Umjetna inteligencija u kontekstu novomedijske manipulacije - forma deep fake videa

Bižaca, Leonarda

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
AKADEMIJA ZA UMJETNOST I KULTURU
ODSJEK ZA KULTURU, MEDIJE I MENADŽMENT

SVEUČILIŠNI PREDDIPLOMSKI STUDIJ
KULTURA, MEDIJI I MENADŽMENT

LEONARDA BIŽACA

UMJETNA INTELIGENCIJA U KONTEKSTU NOVOMEDIJSKE MANIPULACIJE – FORMA *DEEP FAKE* VIDEA

ZAVRŠNI RAD

MENTOR/I:
Mentor: izv. prof. dr. sc. Ivica Šola
Komentor: dr. sc. Snježana Barić-Šelmić, poslijedoktorandica

Osijek, 2023. godina

SADRŽAJ

Sažetak

U brzom napretku tehnologije umjetne inteligencije pojava deepfake videa postala je izazovna tema koja intrigira znanstvenike, stručnjake za medije i društvene znanosti diljem svijeta. Ovaj znanstveni rad istražuje aspekte umjetne inteligencije u kontekstu manipulacije medija putem deepfake tehnologije. Fokusirajući se na duboko lažiranje audiovizualnog sadržaja, rad će analizirati temeljne koncepte generativnih suparničkih mreža (GAN) i dubokog učenja te kako ove tehnike omogućuju kreiranje uvjerljivih lažnih sadržaja. Također, kako bi se dodatno produbilo znanje o utjecaju umjetne inteligencije, u radu će se razraditi povijest i definicija medija, ali i same početke nastanka umjetne inteligencije. Nadalje, rad razmatra implikacije koje deepfake tehnologija ima na sferu medija i društva proučavajući potencijalnu upotrebu u političkom manipuliranju, lažnim vijestima i identitetskoj krađi. Razmatrajući percepciju javnosti o autentičnosti medijskog sadržaja, rad se bavi temama poput ranjivosti društva koje proizlazi iz manipulacija putem deepfake videa. Kroz kritički okvir razmatranja rad će pružiti dublji uvid u izazove i moguće strategije suočavanja s rastućim fenomenom deepfake tehnologije u svijetu medija i komunikacije. Kroz analizu dosadašnjih istraživanja i primjera iz stvarnog svijeta rad će doprinijeti boljem razumijevanju utjecaja umjetne inteligencije na suvremeni medijski svijet te potaknuti daljnje razmišljanje i istraživanje u ovoj važnoj sferi.

Ključne riječi: Deepfake tehnologija, društvene implikacije, etički izazovi, manipulacija medijima, umjetna inteligencija.

Introduction

In the rapid advancement of artificial intelligence technology, the emergence of deep fake videos has become a challenging subject that captivates scientists, media experts, and social scientists worldwide. This scientific paper delves into the fundamental aspects of artificial intelligence in the context of media manipulation through deep fake technology. By focusing on the intricate fabrication of audiovisual content, this paper will analyze the core concepts of Generative Adversarial Networks (GANs) and deep learning, exploring how these techniques facilitate the creation of convincing fabricated content. Additionally, to further enrich our understanding of the impact of artificial intelligence, we will delve into the history and definition of media and the origins of artificial intelligence itself. Furthermore, this paper will contemplate the implications of deep fake technology on media and society. Investigating its potential application in political manipulation, fake news dissemination, and identity theft, the paper will delve into public perception regarding the authenticity of media content. Through a critical analytical framework, this paper will provide deeper insights into the challenges and potential strategies to confront the growing phenomenon of deep fake technology in media and communication. By scrutinizing past research and real-world examples, this paper will better comprehend artificial intelligence's influence on the contemporary media landscape and stimulate further contemplation and exploration in this significant domain.

Keywords: Artificial intelligence, deep fake technology, ethical challenges, media manipulation, societal implications.

1. UVOD

Ubrzan tehnološki napredak posljednjih desetljeća doveo je do formiranja novih paradigmi i inovacija koje su duboko oblikovale način na koji funkcionira suvremeno društvo. Jedno je od najznačajnijih dostignuća tog napretka razvoj umjetne inteligencije (UI). UI kao grana računalne znanosti istražuje mogućnosti stvaranja sustava koji mogu razmišljati, učiti i donositi odluke slične onima koje donosi ljudski um.

Prvo poglavlje ovog znanstvenog rada razmatra osnove umjetne inteligencije. Definiramo UI kao sposobnost računalnih sustava da obavljaju zadatke koji inače zahtijevaju ljudski intelekt. U nastavku istražujemo podjelu UI-a na strojno učenje i duboko učenje. Strojno učenje je tehnika koja omogućuje računalima da uče iz podataka i poboljšavaju svoje performanse tijekom vremena. Duboko učenje je podskup strojnog učenja koji se temelji na neuronskim mrežama i omogućuje računalima da obrade kompleksne obrasce i reprezentacije. Drugo poglavlje fokusira se na razvoj umjetne inteligencije kroz vrijeme. Nadalje, predstavlja razliku između slabe i jake UI. Slaba umjetna inteligencija obuhvaća sustave koji su usmjereni na specifične zadatke i nemaju stvarni razum ili svijest. S druge strane, jaka umjetna inteligencija označava razvoj sustava koji imaju sposobnost dubokog razumijevanja, svijesti i općenitog učenja. (Russell i Norvig, 2010)

Treće poglavlje istražuje pojam medija i analizira njihove prednosti i mane u suvremenom društvu. Mediji su postali ključan kanal komunikacije i informiranja, ali isto tako nose odgovornost za prenošenje istinitih informacija. Četvrto poglavlje posvećeno je detaljnom istraživanju razvoja deepfake videa. Duboki lažni sadržaji postali su značajna tema u svijetu tehnologije i medija. Istražujemo kako duboko učenje omogućuje stvaranje uvjerljivih lažnih video sadržaja putem manipulacije postojećih video i audioelemenata. Posljednje poglavlje fokusira se na problem manipulacije medija, posebno u kontekstu deepfake videa. Analiziramo kako ovi lažni sadržaji mogu ozbiljno narušiti povjerenje u medije i informacije koje konzumiramo. Ističemo potrebu za razvojem tehnoloških i društvenih rješenja kako bismo se nosili s izazovima koje donosi manipulacija medija.

U ovom znanstvenom radu duboko ćemo istražiti koncept umjetne inteligencije, njezin razvoj i primjenu u stvaranju deepfake videa te ćemo analizirati utjecaj tih tehnologija na medije i društvo kao cjelinu.

2. UMJETNA INTELIGENCIJA

Od samih početaka, kao što je općenito poznato, glavni pokretači bili su tehnološki napretci. U početku su bitne tehnologije opće namjene, koje su ujedno i najvažnije te su najviše utjecale na naš život. Parni stroj, električna energija, motor s unutarnjim izgaranjem i iskorištavanje nuklearne lančane reakcije neki su od primjera koji razlikuju tehnologije opće namjene od običnih tehnoloških inovacija. Najvažnija je tehnologija opće namjene u 21. stoljeću upravo umjetna inteligencija. (Europska komisija, 2018)

Umjetna inteligencija (UI) odnosi se na sposobnost digitalnih računala ili računalno kontroliranih robota da obavljaju zadatke koji su obično povezani s inteligentnim bićima, kao što je definirano od strane Copelanda (2014). Norvig i Russell (1995: 3) tvrde da je pitanje izgradnje umjetne inteligencije teško, ali da istraživači imaju čvrste dokaze da je zadatak ostvariv. „Umjetna inteligencija propituje jednu od konačnih zagonetki. Kako je moguće da spor, maleni mozak, biološki ili elektronički, može percipirati, razumjeti i predviđati svijet, te manipulirati svijetom mnogo većim i mnogo kompleksnijim nego što je on? Kako da izgradimo nešto s takvim svojstvima? Ta su pitanja teška, ali, za razliku od putovanja brzinom većom od brzine svjetlosti ili antigravitacijskog uređaja, istraživač na području umjetne inteligencije ima čvrste dokaze da je zadatak moguće ostvariti. Sve što trebamo učiniti je pogledati u zrcalo da bismo vidjeli primjer inteligentnog sustava.“ (Norvig i Russell, 1995: 3)

Prema Kovač (2015: 3) GUI (Artificial General Intelligence, AGI) odnosi se na istraživanja koja se bave stvaranjem strojeva koji su sposobni izvršavati raznolike inteligentne zadatke. Za razliku od specijalizirane umjetne inteligencije, generalna umjetna inteligencija mora biti sposobna upravljati različitim područjima te se nositi s novim zadacima. Zbog toga još uvijek nedostaju precizne definicije područja djelovanja umjetne inteligencije, a fokus je istraživanja na komponentama inteligencije poput učenja, zaključivanja, rješavanja problema, percepcije i jezične uporabe. No drugi se znanstvenici nadovezuju: „Dokle god se ne dokaže da GUI nije moguć, ostaje legitimna tema istraživanja. A uzevši u obzir kompleksnost problema, nema razloga očekivati da GUI postigne cilj u kratkom periodu, pa će sve popularne teorijske kontroverze vjerojatno postojati i nakon što se ostvari GUI, kako je planirano. Činjenica da u istraživanjima ima malo konsenzusa treba nas učiniti opreznijima kada prosuđujemo novu ideju kao potpuno pogrešnu. Kako se već više nego jednom u povijesti događalo, pravi prodor mogao bi proizaći iz nečega što se protivi intuiciji.“ (Goertzel i Wang, 2007: 15)

U članku objavljenom u knjizi „Osnove umjetene inteligencije“, kako je navela Kovač (2015), Schank et al. (1993: 6) navode nekoliko ključnih značajki inteligencije. Dakle, komunikacija je moguća samo među inteligentnim subjektima, dok se unutarnje znanje odnosi na svijest o sebi i svojim potrebama. Znanje o svijetu podrazumijeva sposobnost prepoznavanja i upotrebe informacija iz okoline, kao i upotrebu memorije za procesiranje novih iskustava. Ciljevi i planovi važni su za postizanje željenih ciljeva, dok kreativnost podrazumijeva sposobnost prilagodbe na promjene u okolini i učenje iz iskustva. Schank smatra da je kreativnost ključni element inteligencije te da inteligentan entitet mora biti sposoban učiti kako bi se prilagodio promjenama u okolini. Kreativnost je pretpostavljena u određenim stupnjevima kod svakog inteligentnog entiteta, no sama definicija kreativnosti nedovoljno je precizna. „Kreativnost prije svega znači sposobnost prilagodbe na promjene u okolini te sposobnost učenja iz iskustva. Možemo reći kako entitet koji ne uči vjerojatno nije inteligentan.“ (Schank et al., 1993: 6)

Razvoj računalnih sustava koji mogu obavljati zadatke koji su tradicionalno bili povezani s ljudskom inteligencijom predstavlja polje umjetne inteligencije. Ovo polje koristi različite pristupe, uključujući neuronske mreže, evolucijsko računanje, logičko zaključivanje i drugo. Prema Russellu i Norvigu (2010: 7) umjetna inteligencija odnosi se na stvaranje softverskih sustava koji imaju inteligentne agente sposobne obavljati zadatke koji obično zahtijevaju ljudsku inteligenciju, kao što su percepcija, zaključivanje, učenje i donošenje odluka.

Russell i Norvig u već spomenutom radu „Umjetna inteligencija: Moderan pristup“ navode nekoliko vrsta umjetne inteligencije:

1. Logička umjetna inteligencija – ova vrsta AI koristi logička pravila i zaključivanje za rješavanje problema. Koristi se u stručnim sustavima gdje stručnjaci stvaraju bazu znanja.
2. Algoritamska umjetna inteligencija – ova vrsta AI koristi algoritme pretrage i optimizacije za rješavanje problema, poput igara, planiranja ruta i raspoređivanja.
3. Teorijska umjetna inteligencija – ova vrsta AI usredotočena je na razvoj novih teorija i modela inteligencije, poput teorija učenja, percepcije i kognicije.
4. Biološka umjetna inteligencija – ova vrsta AI koristi modele bioloških sustava, poput neuronskih mreža, evolucijskih algoritama i genetskog programiranja za razvoj novih algoritama i pristupa inteligentnom ponašanju.

2.1.1. Strojno učenje

Umjetna inteligencija, s posebnim naglaskom na strojno učenje, smatra se najvažnijom tehnologijom općenito u današnje vrijeme. Russell i Norvig (2010) u 18. poglavlju „Modernog pristupa“ opisuju strojno učenje kao postupak u kojem računalo automatski poboljšava svoje performanse u izvršavanju zadatka putem iskustva. Djelo „Umjetna inteligencija: Moderni pristup“ navodi primjere primjene strojnog učenja, poput klasifikacije teksta, prepoznavanja slika i preporuka proizvoda. Umjesto da programer ručno programira svaki korak i pravilo u računalnom sustavu, strojno učenje omogućuje računalu da samostalno nauči ta pravila iz podataka.

Strojno učenje sastoji se od triju glavnih komponenti: skupa podataka, modela i algoritama. Skup podataka sastoji se od primjera koje računalo koristi za učenje i izgradnju modela, a model predstavlja matematički opis kako će računalo obraditi i klasificirati nove primjere. Strojno učenje primjenjuje se u različitim područjima kao što su prepoznavanje slika, prevođenje prirodnog jezika, preporučivanje proizvoda, autonomna vožnja, analiza financijskih podataka i slično.

Nadalje, strojno učenje ima utjecaj na tri razine: zadatke i zanimanja, poslovne procese i poslovne modele – izriču Brynjolfsson i McAfee (2017) u radu o uklapanju umjetne inteligencije u organizacijske timove. U posljednjih nekoliko godina strojno učenje postalo je puno učinkovitije i dostupno na širem području nego ikada prije. Strojno učenje spada u područje slabe umjetne inteligencije jer ne razumije govor, već samo povezuje simbole i značenja te identificira uzorke. Glavna je razlika što računalo prepoznaje obrasce u podatcima umjesto da ih programer pruža. Ovo je ujedno i jedno od najbrže rastućih područja u informacijskoj tehnologiji i razvoju umjetne inteligencije, što je omogućeno padom troškova pohrane i obrade podataka u posljednjih nekoliko godina.

2.1.2. Duboko učenje

Akerkar (2019: 3) piše kako se duboko učenje sastoji od mnogih hijerarhijskih slojeva koji obrađuju informacije na nelinearni način, gdje neki koncept niže razine pomaže u definiranju koncepata više razine. Dakle, duboko učenje definirano je kao klasa strojnog učenja koja koristi mnoge slojeve nelinearne obrade informacija za nadzirane i nenadzirane značajke ekstrakcije i transformacije te za analizu uzoraka i klasifikaciju.

Russel i Norvig (2010) u prvom poglavlju svog rada „Umjetna inteligencija: Moderni pristup“ duboko učenje opisuju kao vrstu strojnog učenja koja se temelji na umjetnim neuronskim mrežama s više slojeva. Navode da ova vrsta tehnologije omogućuje računalima da uče složene obrasce u podatcima i donose odluke na temelju tih uzoraka. Umjetne neuronske mreže računalni su modeli inspirirani biološkim neuronima u mozgu. Oni se sastoje od slojeva neurona koji su povezani težinskim vrijednostima. Kada se podatci puste kroz neuronsku mrežu, težinske se vrijednosti podešavaju kako bi se minimizirala greška između stvarnog izlaza i očekivanog izlaza. Nakon mnogo interakcija neuronska mreža može naučiti složene obrasce i znanja iz podataka koji se koriste za obuku.

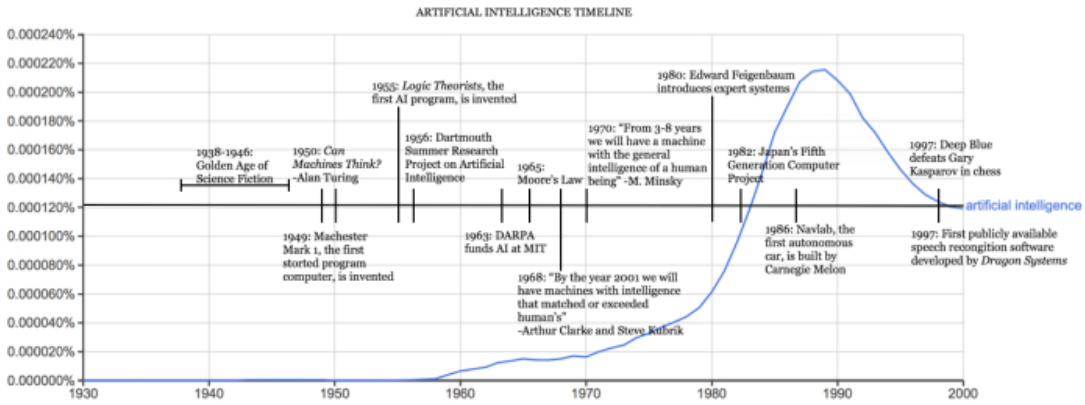
2.2. RAZVOJ UMJETNE INTELIGENCIJE KROZ POVIJEST

Prema članku Harvardova osvrta na povijest UI-a (2017) povijest umjetne inteligencije (UI) obuhvaća niz ključnih događaja i tehnoloških napredaka koji su omogućili razvoj i primjenu inteligentnih sustava. Tijekom desetljeća razvoj UI-a prošao je kroz različite faze i mijenjao se sukladno napretku računalne tehnologije i znanstvenim otkrićima.

Za početak, korijeni umjetne inteligencije sežu u ranu povijest računalnih znanosti. Rockwell (2017) napominje da su još u 1950-ima istraživači poput Alana Turinga i Johna McCarthyja postavljali temelje UI-a. Turing je predložio pojam „Turingova testa" kojim se mjeri sposobnost računalnog sustava da se ponaša poput ljudskog bića, dok je McCarthy skovao pojam „umjetna inteligencija" i organizirao prvu konferenciju o UI-u. Tijekom 1950-ih i 1960-ih godina istraživači su se usredotočili na razvoj logičkih sustava i strojeva za simboliku. Nadalje navodi da je Dartmouth College 1956. godine organizirao „Ljetni institut za umjetnu inteligenciju" koji je okupio ključne ličnosti u području UI-a. Tijekom tog razdoblja razvijeni su prvi programski jezici poput Lispa, koji je postao popularan u istraživanjima UI-a.

Međutim, sljedeća desetljeća, 1970-e i 1980-e, bila su obilježena razočaranjem u području UI-a. Rockwell (2017) tvrdi da istraživanja nisu donijela očekivane rezultate, a ograničenja računalne snage i nedostatak kvalitetnih podataka ograničavali su napredak. Ovo razdoblje poznato je kao „zima umjetne inteligencije" jer je interes za UI smanjen.

No 1990-ih godina interes za UI ponovno se razbuktao zahvaljujući napretku u računalnoj tehnologiji i razvoju interneta. Pojavile su se nove metode i algoritmi koji su poboljšali performanse računalnih sustava. Strojno učenje, posebno neuronske mreže, postale su popularne i omogućile rješavanje složenih zadataka poput prepoznavanja uzoraka i klasifikacije. Slijedilo je intenzivno istraživanje i razvoj različitih područja UI-a. (Rockwell, 2017)



Slika 1. Prikaz razvoja umjetne inteligencije

Izvor: https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

Danas se UI primjenjuje u raznim sektorima poput zdravstva, financija, proizvodnje, obrazovanja, sigurnosti i drugih. Razvijaju se sustavi za samovozeće automobile, robotsku kirurgiju, virtualne asistente, sustave preporuka i drugo. UI sve se više koristi i u raznim aplikacijama poput aplikacija za prepoznavanja govora, prevođenja jezika, prepoznavanja slika i videa, prepoznavanja emocija i slično. Iako je UI danas uvelike razvijen, još uvijek postoje mnogi izazovi u primjeni UI-a u stvarnom svijetu. Problem etičnosti i privatnosti podataka, sposobnost prilagodbe sustava promjenjivim okruženjima te nedostatak transparentnosti i razumljivosti nekih UI sustava samo su neki od izazova s kojima se suočavaju istraživači i inženjeri UI-a. U svakom slučaju, razvoj UI-a otvorio je vrata novim mogućnostima i perspektivama u primjeni računalnih sustava.

Napredak u ovom području utjecat će na naše svakodnevne živote i transformirati način na koji radimo, učimo, putujemo i komuniciramo.

2.3. VRSTE UMJETNE INTELIGENCIJE

Prema Polšeku (2003: 201–205) razvijaju se dva glavna smjera u području umjetne inteligencije. Prvi pristup obuhvaća samostalne programe koji su sposobni međusobno komunicirati i postizati svoje ciljeve. Ovaj pristup obuhvaća discipline poput tradicionalne umjetne inteligencije, sintetičkog života i evolucijskog računalstva. Upravo ova podjela omogućava razlikovanje snažne i slabe umjetne inteligencije. Snažna umjetna inteligencija, koja se ponekad naziva i svjesna umjetna inteligencija, odnosi se na strojeve koji su sposobni za kritičko razmišljanje, osjećanja i razumijevanje vlastitog razmišljanja. Ova razina umjetne inteligencije omogućava replikaciju ljudskih mentalnih karakteristika kao što su instinkti, kreativnost i čak emocije. S druge strane, koncept slabe umjetne inteligencije tvrdi da je glavni doprinos računalstva proučavanju uma u tome da pojačava, oblikuje i testira istraživačke hipoteze. Računala s ovom vrstom inteligencije nisu samo alati za mentalno istraživanje, već dobro dizajnirana računala mogu razmišljati na način sličan ljudima. Funkcionalna hipoteza sugerira da računalo koje izvodi zadatke slične ljudskim kognitivnim procesima zapravo može razumjeti i pripisati mu se zasluge za mentalna i kognitivna stanja, kako je iznio Polšek (2003: 201–205).

2.3.1. Slaba umjetna inteligencija

Slaba umjetna inteligencija (engl. *narrow AI*) vrsta je umjetne inteligencije koja je usmjerena na rješavanje specifičnih zadataka ili problema. (Putica, 2018: 204) Ovi AI sustavi usko su specijalizirani za jednu domenu i nemaju sposobnost općeg razumijevanja ili prilagodbe izvan te domene. Prema Putici (2018: 205) slaba umjetna inteligencija naziva se i ograničenom, a podrazumijeva gradnju više autonomnih sustava ili algoritama sposobnih rješavati problemska područja. Kod ove vrste umjetne inteligencije stroj nije inteligentan, već simulira inteligenciju. Slaba umjetna inteligencija često je prisutna u našem svakodnevnom životu, kao što su virtualni asistenti poput Appleova Siri ili Amazonova Alexa, preporučeni sustavi na platformama za streaming glazbe i videozapisa ili sustavi za prepoznavanje govora i slika. Slabi AI sustavi usko su fokusirani na određenu zadaću ili domenu. Nedostatak općeg razumijevanja slabe umjetne inteligencije podrazumijeva da ne posjeduje opću inteligenciju i nema sposobnost razumijevanja šireg konteksta, već samo uzorka naučenog.

2.3.2. Jaka umjetna inteligencija

Jaka umjetna inteligencija još nije dosegnuta, tj. njen bi razvoj podrazumijevao inteligenciju jednaku ljudskoj (opća umjetna inteligencija) ili pak snažniju od nje (umjetna superinteligencija), uz doseg stanja svijesti. (Hrvatska enciklopedija, 2021) U različitim znanstvenim člancima definicija jake umjetne inteligencije može varirati ovisno o autorima, njihovim istraživačkim pristupima i teorijskim pozadinama. Međutim, općenito gledano, može se reći da jaka umjetna inteligencija (engl. *Strong Artificial Intelligence*) označava koncept da računalni sustavi ili programi imaju sposobnost manifestiranja stvarne svijesti, razumijevanja, svjesnosti i možda čak emocija, slično ljudskom umu. Prema Russellu i Norvigu (2010: 1020) Strong AI je AI sustav koji može izvršiti bilo koji intelektualni zadatak koji čovjek može učiniti. To bi, dakle, zahtijevalo ne samo sposobnost obrade i analize informacija već i razumijevanje jezika, donošenje odluka i posjedovanje općeg znanja o svijetu. Očekuje se da bi jaki AI sustavi mogli obuhvaćati robote koji mogu obavljati širok spektar složenih zadataka, sustave za samostalno učenje, autonomna vozila i druge napredne tehnologije.

Prema Putici (2018: 205) John Searle posebno je kritizirao jaku umjetnu inteligenciju. Searle osporava koncept računalne psihologije, koja tvrdi da je računalo um sa svim kognitivnim stanjima sličnim ljudskima. U svom eseju „Umovi, mozgovi i programi“, kako je u radu navela Putica, Searle argumentira da simulacija nije isto što i replikacija. Centralna točka njegove rasprave, poznate kao „Rasprava o kineskoj sobi“, postala je iznimno utjecajna i u području kognitivne znanosti i filozofije umjetne inteligencije. Searle je 1980. godine stvorio misaoni eksperiment nazvan „Kineska ili Searleova soba“. Kroz ovaj eksperiment ukazao je da manipulacija simbolima ne predstavlja nužno razumijevanje i da upravljanje simbolima ne znači razumijevanje istih. Searleov pristup vraća važnost ljudskim sposobnostima i protivi se ideji da su računala i umjetna inteligencija međusobno zamjenjivi. (Putica, 2018: 205)

Dakle, glavna je razlika između slabe i jake umjetne inteligencije u razini općenitosti i fleksibilnosti. Slabi AI sustavi usko su specijalizirani i ne mogu se nositi sa zadatcima izvan svoje domene, dok bi jaki AI sustavi bili sposobni primijeniti svoje znanje i sposobnosti na širok spektar problema i situacija. (Russel i Norvig, 2010)

Tablica 1. Usporedba jake i slabe umjetne inteligencije

Izvor: https://askanydifference.com/hr/difference-between-strong-ai-and-weak-ai-with-table/

Parametri usporedbe	Jak AI	Slab AI
Inteligencijske sposobnosti	Jaka umjetna inteligencija ima svoj um i pokazuje ljudske kognitivne sposobnosti	Najveća razina snažne umjetne inteligencije odgovara točnoj ljudskoj inteligenciji
Tip	Snažna umjetna inteligencija budućnost je umjetne inteligencije	Slaba umjetna inteligencija je sadašnji oblik umjetne inteligencije
Cilj	Razviti inovativne pristupe za svaki zadatak ili problem pomoću umjetne inteligencije	Brže rješavati probleme i zadatke i izvršiti određeni zadatak
primjena	Još uvijek nije primijenjeno	Modeli osobne pomoći temeljeni na glasu kao što su Siri ili Alexa
Maksimalna razina	Najviša razina slabe umjetne inteligencije je pružanje rješenja unutar unaprijed definiranih odgovora.	Najviša razina slabe umjetne inteligencije je pružanje rješenja unutar unaprijed definiranih odgovora

2.4. MEDIJI OD ŠPILJE DO VIRTUALNOG SVIJETA

Mediji imaju svoje korijene u početcima društvenih zajednica u kojima je postojala potreba za javnim priopćavanjem informacija. Medij je posrednik prijenosa poruke od pošiljatelja poruke do primatelja (Jurčić, 2017: 128).

U području komunikacijskih znanosti prema Jurčić (2017: 128) medij predstavlja tehničko ili fizičko sredstvo putem kojeg se poruke prenose putem kanala. Također se definira kao suvremeno sredstvo za prijenos informacija. Medij je također način izražavanja i komunikacije, sredstvo prenošenja poruka.

Prema Vladimiru Bitiju (1997: 213) ovaj se termin može definirati na najmanje četiri načina:

1. U fiziološkom smislu, medij se odnosi na komunikacijske kanale kao što su slušni, vizualni, taktilni i mirisni, te na njihove međusobne odnose (intermedijalnost).

2. U fizičkom smislu, medij označava materijal putem kojeg se prenosi nova poruka, uključujući jezik, ton i boju.

3. U tehnološkom kontekstu, medij označava sredstvo posredovanja između stvaranja i konzumacije znakova.

4. U sociološkom smislu, medij se shvaća kao institucijski i organizacijski okvir komunikacije, što uključuje aspekte poput politike, gospodarstva, znanosti i obrazovanja. Na ovaj način, pojam medija djelomično se preklapa s pojmom diskursa. Ova je interpretacija šira, ali ipak moguća.

„Nakon tri tisuće godina eksplozije, preko fragmentarnih i mehaničkih tehnologija, zapadni se svijet zgušnjava. Tijekom mehaničkog doba produžili smo svoja tijela u prostoru. Danas, nakon više od stoljeća električne tehnologije, i naš središnji živčani sustav sudjeluje u globalnom zagrljaju, a prostor i vrijeme, barem na našemu planetu, više ne postoje. Veoma brzo približavamo se završnoj fazi čovjekovih produžetaka tehnološkoj simulaciji svijesti, kada će kreativni proces spoznavanja kolektivno i združeno obuhvaćati cijelo ljudsko društvo, jednako kao što smo preko raznih medija već produžili svoja osjetila i živce. Hoće li produžetak svijesti, za čime toliko dugo tragaju oglašivači pojedinih proizvoda, biti "dobra stvar", pitanje je koje dopušta široka rješenja. Slaba je mogućnost da na takva pitanja o čovjekovim produžecima

odgovorimo ako ih ne razmotrimo sve zajedno. Svaki produžetak, bilo kože, ruke ili noge, utječe na ukupan psihički i društveni sklop.“ (McLuhan, 2008: 9)

Prema Marshallu McLuhanu u djelu „Razumijevanje medija“ svi oblici medija (govor, pisanje, telefon, radio, televizija) omogućuju produženje ljudskog bića i ljudskih sposobnosti, povećavajući brzinu i opseg komunikacije te premještajući granice koje čovjek može doseći. Radio i telefon produžetci su slušnog osjeta, televizija produžuje vid i sluh, arhitektura je produžetak ljudske kože kao regulatora temperature. McLuhan tvrdi da, bez obzira na to radi li se o produženju ljudske kože, ruku ili nogu, razvoj tehnologije ima dubok utjecaj na kompleksni psihički i sociološki sustav društva. Ovo nam govori da medije ne možemo promatrati izvan društveno-kulturnog okvira u kojem djeluju i da ih moramo promatrati isključivo kao oblike društveno-kulturne interakcije.

2.4.1. Medijski utjecaj

Medijske su prednosti različitih medija relativne i ne postoji jedan medij koji bi potpuno zasjenio druge. Internet je učinio novine, televiziju i radio manje bitnima, no nijedan medij nije toliko dominantan da bi istisnuo druge. Svaki je medij zanimljiv publici jer zadovoljava specifične potrebe konzumenata. Ljudi koriste medije na različite načine poput uživanja, druženja, praćenja i tumačenja sadržaja. Internetska pojava demokratizirala je i individualizirala medijske sadržaje te omogućila različite načine komunikacije i brzi prijenos informacija.

Prema Jurčić (2017: 133) s pojavom je interneta povjerenje u medijske izvore informacija oslabilo. Svi mediji podložni su tržišnim zakonima, a noviji mediji još više. Manipulacija medija postala je prisutna, a indikatori medijske manipulacije uključuju skraćivanje tema, površno izvještavanje, privlačenje pažnje bizarnostima i poticanje poruka o potrošnji kao primarnoj ljudskoj potrebi. Promjene u znanstvenim perspektivama o utjecaju medija očite su, no i dalje među istraživačima postoji neslaganje oko toga jesu li medijski utjecaji pozitivni ili negativni. Sredstva masovne komunikacije imaju značajnu ulogu u suvremenom društvu i oblikovanju osobnosti.

Glavni je cilj medija zadovoljiti potrebe informiranja o društvenim procesima te prenijeti informativne, edukativne i zabavne sadržaje koji utječu na pojedince. Pozitivan utjecaj medija

ogleda se u stjecanju znanja, razvoju vještina, usvajanju kulturnog nasljeđa i moralnih vrijednosti te razvoju kreativnosti.

S druge strane, mediji također imaju negativan utjecaj poticanjem nasilja, konzumerizma i kriminala. Stavovi znanstvenika o utjecaju medija i dalje su podijeljeni. Na primjer, prema Jurčić (2017: 133) Frankfurtska je škola tvrdila da mediji zaglupljuju mase, a postmodernistički teoretičar Baudrillard (2001: 15) istaknuo je da mediji smještaju suvremenog čovjeka u svemir „simulakra“, narušavajući vlastiti osjećaj stvarnosti. Proučavanje medijskog utjecaja počelo je još početkom 20. stoljeća i mijenjalo se tijekom vremena. Važan je čimbenik u istraživanju publika jer utjecaj medija ovisi i o pasivnosti okoline te osobnoj reakciji pojedinca. Mediji djeluju na više razina: učenje, oblikovanje stavova, emocionalno djelovanje i fiziološke reakcije.

Prema Benabdelouahedu i Dakouanu (2020: 1–6) umjetna inteligencija može manipulirati medijima na različite načine, uključujući:

1. Generiranje lažnih vijesti i članaka – umjetna inteligencija može generirati vijesti, članke i čak cijele web stranice koje su potpuno lažne ili djelomično točne, ali s namjerom manipulacije čitateljima.
2. Manipulacija naslova i podnaslova – umjetna inteligencija može analizirati trendove u čitanju naslova podnaslova te generirati one koji privlače pozornost ljudi. Ovaj proces naziva se *clickbait* i često se koristi za povećanje broja klikova na web stranicama.
3. Personalizacija sadržaja – umjetna inteligencija može analizirati korisničke podatke i prilagoditi sadržaj koji se prikazuje svakom pojedinom korisniku, što može utjecati na njihovu percepciju stvarnosti.
4. Automatizirani botovi – umjetna inteligencija može stvoriti automatizirane botove koji mogu širiti lažne informacije putem društvenih medija i drugih online platformi.
5. Utjecaj na algoritme pretraživanja – umjetna inteligencija može utjecati na algoritme pretraživanja, što može utjecati na to koji će se sadržaj pojaviti na vrhu rezultata pretraživanja. Ovaj proces naziva se *search engine optimization* ili SEO i koristi se za poboljšanje vidljivosti web stranica.
6. Analiza podataka – umjetna inteligencija može analizirati velike količine podataka o čitateljima, korisnicima društvenih medija i drugim izvorima, što omogućuje medijskim tvrtkama da stvore ciljane marketinške kampanje koje mogu manipulirati čitateljima.

Dakle, mediji imaju dualnu narav – mogu pozitivno poticati ponašanje, razvoj društva, toleranciju i maštu, ali također imaju negativne aspekte. Važno je priznati da mediji nisu jedini krivci za negativnosti i da je njihov utjecaj nužno promatrati unutar okvira djelovanja. Mediji nisu isključivo štetni ili korisni, već se mogu manifestirati kao oboje. Oni imaju moć jačanja i rušenja naših stavova, utječu na emocije te mogu izazvati razne reakcije poput radosti, tuge ili poticanja na akciju. Međutim, mediji imaju potencijal biti i koristan izvor informacija i zabave. Oba načina utjecaja oblikuju osobna stajališta i identitet pojedinca.

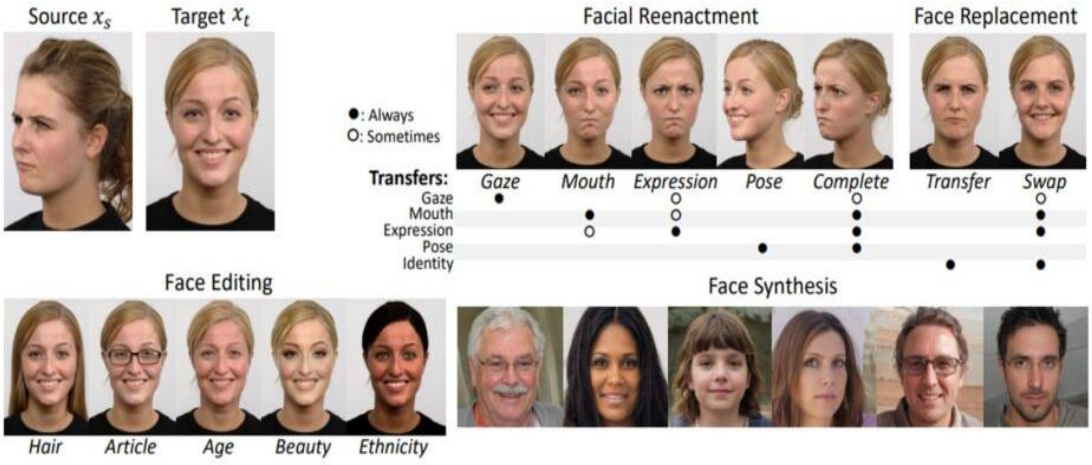
2.5. DUBOKI LAŽNI VIDEOZAPISI – „DEEPFAKES“

Raspoznavanje dubokih laži (*Deepfakes*) postalo je jedno od istaknutih istraživačkih pitanja. Znanstvenici su uložili mnogo truda u pronalaženje rješenja za ovu izazovnu temu. Često su pristupi rješavanju koristili vidljive znakove koji su česti kod većine dubokih falsifikacija. Međutim, najuspješniji pristupi temelje se na suptilnim indikatorima kao što su treptaji očiju, neskladni profili boja te nepravilnosti u izobličenjima lica. Ovi znakovi omogućuju visoku preciznost u otkrivanju velikog broja dubokih falsifikacija, posebno starijih verzija. Unatoč tomu, ova problematika kompleksnija je nego što se čini na prvi pogled. Ona zahtijeva analizu dubokih veza i korelacija koje idu izvan vidljivih znakova. Stoga su se pojavili pristupi koji se temelje na direktnoj klasifikaciji lažnog sadržaja putem sofisticiranih algoritama strojnog i dubokog učenja. Sveukupno, borba protiv dubokih falsifikacija zahtijeva dublje razumijevanje i inovativne metode kako bismo se uspješno suočili s ovim izazovom. (Ivanov et al., 2020: 326)

Općenito, pojam „Deepfakes“ obuhvaća sve digitalno krivotvorene sadržaje koji su stvoreni korištenjem tehnika dubokog učenja. Ovaj izraz nastao je nakon što je korisnik Reddita pod nazivom „Deepfakes“ tvrdio krajem 2017. godine da je razvio algoritam strojnog učenja koji mu je omogućio zamjenu lica poznatih osoba u pornografskim videima. Najproblematičnije primjene Deepfakesa uključuju lažnu pornografiju, širenje lažnih vijesti, izazivanje prijevara te financijske prijevare. Kao posljedica toga, područje istraživanja koje se tradicionalno bavilo općom forenzikom medija sada dobiva novi zamah i posvećuje sve više napora otkrivanju manipulacija lica na slikama i u videima. (Tolosana et al., 2022: 4)

Sve veći interes za otkrivanje lažnih lica ilustriran je povećanjem broja radionica na uglednim konferencijama, međunarodnim projektima poput MediFora koji financira Agencija za napredna istraživanja obrane (DARPA) te natjecanjima kao što je Izazov za forenziku medija (MFC2018) koje je pokrenuo Nacionalni institut za standarde i tehnologiju (NIST). Također, treba spomenuti i Izazov za detekciju Deepfakeova (DFDC) koji je inicirao Facebook, kao i nedavni projekt DeeperForensics. (Tolosana et al., 2022: 4) Ovaj projekt nastavlja istraživanje i razvoj tehnika za detekciju manipulacija lica koristeći najnovije spoznaje iz područja dubokog učenja i računalnog vida. Svi ovi napori i inicijative ukazuju na sve ozbiljniji pristup rješavanju izazova koje predstavljaju manipulacije digitalnim sadržajem, posebno u vezi s lažnim licima.

Također, prije daljnjeg zadiranja u tematiku deepfakeova i vrsta manipulacije istima, važno je točno odrediti definiciju manipulacije. Dakle, manipulacija (prema lat. *manipulus*: rukovet, svežanj) je rukovanje, raspolaganje, upravljanje, upotrebljavanje, stručan način rada; u prenesenom značenju, upravljanje ljudskim ponašanjem; iskorištavanje ljudskih navika i sklonosti u određene svrhe; vješto varanje, podvaljivanje. (Hrvatska enciklopedija, 2021)



Slika 2. Načini rada duboko lažiranih slika ili videozapisa

Izvor: https://datahacker.rs/009-the-creation-and-detection-of-deepfakes-a-survey/

2.5.1. Vrste deepfakeova

1. Cjelokupna sinteza lica

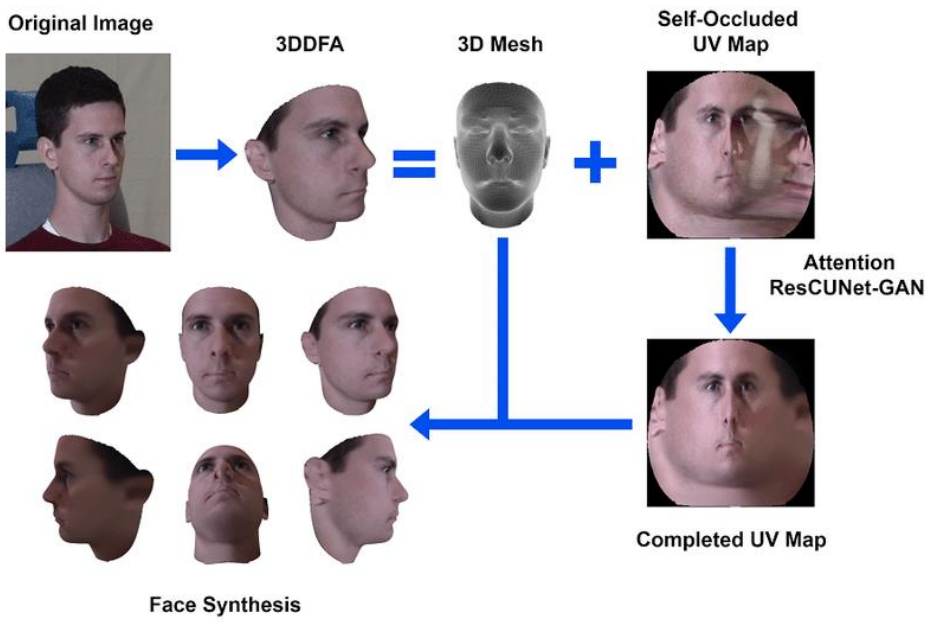
Manipulacije cijelih lica ostvaruju se uz pomoć snažnih GAN-ova (Generative Adversarial Networks). Općenito, GAN se sastoji od dviju različitih neuronskih mreža koje se nadmeću u minimax igri (Tolosana et al., 2022: 5):

- Generator G koji hvata distribuciju podataka i stvara nove primjerke
- Diskriminator D koji procjenjuje vjerojatnost da primjerak dolazi iz skupa podataka za treniranje (stvarni) umjesto iz G (lažni).

Postupak treniranja za G je maksimiziranje vjerojatnosti da D napravi pogrešku, čime se stvaraju visokokvalitetni lažni primjerci. Nakon postupka treniranja D se odbacuje, a G se koristi za stvaranje lažnog sadržaja. Ovaj koncept posljednjih godina intenzivno se primjenjuje za sintezu cijelih lica unaprjeđujući realističnost manipulacija. (Tolosana et al., 2022: 5) Postoje

različite javno dostupne baze podataka namijenjene istraživanju manipulacija cjelokupne sinteze lica. Zanimljivo je primijetiti da se svaka lažna slika može identificirati putem specifičnog otiska GAN-a, slično kao što se prirodne slike prepoznaju putem otiska temeljenog na uređaju (kao što je npr. PRNU). U stvarnosti se čini da ti otisci ovise ne samo o arhitekturi GAN-a već i o različitim primjenama istoga. (Tolosana et al., 2022: 6)

Predstavljena je i nova baza podataka nazvana iFakeFaceDB. Ova baza podataka sadržava 250 000 i 80 000 sintetičkih slika lica koje su prvotno stvorene pomoću tehnika StyleGAN-a i ProGAN-a, redom. Osim toga, kao dodatna karakteristika u usporedbi s prethodnim bazama podataka, kako bi se otežala detekcija lažnih slika, primijenjen je pristup nazvan GANprintR (uklanjanje otiska GAN-a). Ovaj pristup omogućio je uklanjanje otisaka koji su proizvedeni od strane GAN arhitektura, dok je istovremeno očuvana vrlo realistična pojava slika. (Tolosana et al., 2022: 8)



Slika 3. Prikaz sinteze lica

Izvor: https://www.researchgate.net/figure/A-pipeline-process-of-face-synthesis-Using-3DDFA-to-obtain-a-3D-mesh-and-an-incomplete_fig1_345215577

2.5.2. Zamjena identiteta

Ova manipulacija uključuje zamjenu lica jedne osobe u videu (izvor) licem druge osobe (cilj). Za razliku od cjelokupne sinteze lica, gdje se manipulacije izvode na razini slika, kod zamjene identiteta cilj je stvoriti realistične lažne videozapise. (Tolosana et al., 2022: 8)

Od javno dostupnih baza lažnih sadržaja poput UADFV baze podataka pa sve do najnovijih Celeb-DF, DFDC, DeeperForensics-1.0 i WildDeepfake baza podataka, provedene su brojne vizualne nadogradnje koje su značajno povećale realizam lažnih videozapisa. Kao rezultat toga, baze podataka za zamjenu identiteta mogu se podijeliti u dvije različite generacije.

Prva generacija:

- UADFV (2018)
- DeepfakeTIMIT (2018)
- FaceForensics++ (2019)

Druga generacija:

- DeepFakeDetection (2019)
- Celeb-DF (2019)
- DFDC Preview (2019)
- DFDC (2020)
- DeeperForensics-1.0 (2020)
- WildDeepfake (2020)

2.5.3. Sinteza izmiješanih lica

Sinteza izmiješanih lica predstavlja digitalnu manipulaciju lica koja omogućuje stvaranje umjetnih biometrijskih uzoraka lica koji su slični biometrijskim podacima dviju ili više osoba. Ovo znači da bi nova slika lica dobivena ovom metodom uspješno prošla provjeru uspoređivanja s uzorcima lica tih dviju ili više osoba, stvarajući ozbiljnu prijetnju sustavima za prepoznavanje lica. (Tolosana et al., 2022: 13)

Slika 4. Prikaz sinteze lica

Izvor: https://www.gao.gov/blog/deconstructing-deepfakes-how-do-they-work-and-what-are-risks

Sinteza izmiješanih lica je tehnika koja transformira jedno lice u drugo uz glatki prijelaz. Ovo se često koristi u umjetnosti i zabavi te u istraživanjima emocija i percepcije. Algoritmi se koriste za kombiniranje značajki dvaju lica stvarajući novu sliku koja izgleda prirodno. Ova tehnika ima mnogo primjena, od zabave do forenzike. No postoji i tamna strana. Kriminalci je mogu zloupotrijebiti za prijevare prepoznavanja lica i stvaranje lažnih identifikacija. Stručnjaci rade na alatima za otkrivanje ovih manipulacija i educiranju javnosti o opasnostima. Ovo je dinamično područje koje se razvija kako bi se suočilo s naprednim tehnikama manipulacije licem. Sinteza izmiješanih lica općenito se opisuje kao bešavni prijelaz koji transformira jednu sliku lica u drugu. Početno je morfiranje predstavljalo tehniku generiranja slika za potrebe primjene u računalnoj grafici ili u psihološkim istraživanjima. Međutim, tek su u posljednjim godinama postale jasne njegove potencijalne i ozbiljne sigurnosne prijetnje za sustave za prepoznavanje lica (FRS). (Ferrara i Franco, 2022: 117) Ovaj postupak može omogućiti stvaranje lažnih slika koje se mogu uspješno prikazati kao prave osobe, što predstavlja značajan izazov za očuvanje sigurnosti i povjerenja u tehnologije prepoznavanja lica.

Kada je riječ o napadima, uspjeh napada ovisi o tome da izmiješana slika ispunjava dva ključna uvjeta. Prvo, mora zavarati stručnjaka za prepoznavanje tako da izmiješano lice izgleda izrazito slično osobi koja traži dokument, a na slici ne smiju biti prisutni nikakvi elementi koji bi mogli izazvati sumnju, kao što su tragovi manipulacije. Drugo, slika mora prevariti i automatizirani sustav za prepoznavanje identiteta koji se koristi za provjeru identiteta. To znači da izmiješano lice mora biti dovoljno uvjerljivo da se uspješno poklopi s identitetom i kriminalca i suradnika. Ovaj dvostruki zahtjev čini ovakav napad izazovnim i zahtjevnim. (Ferrara i Franco, 2022: 118) Napadi pomoću sinteze izmiješanih lica postaju sve ozbiljnija prijetnja, posebno u kontekstu

sigurnosti i prepoznavanja identiteta. Kao odgovor na ovakve prijetnje istraživači razvijaju nove tehnike i alate za otkrivanje ovakvih manipulacija. Također se radi na poboljšanju svijesti javnosti kako bi se minimizirala ranjivost na ovakve napade. Ovo područje istraživanja ostaje dinamično i nastavlja se razvijati kako bi se suočilo s rastućim izazovima koje donosi tehnologija manipulacije lica.

2.5.4. Manipulacija atributima

Ova manipulacija, također poznata kao uređivanje lica ili dotjerivanje lica, uključuje modifikaciju određenih atributa lica poput boje kose ili kože, spola, dobi, dodavanja naočala itd. Obično se ovaj postupak manipulacije provodi putem GAN-ova, kao što je StarGAN pristup koji je predložen u rad. Primjer je ovakve manipulacije popularna mobilna aplikacija FaceApp. Potrošači mogu koristiti tu tehnologiju kako bi isprobali raznovrsne proizvode kao što su kozmetika, šminka, naočale ili frizure u virtualnom okruženju. (Tolosana et al., 2022: 15)

Slika 4. Prikaz manipulacije atributima

Izvor: https://www.gao.gov/blog/deconstructing-deepfakes-how-do-they-work-and-what-are-risks

2.5.5. Zamjena izraza

Ova manipulacija, također poznata kao reenactment lica, uključuje promjenu izraza lica subjekta. Iako se u literaturi predlažu različite tehnike manipulacije, na primjer, na razini slike putem popularnih GAN arhitektura, u ovoj skupini usredotočujemo se na najpoznatije tehnike kao što su Face2Face i NeuralTextures. Ove tehnike zamjenjuju izraz lica jednog subjekta u videozapisu izrazom lica drugog subjekta. Face2Face tehnika omogućuje preciznu sinkronizaciju izraza lica između izvornog i ciljnog subjekta, dok NeuralTextures koristi tehniku generiranja tekstura kako bi stvorio uvjerljive izraze lica na ciljnom subjektu. Ova vrsta

manipulacije može imati široku primjenu u različitim područjima, uključujući zabavnu industriju, digitalne efekte u filmovima te istraživanje emocionalne komunikacije putem digitalnih medija. (Tolosana et al., 2022: 17)

Slika 5. Prikaz zamjene izraza lica

Izvor: https://vimeo.com/blog/post/video-deepfakes/

2.5.6. Pristupi usmjereni na otkrivanje manipulacija

U posljednjim godinama intenzivno se istražuje razvoj metoda za pouzdanu provjeru integriteta slika. Neki alati otkrivaju fizičke nesukladnosti, kao što su sjene, osvjetljenje ili perspektiva, koje također mogu biti primijećene od strane pažljivih promatrača. Međutim, u većini slučajeva pažljivo izrađene krivotvorine ne ostavljaju vidljive tragove i izgledaju semantički ispravno. Ipak, digitalne manipulacije obično mijenjaju temeljnu statistiku izvornog izvora, ostavljajući tragove koji, iako nevidljivi oku, mogu biti analizirani alatima na razini piksela. U stvarnosti, svaku sliku karakterizira niz značajki koje ovise o različitim fazama njezine povijesti, počevši od samog procesa akvizicije, preko unutarnje obrade kamere (npr. demosaiciranje i kompresija) pa sve do svih vanjskih obrada i operacija uređivanja. Stoga, proučavanjem mogućih odstupanja tih značajki od očekivanog ponašanja, moguće je s visokom pouzdanošću utvrditi je li integritet slike narušen. (Cozzolino i Verdoliva, 2022: 46) Otkrivanje manipulacija slikama predstavlja izazov koji zahtijeva multidisciplinarni pristup, uključujući računalnu viziju, statistiku, forenziku i duboko učenje, kako bi se osigurala autentičnost i povjerenje u digitalnom okruženju.

Najpopularniji pristupi usmjereni na otkrivanje manipulacija slikama traže artefakte koji su povezani s procesiranjem unutar kamere (dokazi temeljeni na kameri) ili s poviješću obrade izvan kamere (dokazi temeljeni na uređivanju). Ovi pristupi često slijede model koji se obično oslanja na statističke analize ili se temelje na ručno izrađenim značajkama te koriste klasične alate strojnog učenja. Konkretno, pristupi koji se fokusiraju na dokaze temeljene na kameri traže znakove koji proizlaze iz samog procesa snimanja i obrade slike unutar kamere. To mogu biti artefakti kao što su kompresija, demosaiciranje ili nesukladnosti svjetla i sjene. S druge strane, pristupi temeljeni na uređivanju analiziraju tragove koje je ostavila bilo koja vanjska obrada ili manipulacija slike nakon što je snimljena. To uključuje promjene boje, kontrasta, oštrine te eventualne dodatke ili brisanja objekata. (Cozzolino i Verdoliva, 2022: 47)

Ovi pristupi često koriste matematičke modele i statističke analize kako bi identificirali odstupanja od očekivanog ponašanja. Također, mogu se koristiti ručno izrađene značajke koje se odnose na specifične obrasce ili karakteristike manipuliranih slika. No svaka metoda temelji se na svojim vlastitim pretpostavkama koje ne moraju uvijek vrijediti za sve vrste manipulacija, što može ograničiti njihovu primjenjivost na određene situacije. S obzirom na stalno napredovanje tehnologije i sofisticiranost manipulacija, istraživači i stručnjaci nastavljaju razvijati nove metode i alate kako bi se otkrivanje manipulacija učinilo što preciznijim i pouzdanijim.

2.6. STVARANJE I SPRJEČAVANJE DEEPFAKEOVA

Nedavna varijacija uznemirujućeg problema *online* dezinformacija krivotvoreni su videozapisi stvoreni pomoću tehnologija umjetne inteligencije, posebno dubokih neuronskih mreža (DNN). Iako izmišljanje i manipulacija digitalnim slikama i videozapisima nisu novi fenomeni, upotreba DNN-ova znatno je olakšala i ubrzala proces stvaranja uvjerljivih lažnih videozapisa. Posebno značajan tip lažnih videozapisa baziranih na DNN-ovima, općenito poznat kao Deepfakeovi, nedavno je privukao veliku pažnju. U Deepfake videozapisima lica ciljane osobe zamjenjuju se licima druge osobe (donora), koja su sintetizirana putem DNN modela. Pritom se zadržavaju izrazi lica i položaji glave ciljane osobe. Budući da su lica intrinzično povezana s identitetom, pažljivo izrađeni Deepfakeovi mogu stvarati iluzije prisutnosti i aktivnosti osobe koje se u stvarnom životu ne događaju. Ova pojava može imati ozbiljne posljedice na političko, društveno, financijsko i pravno područje jer može izazvati duboke zablude i dezinformacije koje se teško mogu ispraviti. S obzirom na brzi razvoj tehnologije, borba protiv ovih manipulacija postaje izazovna, zahtijevajući sve veću suradnju i inovacije kako bi se sačuvala vjerodostojnost i integritet digitalnih informacija. (Li et al., 2022: 72)

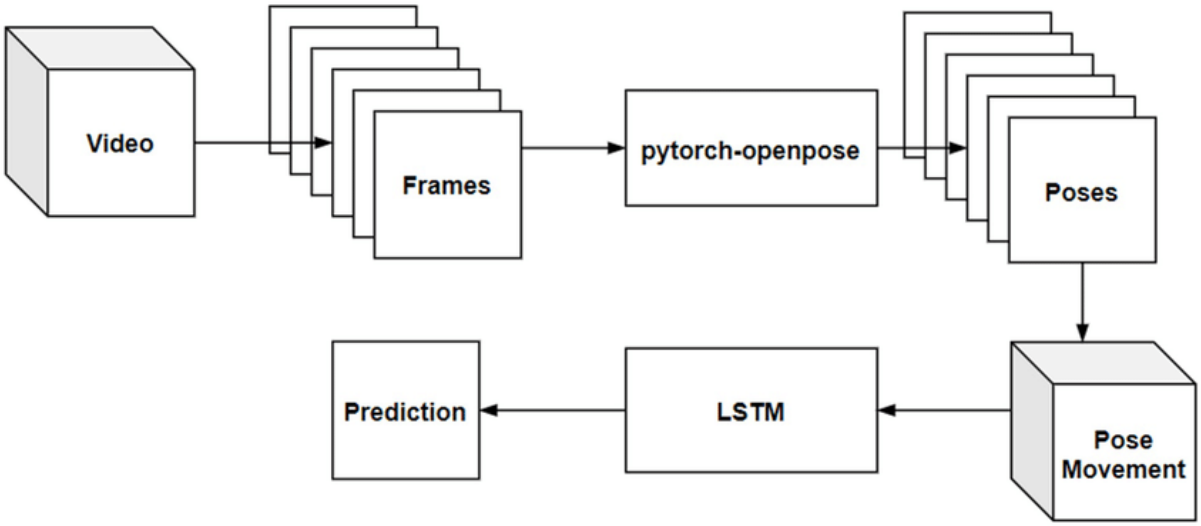
Kada se pažljivo promotre Deepfake videozapisi u postojećim skupovima podataka, primjećuju se značajni kontrasti u vizualnoj kvaliteti u usporedbi s stvarnim Deepfake videozapisima koji se pojavljuju na internetu. Uobičajeni vizualni artefakti koji se pojavljuju u tim skupovima podataka, uključujući loše kvalitetno sintetizirana lica, vidljive granice spajanja, nesklad boja, vidljive dijelove izvornog lica i neujednačene orijentacije sintetiziranih lica. Ovi artefakti vjerojatno su posljedica nepotpunih koraka u postupku sinteze te nedostatka pažljivog odabira sintetiziranih videozapisa prije nego što su uključeni u skupove podataka. (Li et al., 2022: 72)

Nadalje, Deepfake videozapisi s takvom niskom vizualnom kvalitetom teško mogu stvoriti uvjerljiv dojam i vjerojatno neće imati značajan stvarni utjecaj. Stoga, visoka uspješnost detekcije na tim skupovima podataka možda neće biti posebno relevantna kada se detekcijske metode primijene u stvarnom okruženju. Ovaj fenomen ukazuje na važnost razvoja i testiranja detekcijskih alata na stvarnim Deepfake videozapisima kako bi se bolje nosili s izazovima u stvarnom svijetu.

2.6.1. Sprječavanje manipulacije

„Bijela-okvir metoda“ za ometanje stvaranja Deepfakeova temelji se na poremećaju ekstrakcije karakterističnih točaka lica, poznata kao „Landmark Breaker“. Karakteristične točke lica

predstavljaju ključne lokacije važnih dijelova lica, kao što su vrhovi i srednje točke očiju, nosa, usta, obrva i konture. „Landmark Breaker“ strategija je koja cilja na karakteristične točke lica kako bi ometala proces stvaranja Deepfakeova.



Slika 6. Prikaz „Bijele-okvir metode“

Izvor: https://www.researchgate.net/figure/The-proposed-deepfake-detection-method-The-proposed-method-will-extract-the-frames-30_fig3_351187931

Ova metoda uključuje dodavanje adverzalnih perturbacija napadačima karakterističnih točaka lica. Adverzalne perturbacije namjerno su oblikovani šumovi na slici koji su osmišljeni kako bi zavarali duboke neuronske mreže (DNN) koje izdvajaju karakteristične točke lica. Time se stvara nepredvidljivost i nesigurnost u procesu ekstrakcije karakterističnih točaka, što otežava stvaranje preciznih i uvjerljivih videa. (Li et al., 2022: 73) Ova tehnika daje dodatnu zaštitu i otežava napadačima da precizno odrede karakteristične točke lica pri izradi ovog tipa videa. To dalje pomaže u otkrivanju i sprječavanju manipulacija na temelju karakterističnih točaka lica, čime se povećava integritet digitalnih sadržaja i smanjuje potencijalni utjecaj lažnih informacija.

2.7. DALJNJE ISTRAŽIVANJE I RAZVOJ TEHNIKA

Suvremeni tehnološki napretci u automatskom uređivanju video i audiosadržaja, korištenju generativnih suparničkih mreža (GAN) te raširenoj uporabi društvenih medija omogućili su brzu izradu i širenje visokokvalitetnih videozapisa koji su digitalno prerađeni. Ova situacija dovela je do pojave namjernih dezinformacija koje su često nazvane „lažne vijesti“, a koje imaju značajan utjecaj na politički pejzaž u nekoliko zemalja.

Posebno zabrinjavajuća pojava nedavni su videozapisi, često neprikladnog sadržaja, u kojima se lica osoba mogu zamijeniti licima drugih pojedinaca putem neuronskih mreža, poznatih kao Deepfakeovi. Ovakvi videozapisi izazivaju ozbiljne zabrinutosti u javnosti. Dostupan i lako dostupan softver otvorenog koda omogućava svakome da stvori takve zamjene lica, što rezultira velikim brojem sintetski generiranih videozapisa koji se brzo šire putem društvenih medija i vijesti. Ova situacija postavlja značajan tehnički izazov za otkrivanje i filtriranje takvog sadržaja. Navedeni problemi izražavaju potrebu za daljnjim istraživanjem i razvojem tehnika za suzbijanje širenja dezinformacija i manipulacije digitalnim sadržajem. (Korshunov i Marcel, 2022: 97)

Trenutno, područje istraživanja vezano uz deepfakeove još uvijek je relativno nezrelo, no osnovna istraživačka pitanja već su jasno postavljena (Korshunov i Marcel, 2022: 98–113):

1. Povećanje raznolikosti podataka – istraživači se pitaju kako povećati broj dostupnih podataka s različitim tipovima deepfakeova. Raznolikost u podatcima omogućava razvoj boljih detekcijskih metoda koje su otpornije na različite varijacije manipulacija.
2. Prijevara automatiziranog prepoznavanja lica – postavlja se pitanje može li tehnologija deepfakeova prevariti alate za automatizirano prepoznavanje lica. Ako deepfakeovi postanu dovoljno napredni, mogli bi postati izazov za sustave za prepoznavanje lica koji se oslanjaju na karakteristike lica.
3. Prijevara ljudskog vizualnog sustava – interesira istraživače mogu li deepfakeovi dovoljno uvjerljivo prevariti ljudsko oko i vizualni sustav. Ako su deepfakeovi tako realistični da prosječni promatrači ne mogu razaznati manipulaciju, to bi moglo imati ozbiljne implikacije u kontekstu dezinformacija i prijevare.
4. Učinkovita detekcija deepfakeova – veliko je pitanje može li se razviti pouzdana i učinkovita metoda za detekciju deepfakeova. Otkrivanje naprednih manipulacija u

videozapisima izazov je s obzirom na brzinu razvoja tehnologije stvaranja deepfakeova.

S obzirom na brz razvoj tehnologije i njezin potencijalni utjecaj na društvo, istraživanje ovih pitanja ključno je za razumijevanje i suočavanje s izazovima koje deepfakeovi donose.

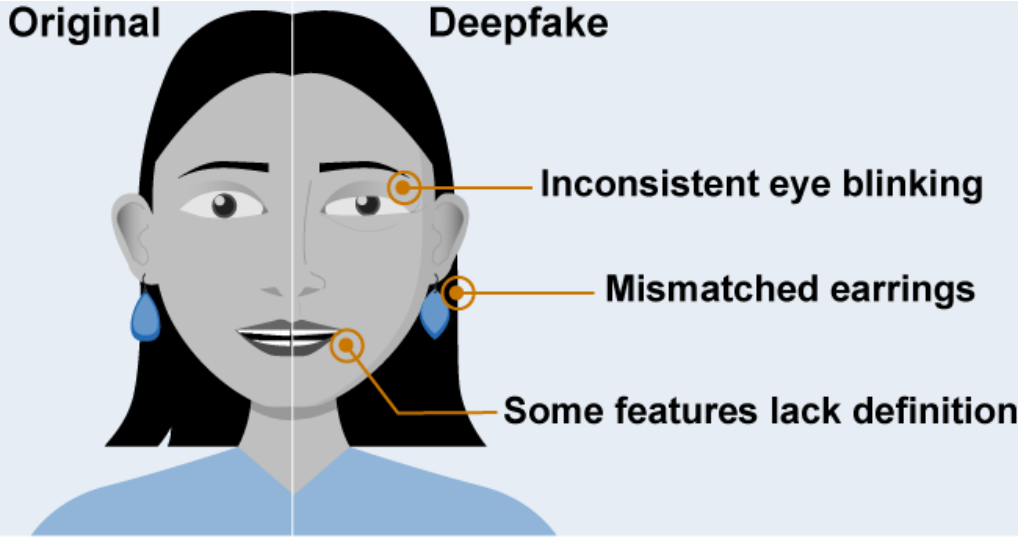
2.7.1. Prepoznavanje i napredak modela

Prepoznavanje lica ima široku primjenu u raznim područjima, kao što su biometrijska autentikacija, upravljanje građanskim identifikacijama te kontrola na granicama. Zahvaljujući nedavnom uspjehu dubokog učenja u području prepoznavanja, postignute su vrlo visoke performanse u biometrijskoj verifikaciji. Ovo je rezultiralo istraživanjem niza naprednih modela za prepoznavanje lica, kao što su VGGFace, Residual Networks (ResNet) i ArcFace. Duboko naučeni modeli usmjereni su na unapređenje biometrijskih performansi čak i u uvjetima teškog degradiranja kvalitete biometrijskih uzoraka, kao što su promjene položaja, osvjetljenja, izraza lica, starenje i različitosti među uzorcima. S poboljšanjem performansi, duboki modeli mogu se primijeniti na identifikaciju subjekata. To znači da se subjekt pretražuje unutar naučenih modela u okruženju gdje su dostupni samo unaprijed poznati uzorci. Također, duboki modeli mogu se koristiti za verifikaciju, gdje se značajke izdvajaju iz dviju slika te se potom uspoređuju kako bi se donijela odluka o njihovoj podudarnosti temeljenoj na prethodno definiranoj granici. (Xu et al., 2022: 140) Ovaj napredak u prepoznavanju lica ima značajne posljedice i izazove, kako u smislu sigurnosti tako i za očuvanje privatnosti i etičke implikacije. Kao rezultat toga, istraživanja se usredotočuju na razvoj sve sofisticiranijih metoda detekcije manipulacija, kao i na educiranje šire javnosti o potencijalnim rizicima i koristima ovih tehnologija.

Napredak u tehnologiji prepoznavanja lica donosi širok raspon mogućnosti i istovremeno postavlja izazove u različitim sektorima društva:

1. Primjena u sigurnosti
2. Biometrijska autentikacija
3. Privatnost i etički aspekti
4. Izazovi u detekciji manipulacija

Općepoznato je da tehnologija prepoznavanja lica nosi mnoge koristi, ali istovremeno postavlja važna pitanja o sigurnosti, privatnosti i etici. Bitno je postići ravnotežu između tehnološkog napretka i društvene odgovornosti kako bi se osiguralo da ova tehnologija donese dobrobit cijelom društvu.



Slika 7. Načini prepoznavanja lažiranog videa

Izvor: https://www.gao.gov/products/gao-20-379sp

2.8. MEDIJI I DUBOKI LAŽNI VIDEOZAPISI

Odnos medija i deepfake videa u kontekstu manipulacije nad medijima i vjerodostojnosti predstavlja izazovno i složeno područje koje duboko utječe na našu sposobnost razlikovanja istinitih informacija od lažnih. Deepfake tehnologija, koja omogućuje izradu uvjerljivih lažnih video sadržaja, postavlja ozbiljna pitanja o integritetu medija i povjerenju javnosti u informacije koje konzumira.

Izvještaj Future Advocacyja (Dokler, 2019) ističe namjeru objavljivanja deepfake videa kao sredstvo za šokiranje i humoristično informiranje javnosti te kako bi se izvršio pritisak na zakonodavce. Kao ključne poteškoće povezane s deepfake videima navode se:

1. Identifikacija deepfake videa: postavlja se pitanje je li društvo sposobno otkriti deepfake videa, bilo tijekom prijenosa ili nakon što su već široko rasprostranjeni putem interneta.
2. Izazovi autentičnosti: fenomen gdje se autentičan videomaterijal može proglasiti deepfakeom, što dodatno komplicira prepoznavanje istinitih sadržaja.
3. Regulacija: postavlja se pitanje kakva ograničenja treba primijeniti na stvaranje deepfake videa i može li se ta regulacija praktično provoditi.
4. Upravljanje posljedicama: kako se nositi s utjecajem deepfake videa kad regulacija nije dovoljna te gdje leži odgovornost za minimiziranje štete.

Iako zabrinutost postoji zbog potencijalne političke zloupotrebe tehnologije, većina objavljenih deepfake videa uključuje pornografski sadržaj. Istraživanje tvrtke Deeptrace iz 2019. identificiralo je 14 678 deepfake videa, a čak 96 % njih bili su pornografski sadržaji stvoreni bez pristanka osoba koje su uključene. (Dokler, 2019) Prema Dokler (2019), s obzirom na politički aspekt, istraživanje također ističe da je svijest o deepfake videima već destabilizirala političke procese jer narušava vjerodostojnost videozapisa u kojima se pojavljuju političari. Također, deepfake videi mogu utjecati na cyber sigurnost omogućujući nove načine napada te potencijalno ojačavajući tradicionalne cyber prijetnje.

U zaključku, istraživanje sugerira da je utjecaj deepfake videa već prisutan na globalnoj razini te se naglašava potreba za poduzimanjem mjera kako bi se pojedinci i organizacije zaštitili od potencijalnih zlouporaba.

Mediji kao glavni nositelji informacija igraju ključnu ulogu u oblikovanju percepcije stvarnosti. No kada deepfake tehnologija ulazi u igru, granica između stvarnog i lažnog postaje zamagljena. Manipulacija medija putem deepfake videa može imati ozbiljne posljedice.

Dakle, može se zaključiti da glavni problemi nastaju zbog sljedećih uzroka:

1. Povjerenje i vjerodostojnost: deepfake videi mogu snažno utjecati na povjerenje javnosti u medije. Kada ljudi nisu sigurni je li video autentičan ili je rezultat manipulacije, povjerenje u medije kao izvore informacija značajno se smanjuje.
2. Brza širenja dezinformacija: deepfake videi mogu se brzo i lako dijeliti putem društvenih mreža i drugih online platformi. Ovo olakšava širenje dezinformacija i lažnih priča koje mogu imati ozbiljne posljedice na društvo.
3. Ugrožavanje ugleda pojedinaca: deepfake videi mogu ozbiljno naštetiti ugledu pojedinaca ili javnih figura. Lažni videomaterijali mogu prikazivati ljude kako izgovaraju ili čine stvari koje nikada nisu učinili, što može imati dugoročne posljedice na njihovu karijeru i život.
4. Teškoće u detekciji: neki deepfake videi toliko su vješto izrađeni da ih je teško razlikovati od stvarnih videosadržaja. To otežava medijima i korisnicima da pouzdano prepoznaju lažne sadržaje.
5. Potreba za novim alatima i protokolima: razvoj deepfake tehnologije iziskuje potrebu za razvojem novih tehnika za detekciju i verifikaciju autentičnosti videosadržaja. Mediji moraju uložiti napore u edukaciju svojih novinara i urednika kako bi se adekvatno nosili s ovim izazovom.
6. Etika i odgovornost: korištenje deepfake tehnologije za manipulaciju medijima postavlja pitanja etičke odgovornosti novinara, izdavača i korisnika. Mediji moraju pažljivo razmotriti kako će se nositi s ovom tehnologijom i kako će osigurati da njihovi sadržaji ostanu vjerodostojni.

U cjelini, odnos između medija i deepfake videa reflektira duboke promjene u načinu na koji percipiramo i konzumiramo informacije. Za medije je važno razviti strategije za zaštitu integriteta svojih sadržaja i povjerenja publike, dok istodobno koriste nove tehnologije za stvaranje inovativnih i autentičnih informacija.

2.9. „VOLODYMYR ZELENSKY“ INCIDENT

Moguće posljedice svega dosad navedenog prikazat će se kroz analizu konkretnog slučaja lažnog videosnimka koji je zahvatio ukrajinsku javnost i stvorio ozbiljne posljedice u kontekstu političke situacije. Konkretno, događaj u kojem je ukrajinska vlada unaprijed upozorila na mogućnost deepfake videa u kojem se ukrajinski predsjednik Volodymyr Zelensky prikazuje kako najavljuje predaju tijekom ruske invazije. Nakon toga pojavio se lažni video na društvenim mrežama i medijima, što je izazvalo brzu reakciju predsjednika i razne mjere na platformama kako bi se ograničilo širenje lažnog sadržaja.

Dakle, na početku ožujka 2022. godine ukrajinska je vlada upozorila da bi neprijatelji mogli stvoriti deepfake video u kojem bi se prikazalo da predsjednik Volodymyr Zelensky najavljuje predaju Rusiji tijekom invazije. Nedugo zatim, na društvenim mrežama poput Facebooka i YouTubea pojavio se lažan video (na slici 8. lijevo je lažno, desno je stvarno) u kojem nepomična verzija Zelenskyja poziva ukrajinske vojnike da odlože oružje, koristeći glas koji se razlikuje od njegova uobičajenog tona.

Situacija je pokazala kako je brza reakcija i sprječavanje širenja lažnih sadržaja moguća kada postoji odgovarajuća priprema i reakcija. Zelensky je imao koristi od svoje uloge u vlasti koja se unaprijed pripremila za takve deepfake napade.

Slika 8. Prikaz „Volodymyr Zelensky“ incidenta

Izvor: https://news.virginia.edu/content/qa-zelenskyy-surrender-hoax-feared-future-deepfakes-here

Njegova brza reakcija na lažni video, podrška televizijske stanice i društvenih mreža zajedno su smanjili širenje videa. Ova situacija pokazuje strategije koje su uvedene kao odgovor na nove prijetnje poput političkih deepfake sadržaja. Iako je ovaj incident brzo razotkriven, deepfake tehnologija i dalje predstavlja ozbiljnu prijetnju. Kako deepfake sadržaji postaju uvjerljiviji i lakši za izradu, politički lideri i druge javne osobe mogu biti meta ovakvih napada. Ovaj incident ukazuje na potrebu za stalnim razvojem tehnologija za prepoznavanje i suzbijanje deepfake sadržaja kako bi se sačuvalo povjerenje i autentičnost informacija.

„On zagađuje informacijski ekosustav i baca sjenu na sve sadržaje, koji se već suočavaju s kompleksnim maglom rata“, rekao je. „Sljedeći put kad predsjednik izađe na televiziju, neki ljudi bi mogli pomisliti, ‚Čekaj malo – je li ovo stvarno?‘ “ – izjavio je jedan od istraživača za intervju. (Allyn, 2022)

Kroz analizu ovog slučaja može se uvidjeti kako se deepfake tehnologija može iskoristiti kao sredstvo dezinformacija, posebno u kriznim situacijama poput oružanih sukoba. Također, koliko je zapravo bitna brza reakcija i priprema vlade, medija i društvenih platformi koje svojim postupcima mogu ograničiti štetne posljedice takvih deepfake sadržaja. U konačnici, ovaj slučaj daje nam uvid u kompleksne izazove vezane uz autentičnost medijskih sadržaja, povjerenje javnosti i sposobnost tehnologije da manipulira percepcijom stvarnosti.

3. ZAKLJUČAK

Umjetna inteligencija (UI) predstavlja revolucionarnu tehničku paradigmu koja izvorima svog potencijala preoblikuje mnoga područja ljudskog djelovanja, uključujući i medije. U ovom znanstvenom radu istražili smo kako se UI, posebno kroz tehnologiju deepfake videa, može koristiti za manipulaciju medijima te kakav utjecaj to ima na širu društvenu stvarnost.

U početnim poglavljima razradili smo ključne koncepte umjetne inteligencije. Strojno učenje i duboko učenje postali su vitalni dijelovi UI-a, omogućavajući računalima da uče iz podataka i obavljaju složene zadatke. Razvoj UI-a, koji obuhvaća razvoj slabih i jakih oblika, dodatno je naglasio dubinu sposobnosti koje računalni sustavi mogu postići. Prelazimo na medije, važan kanal informiranja i oblikovanja javnog mnijenja. Mediji imaju dvostruku ulogu: prenoseći informacije, oblikujući doživljaj stvarnosti. Njihov utjecaj seže dalje od puke komunikacije, obuhvaćajući i kulturne, političke i društvene aspekte. Duboki lažni sadržaji (deepfake videi) ključan su aspekt u ovom radu. Njihova sposobnost manipulacije putem precizne digitalne obrade i sinteze dovodi do ozbiljnih posljedica po integritet medijskih sadržaja. Utjecaj deepfake videa na političke procese i širenje dezinformacija postaje posebno zabrinjavajuć.

Manipulacija medijima putem deepfake videa predstavlja složen izazov. Tehnologija koja može stvoriti uvjerljive lažne sadržaje postavlja pitanja o vjerodostojnosti informacija koje konzumiramo. Povjerenje u medije i njihovu sposobnost prenošenja istinitih informacija smanjuje se. Pitanja otkrivanja deepfake videa, regulacije stvaranja takvih sadržaja te upravljanja štetom koju mogu uzrokovati postaju ključna u postizanju ravnoteže između tehnološkog napretka i zaštite društva. Kako tehnologija napreduje, nužno je pronaći održivo rješenje za prevladavanje izazova manipulacije medijima. Kombinacija tehnoloških inovacija, edukacije i regulatornog okvira mogla bi pružiti osnovu za suzbijanje negativnih učinaka deepfake videa na integritet medija i društva u cjelini. U svjetlu tih izazova, važno je stalno istraživati i razvijati mehanizme kako bi mediji ostali pouzdani izvori informacija te kako bi se spriječila manipulacija koja bi mogla ozbiljno ugroziti javnu sferu.

4. LITERATURA

1. Akerkar, R. (2019) *Artificial Intelligence for Business*. Dostupno, 31.8.2023 na: https://www.studocu.com/row/document/jamaa%D8%A9-alkahr%D8%A9/principle-of-accounting/akerkar-2019-book-artificial-intelligence-for-busin/24891126
2. Allyn, B. (2022) *Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn*. NPR News. Dostupno, 31.8.2023 na: https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia
3. Baudrillard, J. (2001) *Simulacija i zbilja*. Naklada Jesenski i Turk, Hrvatsko sociološko društvo. Dostupno, 31.8.2023 na: https://svetlogike.files.wordpress.com/2014/02/jeanbaudriilard-simulacija_i_zbilja.pdf
4. Benabdelouahed, R. & Dakouan, C. (2020) *The Use of Artificial Intelligence in Social Media: Opportunities and Perspectives*. Expert Journal of Marketing 8, 82–87. Published by Sprint Investify. Dostupno, 31.8.2023 na: https://marketing.expertjournals.com/23446773-806/
5. Biti, V. (1997) *Pojmovnik suvremene književne teorije*. Matica hrvatska, Zagreb. Dostupno, 31.8.2023 na: https://pdfcoffee.com/bitijev-pojmovnik-pdf-free.html
6. Brynjolfsson, E., McAfee, A. (2017) *The Business of Artificial Intelligence: How AI Fits into Your Science Team; What it can – and cannot – do for your organization*. Dostupno, 31.8.2023 na: https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence
7. Copeland, B. (2023) Artificial Intelligence. *Encyclopedia Britannica*. Dostupno. 31.8.2023 na: https://www.britannica.com/technology/artificial-intelligence
8. Cozzolino, D. & Verdoliva, L. (2022) "Chapter 3 (45–71): Multimedia Forensics Before the Deep Learning Era." In *Handbook of Digital Face Manipulation and Detection. From DeepFakes to Morphing Attacks*. Springer Cham. Dostupno, 31.8.2023 na: https://doi.org/10.1007/978-3-030-87664-7
9. Dokler, A. (2019) *Najveće opasnosti lažnih videa napravljenih uz pomoć umjetne inteligencije*. Problem dezinformacija. Dostupno, 31.8.2023 na: https://www.medijskapismenost.hr/najvece-opasnosti-laznih-videa-napravljenih-uz-pomoc-umjetne-inteligencije/

10. Europska komisija. (2018) *Umjetna inteligencija za Europu*. Bruxelles. Dostupno, 31.8.2023 na: https://eur-lex.europa.eu/legal-content/HR/TXT/HTML/?uri=CELEX:52018DC0237&from=ES
11. Ferrara, M. & Franc, A. (2022) "Chapter 6 (117–139): Morph Creation and Vulnerability of Face Recognition Systems to Morphing." In *Handbook of Digital Face Manipulation and Detection. From DeepFakes to Morphing Attacks*. Springer Cham. Dostupno, 31.8.2023 na: https://doi.org/10.1007/978-3-030-87664-7
12. Goertzel, B. & Wang, P. (2007) *Advances in Artificial General Intelligence: Concepts, Architectures, and Algorithms*. IOS Press. Dostupno, 31.8.2023 na: https://www.researchgate.net/publication/271390398_Artificial_General_Intelligence_Concept_State_of_the_Art_and_Future_Prospects
13. *Hrvatska enciklopedija*. Mrežno izdanje. (2021) Manipulacija. Leksikografski zavod Miroslav Krleža. Dostupno. 31.8.2023 na: http://www.enciklopedija.hr/Natuknica.aspx?ID=38638
14. *Hrvatska enciklopedija*. Mrežno izdanje. (2021) Umjetna inteligencija. Leksikografski zavod Miroslav Krleža. Dostupno, 31.8.2023 na: http://www.enciklopedija.hr/Natuknica.aspx?ID=63150
15. Ivanov, N. S., Arzhskov, A. V., Ivanenko, V. G. (2020) *Combining Deep Learning and Super-Resolution Algorithms for Deep Fake Detection*. Department of Computer Systems and Technologies, National Research Nuclear University MEPhI. Dostupno, 31.8.2023 na: https://ieeexplore.ieee.org/document/9039498
16. Jurčić, D. (2017) „TEORIJSKE POSTAVKE O MEDIJIMA – DEFINICIJE, FUNKCIJE I UTJECAJ“, Mostariensia, 21(1), str. 127–136. Dostupno, 31.8.2023 na: https://hrcak.srce.hr/190208
17. Jurčić, D. (2017) „TEORIJSKE POSTAVKE O MEDIJIMA – DEFINICIJE, FUNKCIJE I UTJECAJ“, Mostariensia, 21(1), str. 127–136. Dostupno, 31.8.2023 na: https://hrcak.srce.hr/190208
18. Kalpokas, I. (2021) *Problematising reality: the promises and perils of synthetic media*. SN Soc Sci 1. Dostupno na: https://doi.org/10.1007/s43545-020-00010-8
19. Korshunov, P. & Marcel, S. (2022) "Chapter 5 (97–117): The Threat of Deepfakes to Computer and Human Visions." In *Handbook of Digital Face Manipulation and Detection. From DeepFakes to Morphing Attacks*. Springer Cham. Dostupno, 31.8.2023 na: https://doi.org/10.1007/978-3-030-87664-7

20. Kovač, L. (2015) *Umjetna inteligencija danas*. Diplomski rad. Sveučilište u Rijeci. Dostupno, 31.8.2023 na: https://urn.nsk.hr/urn:nbn:hr:186:606497
21. Li, Y., Sun, P., Qi, H. & Lyu, S. (Year of Publication) "Chapter 4 (71–97): Toward the Creation and Obstruction of DeepFakes." In *Handbook of Digital Face Manipulation and Detection. From DeepFakes to Morphing Attacks*. Springer Cham. Dostupno, 31.8.2023 na: https://doi.org/10.1007/978-3-030-87664-7
22. McLuhan, M. (2008) *Razumijevanje medija*. Golden marketing – Tehnička knjiga. Dostupno, 31.8.2023 na: https://monoskop.org/images/1/15/McLuhan_Marshall_Razumijevanje_medija.pdf
23. Polonijo, B. (2020) *Primjena metoda dubokog učenja*. Diplomski rad. Veleučilište u Rijeci. Dostupno, 31.8.2023 na: https://urn.nsk.hr/urn:nbn:hr:125:964822
24. Polšek, D. (2003) *Zapisi iz treće kulture*. Prvo izdanje. Zagreb: Jesenski i Turk. Dostupno, 31.8.2023 na: https://elektronickeknjige.com/knjiga/polsek-darko/zapisi-iz-trece-kulture/impresum/
25. Putica, M. (2018) „UMJETNA INTELIGENCIJA: DVOJBE SUVREMENOGA RAZVOJA“. Hum, 13(20), str. 198–213. Dostupno, 31.8.2023 na: https://hrcak.srce.hr/219733
26. Rahman, A., Islam, M. M., Moon, M., Tasnim, T., Siddique, N., Shahiduzzaman, M. & Ahmed, S. (2022) "A Qualitative Survey on Deep Learning-Based Deep Fake Video Creation and Detection Method." Australian Journal of Engineering and Innovative Technology. Dostupno, 31.8.2023 na: https://www.researchgate.net/publication/358322160_A_Qualitative_Survey_on_Deep_Learning_Based_Deep_fake_Video_Creation_and_Detection_Method
27. Rockwell, A. (2017) *The history of artifical intelligence*. Special edition on AI, Harvard. Dostupno, 31.8.2023 na: https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/
28. Rose, D. (n. d.) *Artificial Intelligence for Business: What You Need to Know about Machine Learning and Neural Networks*. Dostupno, 31.8.2023 na: https://ptgmedia.pearsoncmg.com/images/9780136556619/samplepages/9780136556619_Sample.pdf
29. Russell, S. i Norvig, P. (1995) *Artificial Intelligence: A Modern Approach*. First edition, Prentice-Hall Inc., New Jersey. Dostupno, 31.8.2023 na: https://theswissbay.ch/pdf/Gentoomen%20Library/Artificial%20Intelligence/General/Artificial%20Intelligence%20A%20Modern%20Approach%20-%20Stuart%20J.%20Russell%20%2C%20Peter%20Norvig.pdf

30. Russell, S. and Norvig, P. (2010) *Artificial Intelligence: A Modern Approach*. Third edition. Pearson Education, Inc., New Jersey. Dostupno, 31.8.2023 na: https://web.cs.ucla.edu/~srinath/static/pdfs/AIMA.pdf
31. Schank, R., Partridge, D., Wilks, Y. (1993) *The foundations of artificial intelligence*, Cambridge University Press. Dostupno, 31.8.2023 na: https://www.cambridge.org/core/books/foundations-of-artificial-intelligence/4EA1C645196F48076AF3A07F56CB0331
32. Searle, J. (2001) Umovi, mozgovi i programi. U: Mišević, N. i Smokrovic, N. *Računala, mozak i ljudski um*. Izdavački centar Rijeka, Rijeka, str. 134–15. Dostupno, 31.8.2023 na: https://hrcak.srce.hr/file/320733
33. Šajer, V. (2019) *Samoprocjena medijske kompetencije studenata nastavničkih studija Filozofskog fakulteta u Rijeci*. Diplomski rad. Sveučilište u Rijeci. Dostupno, 31.8.2023 na: https://urn.nsk.hr/urn:nbn:hr:186:538677
34. Tolosana, R., Rathgeb, C., Vera-Rodriguez, R. & Busch, C. (2022) *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*. Springer Cham, Switzerland. Dostupno, 31.8.2023 na: https://doi.org/10.1007/978-3-030-87664-7
35. Vaccari, C. & Chadwick, A. (2020) *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*. Social Media + Society, 6. Dostupno, 31.8.2023 na: https://journals.sagepub.com/doi/full/10.1177/2056305120903408
36. Xu, Y., Raja, K., Ramachandra, R. & Busch, C. (2022) "Chapter 7 (139–163): Adversarial Attacks on Face Recognition Systems." In *Handbook of Digital Face Manipulation and Detection. From DeepFakes to Morphing Attacks*. Springer Cham. Dostupno, 31.8.2023 sa: https://doi.org/10.1007/978-3-030-87664-7

5. PRILOZI

Popis tablica

Popis slika