistinu postaju revolucionarni akti.

LITERATURA

1. Bacarella, Christian, Timm F. Wagner, Jan Kietzmann, Ian Paul McCarthy (2018), "Social Media? It's Serious! Understanding the Dark Side of Social Media", *European Management Journal*, 36(4), 431-438.
2. Bennett, Lance W., Steven Livingston (2020), *The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States*, Cambridge University Press, Cambridge

3. Beridze, Irakli, James Butcher (2019), "When Seeing is No Longer Believing", *Nature Machine Intelligence*, 1(8), 332-334.
4. Brautović, Mato, Sandra Buratović Maštrapa, Romana John (2020), "Točnost u online medijima: Nedovoljne novinarske rutine u provjeravanju činjenica i ispravcima", *Medijske studije*, 11(21), 66-86.
5. Brown, Nina (2020), "Deepfakes and the Weaponization of Disinformation", *Virginia Journal of Law & Techology*, 23(1), 1-59.
6. Caldera, Elizabeth (2020), "Reject the Evidence of Your Eyes and Ears: Deepfake and the Law of Virtual Replicants", *Seton Hall Law Review*, 50(1), 177-205.
7. Chandler, Simon (2020), "Why Deepfakes Are A Net Positive For Humanity", *Forbes*, izvor: <https://www.forbes.com/sites/simonchandler/2020/03/09/why-deepfakes-are-a-net-positive-for-humanity/#3f950d2a2f84>, pristupljeno 17. 3. 2021.
8. Chesney, Robert, Danielle Citron (2019), "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, 107, 1753-1820.
9. Chesney, Robert, Danielle Citron (2018), "Disinformation on Steroids: The Threat of Deep Fakes", Boston University School of Law, izvor: https://scholarship.law.bu.edu/shorter_works/30, pristupljeno 16. 2. 2021.
10. Chudinov, Anatoly P., Natalya, N. Koshkarova, Natalia B. Ruzhentseva (2019), "Linguistic Interpretation of Russian Political Agenda Through Fake, Deepfake, Post-Truth", *Journal of Siberian Federal University Humanities & Social Sciences*, 12(10), 1840-1853.
11. Dobber, Tom, Nadia Metoui, Damian Trilling, Natali Helberger, Claes de Vreese (2021), "Do Microtargeted Deepfakes Have Real Effects on Political Attitudes?", *The International Journal of Press/Politics*, 26(1), 69-91.
12. Dunaway, Jaime (2018), "Reddit (Finally) Bans Deepfake Communities, but Face-Swapping Porn Isn't Going Anywhere", *Slate*, izvor: <https://slate.com/technology/2018/02/reddit-finally-bans-deepfake-communities-but-face-swapping-porn-isnt-going-anywhere.html>, pristupljeno 12. 3. 2021.
13. Foer, Franklin (2018), "The Era of Fake Video Begins", *Atlantic*, izvor: atlantic.com/magazine/archive/2018/05/realitys-end/556877/, pristupljeno 27. 4. 2021.
14. Hall, Holly Kathleen (2018), "Deepfake Videos: When Seeing Isn't Believing", *Catholic Journal of Law and Technology*, 27(1), 51-76.

15. Hambrick, David, Madeline Marquardt (2018), "Cognitive Ability and Vulnerability to Fake News", *SCI. AM*, izvor: <https://www.scientificamerican.com/article/cognitive-ability-andvulnerability-to-fake-news/>, pristupljeno 17. 2. 2021.
16. Hof, Robert D. (2013), "Deep Learning", *MIT TECH. REV.* izvor: <https://www.technologyreview.com/technology/deep-learning/>, pristupljeno 4. 4. 2021.
17. Karnouskos, Stamatias (2020), "Artificial Intelligence in Digital Media: The Era od Deepfakes", *IEEE Transactions on Technology and Society*, 1(3), 138-147.
18. Keyes, Ralph (2004), *The Post-Truth Era: Dishonesty and Deception in Contemporary Life*, St. Martin's Publishing Group
19. Kietzmann, Jan, Linda W. Lee, Ian Paul McCarthy, Jan Kietzmann (2020), "Deepfakes: Trick or Treat?", *Business Horizons*, 63(2), 135-146.
20. Lewis, James Andrew (2019), "Trust Your Eyes? Deepfakes Policy Brief", Center for Strategic & International Studies, Washington D. C., izvor: <https://www.csis.org/analysis/trust-youreyes-deepfakes-policy-brief>, pristupljeno 22. 2. 2021.
21. Temir, Erkam (2020), "Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism", *SELÇUK İLETİŞİM DERGİSİ*, 13(2), 1009-1024.
22. Vaccari, Cristian, Andrew Chadwick (2020), "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News", *Social Media and Society*, 6(1), 1-13.
23. Vatreš, Amina (2020), *Uloga medija u konstrukciji društva rizika*, magistarski rad, Fakultet političkih nauka Univerziteta u Sarajevu, izvor: <http://fjn.unsa.ba/b/wp-content/uploads/2021/02/ULOGA-MEDIJA-U-KONSTRUKCIJI-DRUSTVA-RIZIKA.pdf>, pristupljeno 18. 4. 2021.
24. Westerlund, Mika (2019), "The Emergence of Deepfake Technology: A Review", *Technology Innovation Management Review*, 9(11), 39-52.
25. Westling, Jeffrey (2019), "Deep fakes: Let's Not Go Off The Deep End", *Techdirt*, izvor: <https://www.techdirt.com/articles/20190128/13215341478/deep-fakes-lets-not-go-off-deep-end.shtml>, pristupljeno 21.1.2021.
26. Zimmermann, Fabian, Matthias Kohring (2020), „Mistrust, Disinforming News, and Vote Choice: A Panel Survey on the Origins and Consequences of Believing Disinformation in the 2017 Parliamentary Election“, *Political Communication*, 37(8), 215-237.

DEEFAKE PHENOMENON: AN ADVANCED FORM OF FAKE NEWS AND ITS IMPLICATIONS ON RELIABLE JOURNALISM

Summary:

In the era of distinctly new media-marked social reality, the process of dissemination of disinformation is becoming an increasingly common phenomenon. As one of the technological innovations that stands out in this field is the deepfake phenomenon – a hyper-realistic digital falsification of video and audio formats. Since it is based on the most sophisticated technology, supported by achievements in the field of artificial intelligence, deepfake unequivocally represents a kind of turning point in the context of production and distribution of fake audio-visual content. To poly perspective point to the multiple repercussions of this phenomenon, the paper provides a deeper insight into the very definition and the specific factors that contribute to the acceptance of deepfake, with special emphasis on its implications on reliable journalism and its social function. The text articulates the role of deepfake in the process of creating and receiving media content, as well as the problem of the emergence of erosion of public trust within a hyperreal media environment.

Keywords: deepfake; disinformation; fake news; erosion of trust; reliable journalism

Adresa autorice

Authors' address

Amina Vatreš

samostalni istraživač

Sarajevo

amina_vatres@live.com

