UDK: 004.738.5:342.721
DOI: 10.5937/PDSC24679M

Dr Darko M. Marković, vanredni profesor
Pravni fakultet za privredu i pravosuđe u Novom Sadu,
Univerzitet Privredna akademija u Novom Sadu
email: darko.markovic@fepps.edu.rs

Dr Mina Zirojević, redovni profesor
Ministarstvo prosvete, Beograd
email: mina.zirojevic@gmail.com

IZAZOVI U REGULISANJU I IDENTIFIKACIJI DEEPFAKE SADRŽAJA

Apstrakt:

Savremeni razvoj digitalnih tehnologija donosi brojne prednosti za svakodnevni život čoveka, otvarajući nove mogućnosti, proširujući i olakšavajući postojeće. Zahvaljujući tome unapređuje se ispunjavanje i privatnih i profesionalnih obaveza na načine koji dovode do povećanja efikasnosti i produktivnosti u radu. Dostupnost digitalnim bazama podataka i raspoloživost naprednih digitalnih alata i tehnika, olakšava naučna istraživanja. Postojanje pristupačnih onlajn platformi doprinosi razmeni ideja i iskustava, osavremenjavanju proizvodnih procesa, pozitivnoj transformaciji industrije zdravstvene zaštite, te napretku nauke u celini. Nažalost, istovremeno uz doprinos razvoju društva u svim segmentima života, otvaraju se i novi prostori za brojne zloupotrebe. S tim u vezi, obim i intenzitet razvoja digitalnih tehnologija umnožavaju postojeće moralne i bezbednosne izazove, proizvodeći istovremeno i nove. Tome značajno doprinosi neadekvatan razvoj odgovarajućih regulatornih odgovora. Izazovi regulisanja i identifikovanja zloupotreba digitalnih tehnologija su naročito izraženi sa razvojem veštačke inteligencije, čijim tehnikama su mogućnosti generisanja lažnih audio i video zapisa, kao i slika, povećane do neslućenih granica. Imajući to u vidu, naš cilj u ovom radu jeste da ukažemo na glavne bezbednosne izazove koje proizvodi tehnologija stvaranja deepfake sadržaja i, u skladu sa tim na ključne regulatorne izazove, kako u nacionalnim tako i u međunarodnim okvirima. Primenom metoda analize sadržaja, indukcije i dedukcije došli smo do zaključka da ukupna moć suprotstavljanja zloupotrebi i kriminalizaciji *deepfake* tehnologije, zahteva dinamičan razvoj alata za

njihovu identifikaciju, uspostavljanje odgovarajućih poslovnih standarda, razvijanje strategija suprotstavljanja, kao i hitno unapređivanje postojećih regulatornih okvira, ne isključujući potrebu za usvajanje i celovitih novih zakonskih rešenja.

Ključne reči: *digitalizacija, digitalne tehnogologije, bezbednosni izazov, pravni okvir, privatnost*

1. Uvod

Utičući na promene načina života, rada i međusobnog komuniciranja, superdinamičan razvoj digitalne tehnologije u značajnoj meri oblikuje savremeno društvo. Mogućnosti za komunikaciju, saradnju i inovacije su bitno povećane zahvaljujući širokoj upotrebi interneta i pametnih telefona, kao i platformi društvenih medija. Ovo omogućava nove oblike samoizražavanja i zajednica kao virtuelnih mesta gde se mogu razmeniti iskustva, podeliti saveti, pronaći i podeliti korisne informacije i pokrenuti razne korisne inicijative. Na taj način, digitalne tehnologije ostvaruju i značajan uticaj na kulturu i identitet.

Dakle, savremeno društvo u eri digitalne tehnologije odlikuju brojne mogućnosti kojima se unapređuju različiti aspekti života, kako na individualnom tako i na društvenom planu. Međutim, sveprisutnost digitalnih tehnologija u našim životima izaziva i brojne brige. Možda i najmanju među njima, ali ne i najmanje važnu, predstavlja promovisanje plitkih interakcija putem socijalnih mreža, tzv. eho komora ili filtriranih balona (pročišćenih mehura)[1]. Dalje, novi oblici rada i automatizacija radnih procesa imaju negativan uticaj na tržište rada, koje sve više karakteriše tendencija sklapanja kratkoročnih ugovora ili slobodnih poslova, kao i radnog angažovanja po rasporedu koji se utvrđuje u poslednjem trenutku. Najveća zabrinutost postoji u vezi sa privatnošśću, nadzorom i mogućom zloupotrebom podataka. Sve to proizvodi nove izazove i rizike, kao što su uznemiravanje na mreži, širenje dezinformacija i sajber kriminal.

Razvoj veštačke inteligencije (eng. *artificial intelligence – AI*) i tehnika mašinskog učenja omogućavaju zloupotrebu multimedijalnih sadržaja (video i audio zapisa, kao i slika), tako što se iz posstojećih originalnih izdvajaju određeni segmenti

[1] Filtrirani balon (eng. *filter bubble*) je kovanica koju je 2010. godine prvi put upotrebio Eli Pariser u svojoj istoimenoj knjizi, tvrdeći da onlajn platforme, kao što su Google, Facebook i Twitter (danas poznat kao mreža X), koriste presonalizovane filtere pomoću kojih oko korisnika stvaraju svojevrsne filtrirane balone, odnosno pročišćene mehure informacija, usmeravajući korisnike na određene konkretne resurse informacija i ograničavajući njihovu sposobnost da razvijaju sopstvena uverenja. Takvo filtiranje se vrši putem onlajn algoritama koji iz prethodnih onlajn aktivnosti korisnika selektivno pogađaju njihove sklonosti i, u skladu sa tim, koje informacije bi želeli dobiti, onemogućavajući ih da se izlože suprotstavljenim gledištima i različitim perspektivama. To je sofisticiran način usmeravanja korisnika na formiranje iskrivljenog pogleda na svet, pošto su izloženi samo informacija koje potvrđuju njihova postojeća uverenja.

i spajaju sa drugim, stvarnim ili izmišljenim, s ciljem izmene osnovnog originalnog sadržaja (promena lica, glasa, sadržaja govora…) na način da novoformirani sadržaj deluje ubedljivo kao original, koji je teško kompromitovati kao lažan. U pitanju je tzv. *deepfake* tehnologija, koja ubrzano napreduje, omogućavajući stvaranje vrlo realističnih lažnih sadržaja, posebno kada su u pitanju video i audio zapisi. Zbog mogućnosti ovakvog sofisticiranog širenja dezinformacija i manipulisanja javnim mnjenjem, pa i kriminalnih prevara, *deepfake* tehnologija donosi značajne izazove za društvo, u prvom redu etičke i bezbednosne. Tome posebno doprinose nužnost kontinuiranog i dinamičnog razvoja pouzdanih tehnika identifikacije i prepreke za uspostavljanje efikasnog regulatornog okvira.

Imajući to u vidu, naš cilj u ovom radu jeste da ukažemo na glavne etičke i bezbednosne izazove koje proizvodi tehnologija stvaranja *deepfake* sadržaja i, u skladu sa tim na ključne regulatorne izazove, kako u nacionalnim tako i u međunarodnim okvirima. Primenom metoda analize sadržaja, indukcije i dedukcijedošli smo do zaključka da ukupna moć suprotstavljanja zloupotrebi i kriminalizaciji *deepfake* tehnologije, zahteva dinamičan razvoj novih za njihovu identifikaciju, uspostavljanje odgovarajućih poslovnih standarda, razvijanje strategija suprotstavljanja, kao i hitno unapređivanje postojećih regulatornih okvira, ne isključujući potrebu za usvajanje i celovitih novih zakonskih rešenja.

2. Deepfake tehnologija – manipulacija putem algoritama veštačke inteligencije

Postoje brojni pokazatelji masovnog prisustva stanovništva planete na društvenim mrežama. Prema istraživanjima iz maja 2024. godine, govorimo o više od polovine (62,2%) planete (5,07 milijardi ljudi), a smo u prošloj godini je registrovano 259 miliona novih korisnika.[2] Zahvaljujući toj činjenici možemo sa značajnom sigurnošću tvrditi da se veliki broj ljudi širom sveta već susreo sa različitim lažiranim multimedijalnim sadržajima na društvenima mrežama. To je prosto neizbežno, jer progresija omasovljenja takvih sadržaja ima tendenciju oštrog rasta zahvaljujući razvoju veštačke inteligencije i mašinskog učenja (eng. *machine learning – ML*). Između njih ne postoji znak jednakosti već se radi o različitim konceptima u oblasti računarskih nauka koje su međusobno povezane. Veštačka inteligencija je širi koncept i obuhvata razvoj inteligentnih mašina koje se odlikuju sposobnošću izvršavanja zadataka iz domena ljudske inteligencije, poput vizuelne percepcije, prepoznavanja govora, donošenje odluka i obrade prirodnog jezika

[2] Chaffey, D. (2024, May 1). Global social media statistics research summary May 2024. *Smart Insights*. Preuzeto sa : https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

(ljudskog jezika i njegovih svojstava).[3] Mašinsko učenje je podskup veštačke inteligencije čiji algoritmi uče iz podataka poteklih iz različitih baza, tekstualnih datoteka, zvučnih i vizuelnih sadržaja. Oni koriste modele za predviđanje ili odluke na osnovu novih podataka zahvaljujući tome što su modeli matematički izrazi odnosa između ulaznih podataka i izlaznih predviđanja. Tokom procesa korišćenja podataka za podučavanje algoritama mašinskog učenja, algoritam prilagođava parametre modela s ciljem postizanja minimalne razlike između predviđenog i stvarnog izlaza.

Mnogi aspekti savremenog društva se danas pokreću primenom tehnologije mašinskog učenja. To se dešava, na primer, kada na Internetu koristimo neki od pretraživača, ili kada filtriramo sadržaje na web lokacijama za e-trgovinu, kao i kada to isto radimo pri izboru sadržaja na društvenim mrežama. Tehnički uređaji, poput digitalnih fotoaparata i pametnih telefona, takođe koriste ovu tehnologiju. Sistemi mašinskog učenja se koriste pri identifikaciji sadržaja na slikama, pri transkribovanju govora u tekst, kao i pri dovođenju u vezu korisnika sa različitim sadržajima na Internetu. Brojne inteligentne aplikacije su izgrađene na stubovima koje čine mašinsko učenje i neuronske mreže.[4] Podskup mašinskog učenja koji koristi veštačke neuronske mreže (eng. *artificial neural networks – ANNs*) s ciljem oponašanja sposobnosti ljudskog mozga da obrađuje podatke i izdvaja obrasce za donošenje odluka, naziva se dubinsko učenje (eng. *deep learning*).[5] Izdvajanje željenih karakteristika je moguće upravo zahvaljujući tome što se veštačka neuronska mreža sastoji iz međusobno povezanih čvorova organizovnah u slojevima, a svaki sloj progresivno izdvaja karakteristike višeg nivoa iz neobrađenih podataka.[6] Modeli dubokog učenja se izlažu velikim skupovima podataka i u njihovoj složenoj strukturi, korišćenjem svojih algoritama, sposobni su da obrađuju podatke sa više nivoa apstrakcije. Duboko učenje unapređuje tehnologiju prepoznavanja govora, detekcije raznih objekata i elemenata unutar njihove strukture, pa samim tim i tehnologiju vizuelnog prepoznavanja objekata. Drugim rečima, zahvaljujući dubokom učenju ostvaren je ogroman napredak u obradi vizuelnih i zvučnih sadržaja, a takođe i u sekvencijalnoj analizi podataka kao što su tekst i govor.

Sama kovanica *deepfake* je nastala i popularizovana pre sedam godina, kombinovanjem engleskih izraza *deep learning* (duboko učenje) i *fake* (lažan, falsifikat), što već samo po sebi govori o tome da *deepfake* tehnologija počiva upravo na potencijalima mašinskog učenja kroz koje je veštačka inteligencija u stanju da

[3] Joiner, I. A. (2018). *Emerging library technologies*. Chandos Publishing, p. 2.
[4] Nagy, Z. (2019). *Osnove veštačke inteligencije i mašinskog učenja*. Beograd: Kompjuter biblioteka, str. xii.
[5] Holdsworth, J., & Scapicchio, M. (2024, June 17). What is deep learning? *IBM*. Preuzeto sa: https://www.ibm.com/topics/deep-learning
[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, p. 441. DOI: 10.1038/nature14539

nauči kako da vrši promene unutar strukture nekog zapisa (foto, video, audio…). Smatra se da je razvoj ove tehnologije počeo kada je veštačka inteligencija putem računarskih modela postigla sposobnost samostalnog analiziranja takvih zapisa. Tom „osamostaljivanju" veštačke inteligencije je prethodio period od 20 godina nakon što je ljudska inteligencija napisala softver pod nazivom „Video Rewrite", uz pomoć kojeg je bilo moguće vršiti izmene vizuelnog prikaza čoveka, uključujući i mimiku usana kako bi se dobio prikaz govora, uz koje bi bio montiran i tonski zapis. Ovaj softver je dalje razvijan i to u pravcu kreiranja algoritama koji će biti sposobni da sami uče. Tokom 2014. godine je napravljen softver koji se sastoji od dve neuronske mreže (generatora i diskriminatora), GANs[7], koje se međusobno nadmeću u generisanju realističnih podataka, kao što su slike ili video zapisi, na primer. U ovaj softver je potrebno uneti određene podatke (video ili zvučni zapis), od kojih on pravi bazu podataka o određenom objektu ili osobi po određenim obrascima. Model generatora vrlo detaljnom analizom uči karakteristike tih podataka (ponašanje konkretnog objekta ili osobe) do te mere da stiče sposobnost njihovog menjanja, stvarajući novi sadržaj za koji nije lako utvrditi da je zapravo isfabrikovan. Analizu sadržaja podataka vrši diskrimintor i utvrđuje da li se radi o originalima ili o generisanim sadržajima. Cilj generatora je da novi generisani sadržaj bude toliko autentičan da diskriminator potroši bar 50% vremena verujući da se radi o originalu. Pošto ljudski um ima manje razvijenu analitičku moć, vremenski period verovanja u originalnost generisanog sadržaja je duži, na čemu se i zasnivaju očekivanja propagande. Već kod ovog softvera je reč o *deepfake* tehnologiji, koja je i u svom začetku neretko bila zloupotrebljavana. Jedan od najstarijih i verovatno najpoznatijih primera zloupotrebe ovog softvera jeste video snimak izmišljenog govora bivšeg predsednika SAD, Baraka Obame.[8] Vlasti savezne američke države Kalifornija su zbog toga 2019. godine zabranile upotrebu ovakvih softvera i tehnologija.

Duboko učenje, posebno korišćenje neuronskih mreža sa mnogo slojeva, danas dominira zahvaljujući čemu se i razvila *deepfake* tehnologija. Kako teče proces stvaranja neke imitacije originala primenom ove tehnologije? Tako što algoritmi koji se unapređuju kroz obavljanje određene aktivnosti, tzv. autoenkoderi, integrišući mašinsko učenje i veštačku inteligenciju u dubokoj mrežnoj arhitekturi prepoznaju glavnu karakteristiku određenog objekta, recimo lica, i prema postavljenim zahtevima generišu određeni *deepfake* izlaz koji imitira original (lice u našem primeru). Što je veća baza podataka na raspolaganju, moć autoenkodera da modifikuje original je takođe veća, dostižući sposobnost ciljanja vizuelnog sistema

[7] GANs – Generative Adversarial Networks (generativne suparničke mreže; negerativne kontradiktorne mreže)

[8] Može se pogledati ovde: https://www.youtube.com/watch?v=MVBe6_o4cMI

mozga za pogrešnu percepciju, poput optičkih iluzija.[9] Ljudski mozak tako postaje prevaren, zaglavljen između realnosti i iluzije, odnosno onesposobljen da razlikuje originalno i lažno.

3. Etički i bezbednosni aspekti zloupotrebe deepfake tehnologije

Kada govorimo o zloupotrebi nečega onda se podrazumeva da to nešto ima i korisnu, pozitivnu stranu. Ako bismo uzeli u razmatranje samo ulogu modela diskriminatora, mogli bismo reći da je primena ove tehnologije korisna pri analizi originalnosti određenih sadržaja. Ali, to bi bilo pravdanje uzroka posledicom. Sagledavajući potencijale ove tehnologije u celini, korisnost njene upotrebe uglavnom vezujemo za oblast zabave i medija. *Deepfake* tehnologija ima korisnu primenu u stvaranju realističnih i impresivnih iskustava u filmskoj industriji i video igrama. Tako, ova tehnologija bi se mogla koristiti za kreiranje novih predstava ili nastupa preminulih glumaca ili muzičara, čime bi se njihovim obožavaocima omogućilo da ih dožive u nekim novim ulogama. Producentima bi to, naravno, donelo novac. Takođe i potencijalno presnimavanje filmova i TV serija na različite jezike, gde bi gledaoci imali potpun i vizuelni i audio doživljaj da glumci govore upravo na maternjem jeziku gledalaca. I u oblasti marketinga, gledaocima se mogu prilagoditi personalizovani sadržaji, kao što su video snimci ili reklame, a u skladu sa njihovim interesovanjima. Ovakva rešenja su dobila čak i ime – sintetički mediji.[10] Zahvaljujući *deepfake* tehnologiji moguće je grafiku video igara učiniti realističnijom u meri u kojoj se igrači mogu osećati kao da su u stvarnom svetu.[11]

Međutim, iako potencijalne koristi od korišćenja *deepfake* tehnologije postoje, moraju se razmotriti etičke i bezbednosne implikacije takve upotrebe. Navešćemo nekoliko načina ugrožavanja etičkih principa i bezbednosti, objedinjeno, jer je teško govoriti o narušavanju etičkog integriteta bez bezbednosnih implikacija.

S obzirom na to da se *deepfake* tehnologijom stvaraju lažni sadržaji, koje je teško razlikovati, ovakav način kreiranja medijskih sadržaja omogućava širenje lažnih informacija i manipulisanje javnim mnjenjem. Svrha zloupotrebe ove tehnologije može biti različita, kako u odnosu na pojedince tako i na preduzeća, organizacije i javne institucije, uključujući i državu. Montiranjem multimedijalnih

[9] Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), p. 136. https://doi.org/10.1016/j.bushor.2019.11.006

[10] Preminger, A., & Kugler, M. B. (2024). The right of publicity can save actors from deepfake Armageddon. Berkeley Technology Law Journal, Northwestern Public Law Research Paper No. 23–52, p. 102. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4563774

[11] Thies, J., Zollhöfer, M., Nießner, M., Stamminger, M., Theobalt, C., & Nießner, M. (2019). Deferred Neural Rendering: Image Synthesis Using Learned Gradient Descent. *ACM Transactions on Graphics*, *38*(6), p. 2.

sadržaja, kojima se lažno prikazuju negativni efekti upotrebe nekih proizvoda široke potrošnje, diskredituju se njihovi proizvođači i trgovinski lanci. Putem montiranih lažnih fotografija, video snimaka i audio zapisa, kojima se neka osoba prikazuje u negativnom svetlu, može se u ozbiljnoj meri narušiti njena reputacija i na taj način ta osoba diskreditovati u očima neposrednog okruženja, obrazovne ili radne sredine, pa i javnosti u celini. Shodno tome, ovakvi sadržaji se mogu koristiti za širenje lažnih informacija i propagande na političkoj sceni, uključujući i izborni proces. Pored toga, kreiranje lažnih multimedijalnih sadržaja se vrši i radi dovođenja neke osobe u kompromitujuću situaciju radi finansijske ucene i iznude.

Tehnologija *deepfake* se može koristiti i za sajber napade. Tako, može se koristiti za tzv. pecanje (*phishing*) – kreiranjem lažnih audio ili video snimaka rukovodilaca ili kolega u radnom kolektivu, kako bi se prevarom otkrile osetljive informacije ili naveli zaposleni da kliknu na zlonamerne linkove. Osim toga, zaposleni se putem lažno kreirane poruke određenog autoriteta može navesti da izvrši uplatu na račun dobavljača, a broj računa koji bi mu bio dat za uplatu bi zapravo bio račun napadača. Na sličan način, pojedinci kao fizička lica u privatnom životu, mogu se navesti da poveruju da im se obraćaju bliske osobe za novčanu pomoć uplatom na takođe lažni račun. Zatim, putem krađe identiteta. Simulacijom glasa mogu se razbiti glasovne lozinke na ulazima u određene objekte ili prostorije, kao i pristupiti restriktivnim sadržajima u mobilnim telefonima i računarima kada se takvi uređaji „otključavaju" glasovnim komandama. Takođe, ova tehnologija se može koristiti za sajber napade širenjem zlonamernog softvera (*malware*).

Navedeni primeri zloupotrebe *deepfake* za posledicu imaju razvijanje nepoverenja u digitalni sadržaj i medije,[12] a sledstveno tome i narušavanje reputacije medija[13] i integriteta demokratskih institucija. Da bi se štetni efekti zloupotrebe ove tehnologije predupredili nužno je uspostavljanje posebnih standarda u svim oblastima njene primene, kao i iznalaženje regulatornih rešenja, kako ojačavanjem postojećih propisa tako i usvajanjem novih, što bi bio direktan zakonodavni odgovor. I sa izgrađenim pravnim okvirom sa kojim bi se sprečile ili sankcionisale zloupotrebe potencijala deepfake tehnologije, potrebno je raspolagati alatima i znanjem neophodnim za identifikovanje njenih sadržaja.

[12] Nguyen, T. T. et al. (2022). Deep Learning for Deepfakes Creation and Detection: A Survey. *Computer Vision and Image Understanding*, 223, 103525, 1-19. DOI:10.1016/j.cviu.2022.103525

[13] Citron, D. K, & Chesney, R. (2019). Deep fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review, 107*(6), p. 1774.

4. Identifikovanje deepfake sadržaja – tehnički izazovi

Ubrzan napredak *deepfake* tehnologije poslednjih godina značajno otežava razlikovanje pravih i lažnih sadržaja, posebno kada su u pitanju medijski sadržaji. Za to su u prvom redu zaslužni algoritmi dubokog učenja koje koristi ova tehnologija za generisanje sintetičkih medija, kao što su tekst, slika, video i audio zapis, u forme koje izgledaju kao autentične. Otkrivanje ovih algoritama je jedan od glavnih tehničkih izazova u razotkrivanju upotrebe *deepfake* tehnologije. Iz dosadašnjih istraživanja u ovoj oblasti, izdvaja se nekoliko različitih tehnika koje su se pokazale kao najefikasnije u otkrivanju nedoslednosti u *deepfake* sadržajima.

Kreiranja lažnog teksta i fotomontaže su najstariji oblici manipulacije originalnim sadržajima. Danas, zbog sofisticiranosti *deepfake* tehnologije, uočavanje razlika je pravi izazov, ali ne i nemoguća misija, s obzirom na to da postoji nekoliko tehnika za njihovo otkrivanje. Kod fotografija se traže nedoslednosti u osvetljenju, senkama i perspektivi, a posebno na prelazima, odnosno linijama spajanja. Takođe, pretraživanjem na internetu je moguće pronaći pravu sliku i uporediti je sa originalom. Analiza teksta ne podrazumeva samo uočavanje oblika slova i konzistentnosti teksta, već i nedoslednosti u jeziku, gramatici sintaksi. Generisan tekst putem algoritama mašinskog učenja se otkriva analizom učestalosti i distribucije određenih reči ili fraza.

Kod video zapisa se vrši analiziranje vizuelnih artefakata[14] i nedoslednosti u *deepfake* snimcima, kao što su obrasci treptanja, osvetljenja i senke, izrazi lica i pokreti glave.[15] Korišćenjem algoritama za mašinsko učenje se vrši analiza teksture, boja i oblika u strukturi vidoe snimka. Zbog nesavršenosti algoritama koji se koriste za repliciranje pokreta glave, moguće je otkriti neprirodne pokrete vrata i glave. Markiranjem neprirodnih treptaja, nedoslednosti u izrazima lica (neprirodna, odnosno usporena ili preterana mimika), posebno analiziranjem pokreta delova lica (oči, nos, usta) ukazuje se na mogućnost da je video zapis veštački kreiran. Slična situacija je i sa algoritmima koji repliciraju senke i osvetljenja. Takođe, video zapisi nastali ovom tehnologijom mogu imati izobličenja i treperenja kojih nema u originalnom videu. Analiziranje zvuka kod video i audio zapisa se vrši na sličan način – traže se izobličenja zvuka, odnosno ispituju se audio karakteristike deepfake sadržaja, naročito visina, ritam i intonacija govora, uključujući i boju glasa.[16]

[14] Artefakti su predmeti koji su nastali ljudskom delatnošću, bilo da je reč o stvaranju ili prepravljanju.
[15] Ciftci, N., Kantarci, M., & Erdogmus, K. (2020). FakeCatcher: Detection of audio Deepfakes using biologically-inspired spectro-temporal features. arXiv preprint arXiv:2006.00810.
[16] Više o tome u Shaaban, O., Yildirim, R., & Alguttar, A. (2023). Audio deepfake approaches. *IEEE Access PP*(99): 1–1. DOI: 10.1109/ACCESS.2023.3333866

Za svaki od ovih medijskih sadržaja, za koji se sumnja da je izmanipulisan, pored primene navedenih tehnika, vrši se i provera činjenica, odnosno informacija koje sadrže, kako bi se utvrdilo da li su autentični. Sve tehnike otkrivanja prisustva deepfake tehnologije takođe koriste algoritme mašinskog učenja, koji su prethodno obučavani na skupu podataka autentičnih i lažnih sadržaja, kako bi se razvio model za njihovo precizno otkrivanje. Greške koje algoritmi prave pri kreiranju lažnih sadržaja mogu se po istom obrascu pojaviti i pri upotrebi tehnika otkrivanja. To podrazumeva da uporedo sa razvojem deepfake tehnologije moraju se razvijati i tehnike za otkrivanje lažnih sadržaja.

5. Regulatorni izazovi

Da bi napori uloženi na razvijanje tehnika za otkrivanje lažnih sadržaja sačinjenih *deepfake* tehnologijom, kao i na razvijanje i sprovođenje bezbednosnih strategija u ovoj oblasti dali efikasne rezultate, neopodno je postojanje relevantnog pravnog okvira. Evropska unija je prva usvojila veliki regulator upotrebe veštačke inteligencije – Zakon o veštačkoj inteligenciji, kojim je stvorila pravni okvir za oblasti primene veštačke inteligencije, s ciljem umanjivanja štete koja može nastati njenim zloupotrebama.[17] Međutim, ni ovaj zakon ne pokriva sve rizike zloupotrebe *deepfake* tehnologije, jer ne reguliše „aplikacije koje nisu eksplicitno zabranjene ili navedene kao visokorizičme“.[18]

Ako se osvrnemo na oko možda i najbezazleniju (zlo)upotrebu ove tehnologije – „vaskrsavanje“ preminulih ličnosti u filmskoj industriji – suočavamo se pitanjem vlasništva prava i saglasnosti, dok se u pogledu kreiranja peronalizovanog sadržaja postavlja pitanje privatnosti i zaštite podataka. Na neka od ovih pitanja odgovori su obuhvaćeni kompleksom međunarodnih i nacionalnih pravnih instrumenata, naročito u pogledu prava na privatnost, odnosno zaštitu podataka o ličnosti. Međutim, međunarodnim aktima kojima se štiti pravo na slobodu mišljenja i govora, koje je ugrađeno i u ustave država, se može zamagliti granica između pravnog i protivpravnog. Neki od oblika *deepfake* sadržaja mogu biti zaštićeni upravo ovim aktima o slobodi izražavanja, zbog čega je neophodno izvršiti uravnoteženje potrebe za zaštitom pojedinaca od obmanjujućih sadržaja sa potrebom za zaštitom prava na slobodu govora. Kao prepreka se javlja nepostojanje univerzalno prihvaćene definicije *deepfake* sadržaja, zbog čega različite jurisdikcije mogu imati rezličite

[17] Mladenov, M. (2023). Human vs. Artificial intelligence – EU's legal response. *Pravo – teorija i praksa*, 40(1), p. 32. DOI:10.5937/ptp2300032M

[18] The EU Artificial Intelligence Act – Up-to-date developments and analyses of the EU AI act. Preuzeto 5. jula 2024. sa https://artificialintelligenceact.eu

definicije i pravne okvire za rešavanje potencijalnih bezbednosnih i pravnih problema nastalih primenom ove tehnologije.

Falsifikatori, sada potpomognuti razvojem sofisticiranih tehnologija, igraju značajnu ulogu brisanju državnih granica za kriminal.[19] Sadržaji kreirani putem *deepfake* tehnologije prelaze državne granice, što otežava otkrivanje mesta gde su nastali. To je veliki izazov za agencije za sprovođenje zakona pri sprovođenju istrage, ali i krivičnog gonjenja u vezi sa kreiranjem i plasiranjem ovih lažnih sadržaja.[20] Takođe, to izazovom čini i pitanje nadležnosti, ne toliko u pogledu odgovornosti, jer se ona definiše propisivanjem oblika izvršenja, koliko u pogledu različitosti zakona i propisa u vezi sa *deepfake* tehnologijom, što je u vezi sa već pomenutim problemom nepostojanja opšteprihvaćene definicije. I samo spovođenje zakona i propisa o *deepfake* tehnologiji može predstavljati veliki izazov za agencije za sprovođenje zakona, jer ne samo da se ovi sadržaji sačinjavaju anonimno već se u tajnosti i šire, što otežava identifikovanje i krivično gonjenje kreatora i distributera lažnih sadržaja. Kao rešenje se nameće neophodnost uspostavljanja trajnih i čvrstih oblika saradnje i koordinacije vlada i agencija za sprovođenje zakona pri sprovođenju aktivnosti na razvijanju efikasnih propisa za pravno regulisanje *deepfake* tehnologije. Ovakva saradnja uključuje razmenu informacija i najboljih praksi, razvoj zajedničkih standarda i protokola, kao i zajednički rad na rešavanju novih pretnji, što zahteva uključivanje u ove procese i tehnoloških kompanija, obrazovnog sistema i svih raspoloživih kapaciteta za podizanje svesti javnosti.

Konačno, kreatori zakona se suočavaju sa velikim izazovom razvijanja efikasnih propisa koji mogu da idu u korak sa dinamičnim razvojem *deepfake* tehnologije. Prevazilaženje ovog problema je moguće razvijanjem fleksibilnih i prilagodljivih regulatornih okvira, koji mogu da prate ovako dinamičan razvoj *deepfake* tehnologije. To podrazumeva korišćenje regulatornih sendbokseva[21] ili drugih pristupa za proveru novih propisa pre stupanja na snagu.

[19] Marković, D.M. (2021). The role and importance of border police in detecting forged documents. *Nauka i društvo, 7*(1), p. 216. DOI:10.5281/zenodo.5572518

[20] Matijašević, J., & Dragojlović, J. (2021). Metodika otkrivanja krivičnih dela računarskog kriminaliteta. Kultura polisa, 18(2), 51–63. DOI:10.51738/Kpolisa2021.18.2p.1.04

[21] Sandbox – mesto na koje se izdvajaju maliciozni sadržaji i delovi kodova od ostatka okruženja, kako bi se kasnije mogli analizirati i utvrditi da li predstavljaju pretnju sistemu.

6. Zaključak

Među najmarkantnijim odlikama savremenog društva u eri digitalne tehnologije jeste povećan značaj dostupnosti podataka, koji predstavljaju vredan informacioni i analitički resurs, čijim raspolaganjem i pravilnom upotrebom se može poboljšati donošenje odluka, personalizacija usluga, pokretanje inovacija i uspešnije sprovođenje naučno-istraživačkih poduhvata. Zahvaljujući dinamičnom razvoju digitalnih tehnologija javljaju se novi oblici rada na daljinu a postojeći dobijaju mogućnost unapređenja kroz automatizaciju. Takođe, poboljšava se udobnost života, uključujući poboljšanje zdravstvene zaštite i povećanje efikasnosti medicinskih tretmana.

Međutim, sofisticirane tehnologije donosi i sofisticirane mogućnosti njihove zloupotrebe. Zahvaljujući brzom razvoju mašinskog učenja i veštačke inteligencije, *deepfake* tehnologija je u poslednjih sedam godina doživela veliki napredak, omogućavajući stvaranje vrlo realističnih lažnih sadržaja, kako u formi teksta i slike, tako i u obliku audio i video zapisa, koji se mogu koristiti za prevaru i manipulaciju ljudima. Stoga, zloupotreba *deepfake* tehnologije, radi širenja dezinformacija, narušavanja reputacija i radi prevara, predstavlja snažan izazov za bezbednost pojedinaca, organizacija i društva u celini, uz istovremeno kršenje pozitivnih etičkih principa.

Dinamična i evoluirajuća priroda *deepfake* tehnologije otežava identifikovanje sadržaja koji se njome kreiraju, posebno imajući u vidu da tehnike otkrivanja lažnih sadržaja nisu pouzdane, zbog čega postoji mogućnost da ne budu efikasne u otkrivanju najnaprednijih *deepfake* sadržaja. Ovo se reflektuje na efikasnost rešavanja jedinstvenih izazova koje postavlja ova tehnologija, kao što je teškoća u identifikaciji i verifikaciji autentičnosti, posebno u okolnostima kada zakonodavni okvir čine tradicionalni regulatorni pristupi.

Kao najznačajniji regulatorni izazovi koje postavlja zloupotreba *deepfake* tehnologije u prvom planu su pitanja nadležnosti, pitanja slobode govora i nužnost međunarodne saradnje. Razvijanje efikasnih propisa koji mogu da prate dinamični razvoj ove sofisticirane tehnologije zahteva saradnju, koordianciju, dinamičnu regulativu, međunarodnu saradnju. S obzirom na to da ni Zakon o veštačkoj inteligenciji, kao prvi veliki regulator veštačke inteligencije, koji je usvojila Evropska unija, ne obuhvata sve visokorizične aplikacije, nije teško zaključiti da dinamičnost regulative neće biti zadovoljena samo prilagođavanjem postojećih propisa novim okolnostima, već i usvajanjem celovitih novih zakonskih rešenja. Izazovi u regulisanju i identifikaciji *deepfake* sadržaja zahtevaju višestrani pristup. Kreatori javnih politika, zakonodavci, tehnološke kompanije i sve zainteresovane strane moraju raditi zajedno kako bi razvili efikasne strategije za rešavanje ovih izazova i ublažavanje rizika od *deepfake* tehnologije. Obrazovanje i razvijanje svesti javnosti su takođe nužan element tih procesa.

Darko M. Marković, PhD, Associate Professor
Faculty of Law for Commerce and Judiciary in Novi Sad,
Business Academy University in Novi Sad

Mina Zirojević, PhD, Full Professor
Ministry of Education, Belgrade

CHALLENGES IN REGULATING AND IDENTIFYING DEEPFAKE CONTENT

Abstract:

The modern development of digital technologies brings numerous advantages to a person's everyday life, opening up new opportunities, expanding and facilitating existing ones. Thanks to this, the fulfilment of both private and professional obligations is improved in ways that lead to increased efficiency and productivity at work. The availability of digital databases and the availability of advanced digital tools and techniques facilitate scientific research. The existence of accessible online platforms contributes to the exchange of ideas and experiences, modernization of production processes, positive transformation of the health care industry, and progress of science as a whole. Unfortunately, simultaneously to contributing to the development of society in all segments of life, new areas for numerous abuses are opening up. In this regard, the scope and intensity of the development of digital technologies multiply the existing moral and security challenges while producing new ones at the same time. The inadequate development of appropriate regulatory responses significantly contributes to this. The challenges in regulating and identifying abuses of digital technologies have been particularly pronounced with the development of artificial intelligence, whose techniques have increased the possibilities of generating fake audio and video recordings, as well as images, to unimaginable limits. Having that in mind, our goal in this paper is to point out the main security challenges produced by the technology of creating deepfake content and, accordingly, the key regulatory challenges, both in national and international frameworks. Applying the methods of content analysis, induction and deduction we have come to a conclusion that the overall power of opposing the abuse and criminalization of deepfake techniques requires dynamic development of tools for their identification, establishment of appropriate business standards, development of protection strategies, as well as urgent improvement of the existing regulatory frameworks, not excluding the need to adopt completely new legal solutions.

Keywords: *digitization, digital technologies, security challenge, legal framework, privacy*

8. Reference

1. Chaffey, D. (2024, May 1). Global social media statistics research summary May 2024. *Smart Insights*. Preuzeto sa : https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/
2. Chesney, R., & Citron, D. K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, *107*(6), 1753-1794.
3. Holdsworth, J., & Scapicchio, M. (2024, June 17). What is deep learning? *IBM*. Preuzeto sa: https://www.ibm.com/topics/deep-learning
4. Joiner, I. A. (2018). *Emerging library technologies*. Chandos Publishing.
5. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006
6. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. DOI:10.1038/nature14539
7. Marković, D.M. (2021). The role and importance of border police in detecting forged documents. *Nauka i društvo*, *7*(1), 202-217. DOI:10.5281/zenodo.5572518
8. Matijašević, J., & Dragojlović, J. (2021). Metodika otkrivanja krivičnih dela računarskog kriminaliteta. Kultura polisa, 18(2), 51–63. DOI:10.51738/Kpolisa2021.18.2p.1.04
9. Mladenov, M. (2023). Human vs. Artificial intelligence – EU's legal response. *Pravo – teorija i praksa*, *40*(1), 32–43. DOI:10.5937/ptp2300032M
10. Nagy, Z. (2019). *Osnove veštačke inteligencije i mašinskog učenja*. Beograd: Kompjuter biblioteka.
11. Nguyen, T. T, Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q. V., & Nguyen, C. M. (2022). Deep Learning for Deepfakes Creation and Detection: A Survey. *Computer Vision and Image Understanding*, 223, 103525, 1–19. DOI:10.1016/j.cviu.2022.103525
12. Preminger, A., & Kugler, M. B. (2024). The right of publicity can save actors from deepfake Armageddon. *Berkeley Technology Law Journal*, Northwestern Public Law Research Paper No. 23–52, 101–158. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4563774
13. Shaaban, O., Yildirim, R., & Alguttar, A. (2023). Audio deepfake approaches. *IEEE Access PP*(99): 1–1. DOI: 10.1109/ACCESS.2023.3333866
14. The EU Artificial Intelligence Act – Up-to-date developments and analyses of the EU AI act. Preuzeto 5. jula 2024. sa https://artificialintelligenceact.eu

15. Thies, J., Zollhöfer, M., Nießner, M., Stamminger, M., Theobalt, C., & Nießner, M. (2019). Deferred Neural Rendering: Image Synthesis Using Learned Gradient Descent. *ACM Transactions on Graphics*, *38*(6), 1-14. Preuzeto 24. juna 2024. sa https://niessnerlab.org/papers/2019/11neuralrendering/thies2019neural_preprint.pdf