Tehnološki aspekti informacijskog poremećaja

Marković, Mijo

Sveučilište Josipa Jurja Storssmayera

Filozofski fakultet Osijek

Preddiplomski studij informatologije

Mijo Marković

Tehnološki aspekti informacijskog poremećaja

Završni rad

Mentor: doc. dr. sc. Snježana Stanarević Katavić

Osijek, 2022.

Sveučilište Josipa Jurja Strossmayera

Filozofski fakultet Osijek

Odsjek za informacijske znanosti

Preddiplomski studij informatologije

Mijo Marković

Tehnološki aspekti informacijskog poremećaja

Završni rad

Društvene znanosti, informacijske i komunikacijske znanosti, informacijski sustav i informatologija

Mentor: doc. dr. sc. Snježana Stanarević Katavić

Osijek, 2022.

Prilog: Izjava o akademskoj čestitosti i o suglasnosti za javno objavljivanje

Obveza je studenta da donju Izjavu vlastoručno potpiše i umetne kao treću stranicu završnog odnosno diplomskog rada.

IZJAVA

Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napravio te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citati s napisanim izvorom odakle su preneseni.
Svojim vlastoručnim potpisom potvrđujem da sam suglasan da Filozofski fakultet Osijek trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta Osijek, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

U Osijeku, datum 9.9.2022.

Mijo Marković, 0122231998
ime i prezime studenta, JMBAG

Sadržaj

Sažetak

Informacijski je poremećaj fenomen koji se pojavljuje na tri osnovna načina, i to u obliku pogrešnih informacija, dezinformacija i zlonamjernih informacija. To je fenomen koji je sredinom prošlog desetljeća intenzivno privukao pozornost znanstvenika i javnosti uslijed globalnih političkih događaja na koje je imao značajan utjecaj. Informacijski se poremećaj najlakše i najbrže širi putem medija i društvenih mreža te su znanstvenici ubrzo počeli upućivati na važnost istinitosti, relevantnosti i odgovornog ponašanja kada je u pitanju dijeljenje informacija. Osim medija i društvenih mreža, širenju informacijskog poremećaja, pridonosi i umjetna inteligencija, odnosno automatizacijske tehnologije za obradu informacija. Automatizacijske tehnologije koriste se za kreiranje i širenje dezinformacija uz minimalni ljudski napor. Najraširenij oblici automatizacijskih tehnologija koji se koriste za širenje dezinformacija na internetu su botovi i duboke krivotvorine. Nadalje, kako korisnici ostavljaju golemu količinu podataka svojom aktivnošću online, razni poslovni ili politički akteri iskorištavaju njihov digitalni otisak u svrhe ciljanog profiliranja, segmentiranja i mikrociljanja kako bi im mogli servirati sadržaje i oglase s ciljem ostvarivanja svojih komercijalnih ili političkih interesa. Uporaba automatizacijskih tehnologija podrazumijeva uporabu algoritama, računalnih funkcija koji imaju sposobnost samostalnog prilagođavanja, a ključna su sastavnica funkcioniranja društvenih mreža. Problem se javlja kada algoritmi pridonose stvaranju filter mjehurića i algoritamskih pristranosti te na koji način utječu na jačanje stavova i polarizaciju korisnika ili širenje predrasuda. Dok su istraživanja pokazala kako filter mjehurići i eho komore nisu onoliki problem za kakav su se u teoriji smatrali, algoritamske pristranosti u tražilicama ili u poslovnom odlučivanju mogu dovesti do ozbiljnih posljedica u vidu širenja predrasuda ili donošenja odluka utemeljenih na pristranim izvorima informacija.

Ključne riječi: informacijski poremećaj, automatizacijske tehnologije, botovi, duboke krivotvorine, mikrociljanje, filter mjehurić, algoritamske pristranosti

1. Uvod

U radu će se tumačiti na koji način automatizacijske tehnologije doprinose jačanju informacijskog poremećaja. Informacijski poremećaj je fenomen koji se definira kao kreiranje i dijeljenje lažnih i/ili netočnih informacija sa ili bez namjere nanošenja. U prvom poglavlju rada pojasnit će se što je informacijski poremećaj, u kakvim se oblicima pojavljuje, kada se pojam prvi puta pojavio te tko su kreatori informacijskog poremećaja. Glavna pitanja koja će se obraditi jesu što se sve podrazumijeva pod informacijskim poremećajem, kada i zbog čega javlja interes znanstvenika za proučavanje informacijskog poremećaja te koje posljedice mogu biti prouzročene širenjem informacijskog poremećaja. U sljedećem su poglavlju opisani automatizirani oblici tehnologije koji mogu utjecati na širenje informacijskog poremećaja. Kao primjere zloupotrebe automatizacijskih tehnologija navest će se botove i duboke krivotvorine tzv. „deep fakes“, te narušavanje privatnosti korisnika kroz prikupljanje podataka koje korisnici ostavljaju na internetu čime se omogućuje manipuliranje sadržaja koji se korisnicima prikazuje. Sljedeće poglavlje govori o legitimnoj uporabi automatizirane tehnologije što također može pojačati utjecaj lažnih i netočnih informacija kroz fenomene filter mjehurića i eho komore. U istom se poglavlju tumači koncept algoritamskih pristranosti i posljedice do kojih one mogu dovesti. U posljednjem je poglavlju opisano na koje se načine kontrolira i moderira uznemirujući i obmanjujući sadržaj na internetu te koja je uloga automatizacijskih tehnologija u rješavanju navedenog problema. U fokusu ovoga rada jest tumačenje uloge tehnologije u informacijskom poremećaju, od toga na koji način tehnološka rješenja koja su kreirana za olakšavanje obrade informacija mogu doprinositi jačanju informacijskog poremećaja, do toga kako se ta ista rješenja koriste u njegovu suzbijanju.

2. Što je informacijski poremećaj

Informacijski je poremećaj čije je korištenje Vijeće Europe preporučilo umjesto izraza "lažne vijesti" 2017. Godine, a može se pojaviti u tri oblika:

- pogrešne informacije ili misinformacije – netočne informacije koje korisnik širi ne znajući da su netočne
- dezinformacije – netočne informacije koje su namjerno stvorene kako bi načinile štetu
- zlonamjerne informacije ili malinformacije – istinite informacije koje su najčešće privatno vlasništvo koje se dijele sa svrhom nanošenja štete korisniku[1]

Pojam je prvi puta upotrjebljen 2017. godine u izvješću Vijeća Europe o problematici lažnih vijesti, a od tada se preporučuje da se pojam „informacijski poremećaj" koristi umjesto pojma „lažne vijesti".[2] Informacijski je poremećaj pridobio veću pozornost znanstvenika nakon 2016. godine i dvaju političkih događaja koja su obilježena intenzivnim širenjem lažnih i netočnih informacija, a to su američki izbori na kojima je pobijedio Donald Trump i referendum o izlasku Velike Britanije iz Europske unije, tzv. Brexit.[3]

Iako se o informacijskom poremećaju intenzivnije promišlja i govori od sredine prošlog desetljeća, njegovi pojavni oblici datiraju još iz doba antike, a od izuma tiskarskog stroja lažne i netočne informacije šire se i službenim kanalima i glasilima jer u vrijeme pojavih prvih novina u današnjem smislu riječi početkom 17. stoljeća nije postojao koncept novinarske etike. U povijesti novinarstva poznat je primjer izvještavanja o katastrofi koja je zadesila grad Lisabon 1755. godine, na dan Svi svetih. Grad je u isto vrijeme pogodio potres, tsunami i požar te je navedena katastrofa bila središnja tema novina veće tiraže tijekom 1756. godine. Objavljivanje su priče o iskustvima građana Lisabona tijekom potresa koje se nikada nisu dogodile. Ti su izvještaji sadržavali elemente fantazije, drame i različita naklapanja, a bili su glavni izvor informacija za

[1] Usp. Kandel, Nirmal. Information Disorder Syndrome and Its Managment. // Journal of Nepal Medical Asociation 58, 224 (2020), str. 281. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7580464/pdf/JNMA-58-224-280.pdf

[2] Usp. Wardle, Claire; Derakhshan, Hossein. Information disorder: Toward an interdisciplinary framework for research and policy making, 2017. URL: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

[3] Usp. Frau-Meigs, Divina. Information Disorders: Risks and Opportunities for Digital Media and Information Literacy. // Medijske studije 10, 19 (2019), str. 16-18. URL: https://hrcak.srce.hr/file/330016

nepismene. Objavljivanje navedenih priča pridonijelo je komercijalnom uspjehu mnogih tiskovina u to vrijeme, uključujući vjerske tekstove koji su pripisali potres Božjem gnjevu kao božansku osvetu grešnicima.[4]

Premda su se lažne i netočne informacije i propaganda koristili još i prije pojave tiska, moderno informacijsko okruženje koje se bazira na digitalnim tehnologijama omogućilo je masovno širenje informacija, ali isto tako i dezinformacija.

3. Utjecaj digitalnog informacijskog okruženja na informacijski poremećaj

Kako dolazi do gotovo svakodnevnog razvoja tehnologija, danas je korisnicima medija moguće pružiti nevjerojatnu količinu informacija na dnevnoj razini. Izravan pristup, neposrednost i interakcija sa sadržajem neupitne su prednosti suvremenog informacijskog okruženja. Isprva je tradicionalnim medijima kao što su novine i televizija bila moguća kontrola diseminacije informacija i informiranja građana, ali s pojavom interneta i društvenih mreža, tradicionalni „gatekeeperi“ izgubili su mogućost kontrole sadržaja koji se objavljuju u javnom prostoru. Nadalje, tradicionalni su mediji uslijed prelaska na digitalne platforme promijenili svoj model poslovanja prešavši s pretplatnog na reklamni model. Kako se reklamni model poslovanja bazira na financiranju mrežnih mjesta temeljem broja posjeta koje primaju od korisnika, mediji su postupno počeli razvijati fleksibilan odnos s istinom, pribjegavati senzacionalizmu i “clickbait” stilu kako bi privukli korisničke klikove.[5] No mediji nisu jedini koji su u digitalnom okruženju promijenili svoj odnos prema istini i činjenicama potaknuti ekonomskim interesima. U privatnom komercijalnom sektoru također se koriste mogućnosti koje pruža digitalno okruženje za manipulaciju informacijama u svrhu unapređenja prodaje i izgradnje pozitivne reputacije. U tom se kontekstu najčešće kao problem navode lažne korisničke ocjene o proizvodima i tvrtkama. U istraživanju u kojem je sudjelovalo nešto više od 1000 korisnika, koje je provela američka istraživačka kompanija Dimensional Research 2013. godine, ustanovljeno je da na 90% kupaca pozitivna ocjena nekog proizvoda ima veliki utjecaj prilikom donošenja odluke hoće li određeni proizvod kupiti ili ne, dok je ustanovljeno da je na 86% kupaca negativna ocjena utjecala hoće li

[4] Usp. Araújo, Ana Cristina The Lisbon Earthquake of 1755, 2006.
URL: https://www.brown.edu/Departments/Portuguese_Brazilian_Studies/ejph/html/issue7/html/aaraujo_main.html
[5] Usp. Dezinformacije i propaganda na internetu, CERT.hr-PUBDOC-2019-3-375, 2019.
URL: https://www.cert.hr/wp-content/uploads/2019/03/dezinformacije_i_propaganda_na_Internetu.pdf

kupiti neki proizvod.[6] Načini kako se ocjene mogu lažirati kreću se od prisiljavanja vlastitih zaposlenika na ostavljanje pozitivnih komentara do plaćanja specijaliziranih tvrtki koje posebnim tehnologijama taj posao čine još učinkovitije.[7]

Ipak, najveću raspoloživu moć za širenje dezinformacija i utjecaj na javno mišljenje imaju države i institucije koje su s njima povezane. Takav problem je izraženiji su državama s manje razvijenim demokratskim ustrojem kao što su Kina i Saudijska Arabija. Primjerice, dokazano je da u Kini postoji svojstvena „vojska trolova“, koja će u slučaju da se na društvenim mrežama ili forumima pojavi negativni komentar o nečemu što nije dopušteno kritizirati, kao npr. vlada, u malom vremenskom roku reagirati velikim brojem suprotstavljenih komentara koji dolaze s lažnih profila.[8] Istraživanja Sveučilišta u Oxfordu iz 2019. godine otkrila su kampanje za manipuliranje javnim mnijenjem na mrežnim platformama u 70 zemalja svijeta.[9]

Internet omogućuje širenje netočnih i lažnih informacija informacijskim ponašanjem samih korisnika medija i društvenih medija, s jedne strane. Pojavom naprednijih tehnologija postalo je moguće dezinformacije još više učiniti uvjerljivima pomoću tehnologija za manipuliranje audio-vizualnih sadržaja kojima se stvara potpuno izmišljeni sadržaj, a korisnicima dodatno otežava razlikovati istinitu od lažne informacije. Međutim, osim opisanih problema koji su kombinacija digitalnog okruženja i ljudskog djelovanja u istom, napredna tehnološka rješenja u vidu automatizacijskih tehnologija, odnosno umjetne inteligencije koja je oblikovana da bi se olakšala i unaprijedila obrada informacija, mogu doprinijeti širenju informacijskog poremećaja. U sljedećem poglavlju opisati će se primjeri zlouporabe automatizacijskih tehnologija u upravljanju informacijama.

---

[6] Usp. Gesenhues, Amy. Survey: 90% Of Customers Say Buying Decisions Are Influenced By Online Reviews, 2013. URL: https://martech.org/survey-customers-more-frustrated-by-how-long-it-takes-to-resolve-a-customer-service-issue-than-the-resolution/

[7] Usp. Dezinformacije i propaganda na internetu. Nav. dj. Str. 7-8.

[8] Usp. Isto. Str. 6.

[9] Usp. Nenadić, Iva; Vučković, Milica. Dezinformacije: Edukativna brošura i vježbe za razumijevanje problema dezinformacije. Str. 9-10. URL: https://www.medijskapismenost.hr/wp-content/uploads/2021/04/brosura-Dezinformacije.pdf

3.1. Zlouporaba automatizacijskih tehnologija u upravljanju informacijama: botovi i duboke krivotvorine

Jedan od najmoćnijih alata koji kreatori dezinformacija imaju u svom arsenalu su automatizirane tehnologije koje obavljaju procese širenja dezinformacija umjesto njih. To je moguće ostvariti na nekoliko načina, a najrašireniji način jest korištenje botova. Botovi su softverski programi koji obavljaju automatizirane zadatke s ciljem da imitiraju ljudske postupke. Botovi se ne koriste isključivo u negativne svrhe. Oni su zapravo jako rašireni na Internetu te se koriste u radu službi za korisnike, u različitim aplikacijama u kojima simuliraju osobnu komunikaciju i sl. Ali kao i s ostalim tehnologijama koje pružaju mogućnost iskorištavanja, nije trebalo dugo da se prepozna potencijal za iskorištavanje botova u zlonamjerne aktivnosti.[10] U kontekstu informacijskog poremećaja botovi se najčešće koriste za kreiranje lažnih korisničkih profila koji se zatim koriste za podržavanje ili diskreditiranje određenog mišljenja ili teme objavljivanjem poruka na društvenim mrežama čime se iskorištava ljudska sklonost naginjanju mišljenju većine što također može imati utjecaja i na osobe koje se ne slažu s temom rasprave.[11] Američki stručnjaci za sigurnost smatraju kako je 10% sadržaja na društvenim mrežama i oko 62% sadržaja općenito na Internetu stvoreno od strane botova.[12] U literature se navode tri vrste botova:

- propagandni botovi – služe za širenje misinformacija o nekoj temi, koriste veliki broj poruka kako bi što više utjecali na mišljenje korisnika, moguće ih je koristiti za pozitivan ili negativan utjecaj, a predstavljeni su kao profili stvarnih osoba
- botovi sljedbenici – služe za simuliranje velikog broja korisnika koje se koriste za pridodavanje lažnog kredibiliteta nekoj ideji, a često se koriste u političke svrhe
- botovi barikade – služe za ometanje i onemogućavanje civilizirane komunikacije provociranjem i masivnog „spamanja“ poruka kako bi se što teže stvarne poruke razlikovale od lažnih.[13]

Dokazano je da su se botovi koristili tijekom političkih kampanja i izbora tijekom Brexita, njemačkih izbora 2017. godine, referenduma o samostalnosti Katalonije 2017. i nekolicini drugih

[10] Usp. What are bots? – Definiton and Explanation.
URL: https://www.kaspersky.com/resource-center/definitions/what-are-bots
[11] Usp. Dezinformacije i propaganda na internetu. Nav. dj. Str. 4.
[12] Usp. Powers, Shawn; Kounalakis, Markos. Can Public Diplomacy Survive the Internet?: Bots, Echo Chambers and Disinformation, 2017. Str. 18. URL: https://www.state.gov/wp-content/uploads/2019/05/2017-ACPD-Internet.pdf
[13] Usp. Dezinformacije i propaganda na internetu. Nav. dj. Str. 9-10.

primjera.[14] Kroz provedena istraživanja, ustanovljeno je da se za 8,5% na Twitteru i 7% profila na Facebooku smatra da su bot profili.[15] Za primjer, način na koji Twitter botovi funkcioniraju je taj da kao input imaju definiran jedan ili više izvora iz kojeg sakupljaju novosti koje zatim „tweetaju“, a to mogu biti druge mrežne stranice, baze podataka, arhivi, neki drugi Twitter profil i sl. Područja za koja se botovi na Twitteru najviše koriste su politika, financije, crna kronika i zabava, prognoza vremena i stanje u prometu.[16] Problem s botovima na Twitteru je dosegao toliku razinu da je Twitter odnedavno počeo ispod imena profila stavljati oznaku „automatski“ kako bi se lakše razlikovali objave stvarnih korisnika od objava botova (Slika 1).[17]

Slika 1. tweet automatiziranog Twitter računa[18]

Kreatori dezinformacija moraju biti kreativni u svome nastojanju da zavaraju velik broj ljudi tako da moraju koristiti različite alate, a osim botova, posljednjih su se godina počele koristiti duboke krivotvorine, tzv. „deepfakes“. Duboke krivotvorine su digitalni, najčešće audio-vizualni sadržaj, manipuliran i uređen da se što manje razlikuje od stvarnog sadržaja. Najčešće se koriste kao oblici zabave, ali isto i u svrhe prevare korisnika, u političkim prijeporima ili za osvetu. Već postoje i besplatni i komercijalni softveri koji služe za kreiranje krivotvorina te se smatra da će do 2030. godine *deepfake* softveri postati toliko napredni da neće postojati način

[14] Usp. Dumbrava, Costica. Key social media risks to democracy: Risks from surveillance, personalization, disinformation, moderation and microtargeting. Brussels: European Parliamentary Research Service, 2021. URL: https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/698845/EPRS_IDA(2021)698845_EN.pdf
[15] Usp. Lokot, Tetyana; Diakopoulos, Nicholas. News Bots: Automating news and information dissemination on Twitter. // Digital Journalism 4, 6 (2016), str. 2.
URL: http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/newsbots_final.pdf
[16] Usp. Isto. Str. 6-7.
[17] Usp. About automated account labels. URL: https://help.twitter.com/en/using-twitter/automated-account-labels
[18] Peters, Jay. Twitter is now letting developers add labels to bot accounts, 2022. URL: https://www.theverge.com/2022/2/16/22937435/twitter-labels-automated-bot-accounts

kako razlikovati krivotvoreni sadržaj od stvarnog, a to se ne odnosi samo na ljude, nego i na strojeve. Realistični prikazi ljudi koji rade ili govore nešto što nikada nisu napravili mogu imati velike posljedice na kredibilitet i povjerljivost osoba i ustanova ili pak utjecati na ishod političkog dvoboja, za što se ovakvi alati i najčešće koriste, ako se ne uoči da se radi o krivotvorini.[19]

Stručnjaci su iznijeli četiri načina kako se suprostaviti dubokim krivotvorinama:

1. korištenje forenzičkih alata kako bi se prepoznali neuobičajeni uzorci na umjetno kreiranim licima
2. autentifikacija digitalnog sadržaja svojstvenim žigom koji bi se stavio u trenutku nastanka sadržaja
3. potpuni nadzor osobe i njenog kretanja i ponašanja, ali ta ideja neće nikada oživjeti s obzirom da krši jedno od osnovnih ljudskih prava, pravo na privatnost
4. korištenje istih alata koji kreiraju krivotvorine za njihovo prepoznavanje i njihovo implementiranje u algoritme pretraživača i društvenih medija.[20]

Korištenje navedenih alata popularno je iz dvaju razloga: svatko ih može koristiti zbog njihove jednostavnosti bez obzira na razinu snalaženja s tehnologijom jer cijeli proces obavlja stroj pomoću algoritama koji se temelje na umjetnoj inteligenciji. Duboke krivotvorine funkcioniraju tako što koriste veliki broj slika i videa s interneta osobe koju se želi krivotvoriti koje se potom kombiniraju kako bi se dobio što kvalitetniji prikaz neke osobe, što znači da što je osoba popularnija, to će njezina krivotvorina biti uvjerljivija.[21] Samo postojanje mogućnosti i jednostavnosti kojom je moguće kreirati krivotvorinu bilo koje osobe predstavlja opasnost osobnoj privatnosti, a činjenica da se ni na koji način korištenje krivotvorina ne može zaustaviti nije nešto što bi se trebalo olako shvatiti. Koliko god da su algoritmi za kreiranje krivotvorina učinkoviti, još uvijek nisu dosegli svoj vrhunac tako da još uvijek postoje metode za njihovo prepoznavanje. Naime, ti algoritmi imaju nekoliko nedostataka, a najuočljiviji je kada je potrebno

[19] Usp. Kertysova, Katarina. Artificial Intelligence and Disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered // Security and Human Rights 29, 1-4 (2018), str. 66-67. URL: https://brill.com/downloadpdf/journals/shrs/29/1-4/article-p55_55.xml
[20] Usp. Isto. Str. 15.
[21] Usp. Albahar, Marwan; Almalki, Jameel. Deepfakes: threats and countermeasures: systematic review // Journal of Theoretical and Applied Information Technology 97, 22 (2019), str. 3243. URL: http://www.jatit.org/volumes/Vol97No22/7Vol97No22.pdf

krivotvoriti treptanje očiju koje još uvijek nije moguće krivotvoriti da izgleda ljudski. Također je moguće tražiti detalje u boji očiju i zubima.[22]

Nadalje, automatizacijske tehnologije osim što olakšavaju kreiranje i dijeljenje lažnih i netočnih informacija, također omogućuju personalizirano ciljanje svakog pojedinog korisnika na internetu čime se korisnicima prezentira onaj sadržaj koji algoritmi odaberu, što može pojačati njihova vjerovanja i stajališta te polarizaciju korisnika medija bez da su korisnici toga svjesni. Poseban je problem kada se korisnike cilja dezinformacijama. O navedenim problemima govori sljedeće poglavlje.

3.2. Narušavanje privatnosti korisnika i manipulacija

Korisnici Interneta danas nemaju previše filtera prilikom objavljivanja na društvenim mrežama. Objavljuje se sve, od toga što osoba radi u određenom trenutku, kako se osjeća, gdje se nalazi i sl. Sav taj materijal, prigodno nazvan digitalni otisak, čuva se na privatnim serverima društvenih mreža, a može se koristiti za različite svrhe narušavanja privatnosti i manipuliranja korisnika. Digitalni otisak ne obuhvaća samo sadržaj koji se objavljuje, nego podrazumijeva i lajkove, komentar i dijeljenja, a isto i sadržaj koji netko drugi objavi o određenoj osobi. Digitalni otisak može se ostavljati pasivno i aktivno. Pasivni digitalni otisak podrazumijeva prikupljanje podataka u tajnosti, kako što je otkrivanje korisnikove lokacije putem IP adrese, dok se aktivni digitalni otisak odnosi na postupke kada korisnik samostalno ostavlja osobne podatke.[23]

Kršenje privatnosti korisnika obuhvaćaju narušavanje privatnosti kroz prikupljanje podataka o korisniku bez njegovog pristanka profiliranje i segmentaciju korisnika te

---

[22] Isto. Str. 3247-3248.

[23] Usp. Arakerimath, Anjana R; Gupta, Pramod Kumar. Digital Footprint: Pros, Cons and Future // International Journal of Latest Technology in Engineering, Managment & Applied Science 4, 10 (2015), str. 52-53. URL: https://d1wqtxts1xzle7.cloudfront.net/39569967/52-56-with-cover-page-v2.pdf?Expires=1659999043&Signature=gzkBPmSddBC~DWGKe8823Zotqa3NhQG26vAYnOBcQZbNNv38-XPCdtklVDm7B2cPveG-QdL1FSgELX0b6kBznY4mpTKsYmFLxDlHBKsrZJHebCQXgZLZSkLnLgFU9ZvjNZX~6SKkjbIdqvHHLxOTzQN41ZbLhlW3P4k3Xot4rs-IXX0ItcNotBfMPoYzOOu-Zh4IlpXmALd3pqTp17yYAO5rL5rHBNQyY8KKjTGfQlv82MfXOvCvfcK~xAPOqTwB6Dj4K3Q18EAQJIQoJgFxX1tflZnIulUgrjtsh7ZAfD3fGzofWqk81wEecpkW4RU3Te-~xUNmKINFO5JVR28CzA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

hiperpersonalizirano ciljanje i mikrociljanje. S obzirom na sličnost ovih fenomena, oni će biti grupirani zajedno te će se tako i opisati.

Profiliranje korisnika odvija se u tri koraka, prvi od kojih je ekstrakcija korisnih podataka korisnika s interneta i njegovih profila na društvenim mrežama, a omogućeno je toliko detaljno praćenje da je moguće bilježiti kako korisnik koristi miš tj. kako pozicionira miš na tekst, na što klikće te na što ne klikće, kojom brzinom lista tekst te kako si prilagođava veličinu prozora web preglednika. Poslije toga slijedi integracija profila u kojem se brišu svi duplicirani podaci. Nakon uređivanja podataka, korisnici se dijele u zajedničke grupe interesa, tj. korisnici se filtriraju. Samo filtriranje provodi se na dva načina, prvi od kojih je filtriranje prema sadržaju koje funkcionira na sistemu ponude određenih proizvoda korisnicima koji su istim proizvodima dali slične ocijene, dok drugi način grupira korisnike u skupine prema tome koje proizvode koriste. Ako su se korisnici slagali oko više proizvoda prije, šanse su da će se oko njih slagati i u budućnosti, stoga ova tehnika filtriranja jako ovisi o kvaliteti stvaranja korisničkih grupa.[24] Mikrociljanje ili hiperpersonalizirano ciljanje je proces kreiranja prilagođenog sadržaja i pružanje tog sadržaja u skladu s interesima korisnika. Mikrociljanje na društvenim medijima omogućuje da se na osnovi različitij čimbenia, kao što su dob, spol, hobiji i sl., ponude reklame koje odgovaraju njegovim interesima. Oglašivači na taj način koriste manje sredstava za efektivnije reklamiranje, a u isto vrijeme prikupljaju podatke koji služe za izvještaje o učinkovitosti reklama, obraćaju li korisnici na njih pažnju i koliko često na njih klikću.[25]

Učinkovito mikrociljanje zahtjeva nekoliko koraka. Prvi korak je definiranje želja i potreba određene grupe korisnika, a to se određuje testiranjem i proučavanjem tržišta. Sljedeći korak podrazumijeva kreiranje proizvoda koji najbolje pokriva otkrivene želje i potrebe ciljanih korisnika i zadnji korak je kvalitetna prezentacija proizvoda na tržištu kako bi se što više povećale šanse za odabir od strane korisnika. U slučaju Facebooka, koji od korisnika traži ime, prezime, lokaciju, datum rođenja i spol, sve navedeno već može poslužiti za mikrociljanje bez da je osoba ostavila lajk ili komentar ili objavila neki sadržaj.

---

[24] Usp. Kanoje, Sumitkumar; Girase, Sheetal; Mukhopadhyay, Debajyoti. User Profiling Trends, Techniques and Applications // International Journal of Advance Foundation and Research in Computer 1, 1 (2014), str. 2-4. URL: https://arxiv.org/ftp/arxiv/papers/1503/1503.07474.pdf

[25] Usp. Barbu, Oana. Advertising, Microtargeting and Social Media, 2013. str. 44-46. URL: https://www.sciencedirect.com/science/article/pii/S187704281406385X

Mikrociljanje može biti jako efektivan marketinški alat, ali se postavljaju pitanja o etičnosti korištenja. Iako se tako ne čini prikupljanje korisničkih podataka, profiliranje i segmentacija korisnika i mikrociljanje legitimne su radnje, sve dok se ne koriste za potrebe manipuliranja. S jedne strane, korisnicima je prezentiran isključivo sadržaj koji ih zanima, ali opet samim time što vide samo jedan dio sadržaja, znači da ne vide cijelu sliku te se ne ostavlja dovoljno mjesta za kritičko promišljanje.[26]

Jedan od najpoznatijih primjera ilegalnog prikupljanja podataka o korisnicima u svrhu mikrociljanja afera je koja je pogodila Facebook 2016. godine. gdje je utvrđeno da su tijekom godina razne tvrtke, Cambridge Analytica kao najpoznatija, koristile Facebookove podatke od 87 milijuna korisnika za ciljano političko oglašavanje s ciljem utjecanja na političke preferencije koje je svoj vrhunac doseglo tijekom Brexita i američkih predsjedničkih izbora 2016. godine. U slučaju Cambridge Analytice, oni su podatke prikupljali preko kviza osobnosti, koji nakon povezivanja s Facebook profilom, omogućuje kreatoru aplikacije pristup profilu osobe koja je riješila kviz, a isto tako popisu prijatelja i lajkovima, a s time su počeli 2014. godine. Godinu dana kasnije, u Facebooku su prvi puta doznali da se podaci koji su u njihovu vlasništvu dijele, te su smjesta uklonili aplikaciju te zahtijevali da Cambridge Analytica formalno da izjavu da su obrisani svi prikupljeni podaci. Tijekom narednih godina, uslijedio je niz optužbi od raznih strana protiv Facebooka što je uvjetovalo da Mark Zuckerberg svjedoči pred američkim Kongresom 2018. godine. Cijela ova afera još uvijek nije riješena. Glede Brexita, na Facebooku su, u sklopu kampanja kao što su VoteLeave i BeLeave, objavljivani razni oglasi koji su se ciljano prikazivali odabranim korisnicima kako bi ih naveli na glasanje za izlazak iz Europske unije za što je utrošeno oko 2.7 milijuna funti (Slika 2.). Oglasi su se najčešće fokusirali na specifične probleme do kojih bi došlo ako bi Velika Britanija ostala u Europskoj uniji kao što je imigracija koja bi zatim dovela do nedovoljno radnih mjesta, prekomjerno slanje britanskog novca u EU fondove i sl.[27]

[26] Isto. Str. 47-48.
[27] Usp. Vote Leave's targeted Brexit ads released by Facebook, 2018. URL: https://www.bbc.com/news/uk-politics-44966969

Slika 2. Primjer oglasa na Facebooku tijekom Brexita[28]

Opisani primjeri prikupljanja podataka o korisnicima u svrhu ciljanog političkog oglašavanja ilustriraju domet i razmjere u kojima je moguće provoditi dezinformacijske kampanje u suvremenom tehnološkom okruženju.

4. Legitimna uporaba automatizacijskih tehnologija

Automatizacijske tehnologije baziraju se na algoritmima, odnosno računalnim funkcijama koje se implementiraju u sustav računala, a njihova je specijalnost mogućnost samostalnog prilagođavanja u slučaju potrebe te su ključna sastavnica funkcioniranja i poslovnog modela društvenih mreža, tražilica i ostalih komercijalnih sustava. Algoritmi koriste prethodna pretraživanja i aktivnost korisnika kako bi im se ponudio sadržaj koji ih zanima. Nakon Brexita i europske migrantske krize, javilo se puno zabrinutih glasova koji su isticali negativne strane algoritama na društvenim mrežama. U tom se kontekstu problematiziraju dva srodna fenomena, a to su filter mjehurić i eho komora, te algoritamske pristranosti.[29]

[28] Isto
[29] Usp. Dokler, Ana. Ako koristite Google, Facebook ili Twitter, morate znati kako djeluju algoritmi, 12. 11. 2019. URL: https://www.medijskapismenost.hr/ako-koristite-google-facebook-twitter-trebate-znati-kako-djeluju-algoritmi/

4.1. Filter mjehurići i eho komore

Oba fenomena opisuju slične koncepte, ali dok filter mjehurići predstavljaju fenomen gdje algoritmi selektiraju koji se sadržaj prikazuje korisniku, tj. prikazuju onaj sadržaj za koji se smatra da je korisniku zanimljiv, eho komore, koje imaju sličan efekt, opisuju isključivo grupu istomišljenika koja se "hrani" informacijama koja potvrđuje njihova vjerovanja. Prije nego što protumačimo zašto ova dva fenomena nisu tolika prijetnja kolikom se smatraju, bitno je pojasniti pojam kognitivne pristranosti. Kognitivna je pristranost sistematička greška u razmišljanju koja se javlja kada ljudi analiziraju neku informaciju koja utječe na odluke i razmišljanja koja donose, a često su rezultat pokušaja pojednostavljivanja mentalne obrade informacija. Neki od znakova da je osoba pod utjecajem kognitivne pristranosti uključuju: praćenje samo onih vijesti koja potvrđuju vlastita mišljenja, pripisivanje tuđeg uspjeha pukoj sreći, pretpostavljanje da sve osobe imaju isto mišljenje koje je identično vlastitom i sl. Postoji prostor za daljnju raspravu o kognitivnoj pristranosti, ali kako nije direktna tema ovog rada, zadržavamo se na ovom kratkom opisu.[30] Filter mjehurići predstavljaju stanje informacijske izolacije u kojoj algoritam nekoj osobi prikazuje primarno informacije kojima je ta osoba sklona, a temelji se na prethodnim pretraživanjima i ponašanju online. Tijekom godina na filter mjehuriće gledao se kao na veliki problem za demokratske procese. Provedeno međunarodno istraživanje o učincima filter mjehurića iz 2020. godine pokazalo je kako se od društvenih medija, za novosti najviše koristi Facebook, da ispitanici koji društvene mreže ne koriste isključivo za vijesti, bivaju nenamjerno izloženi vijestima na društvenim mrežama, što opet povećava broj vijesti kojima se koriste za razliku od ispitanika koji ne koriste društvene mreže. Rezultati su pokazali da je korištenje različitih izvora vijesti bilo prisutnije kod mlađih ispitanika i kod ispitanika koji se nisu previše zanimali za vijesti te na aplikacijama YouTube i Twitter. Na kraju istraživanja iznesen je zaključak da ne postoje dokazi da algoritmi, odnosno teorijski koncept filter mjehurića dovodi do intelektualne izolacije, no da može doprinijeti dodatnoj polarizaciji neistomišljenika.[31]

Kada je riječ o digitalnim eho komorama, one se definiraju kao okruženja u kojima je korisnik okružen svojim istomišljenicima koji onda zajedno jedni drugima pojačavaju učinak

[30] Usp. Cherry, Kendra. What Is Cognitive Bias?, 19. 7. 2020. URL: https://www.verywellmind.com/what-is-a-cognitive-bias-2794963
[31] Usp. Fletcher, Richard. The truth behind filter bubbles: Bursting some myths. // Reuters Institute, 24. 1. 2020. URL: https://reutersinstitute.politics.ox.ac.uk/news/truth-behind-filter-bubbles-bursting-some-myths

kognitivne pristranosti. Dolazi se do zaključka da je glavna razlika između filter mjehurića i eho komora ta da filter mjehurići zatvaraju jednu osobu u krug intelektualne izolacije dok se u eho komorama zatvara grupa ljudi sličnih interesa i stavova.

Slično kao u slučaju s filter mjehurićima, znanstvenici smatraju kako su navodi o opasnostima eho komora pretjerani.[32] Nasuprot popularnom mišljenju, znanstvenici smatraju da korisnici već imaju razvijene navike kako zaobići zastajanje u eho komorama. Specifičan problem koji se javlja, u pogledu političke informiranosti i angažmana, je taj što korisnici ako to žele, mogu jednostavno zaobići novosti političke prirode, što pridodaje tome da su manje informirani o političkim događanjima i da su manje šanse da će izaći i glasati kada za to dođe vrijeme. Takvo nešto nije osobito pogodno za održavanje moderne demokracije.

Provedeno istraživanje o učincima filter mjehurića iz 2018. godine pokazalo je da su osobe koje su uključene u političke rasprave online sklonije provjeravanju dodatnih resursa kako bi utvrdili točnost prezentiranih činjenica te su također sklonije mijenjanju mišljenja ako se suoče s dovoljno uvjerljivim argumentom ili dokazom. Na društvenim mrežama će i dalje postojati mjesta gdje se izmjena informacija vrši tako što se potvrđuju međusobno stajališta, ali u današnje vrijeme, jako malen broj korisnika svoje informacije dobiva iz jednog izvora. Uz sve rečeno, ne bi škodilo kada bi velike društvene mreže, Twitter i Facebook pogotovo, promovirale medijsku pismenost s obzirom na veličinu korisničke publike.[33] Twitter je poduzeo korak tako što je na profilima koji su u službi vlade stavio oznaku da pripadaju vladi određene države (Slika 3) te je na članke čiji je izvor državno vlasništvo također stavio sličnu oznaku (Slika 4).[34] Isto tako, ako se pokaže da je neka netočna informacija postala popularna tema na Twitteru, na vrhu ”Pretražite” sekcije pojaviti će se članak koji objašnjava da ta informacija nije točna. Takav se primjer nedavno dogodio s fotografijom bivšeg košarkaša Earvina ”Magica” Johnsona gdje je prikazan kako daruje krv, a koji je također jedna od najpoznatijih osoba oboljelih od HIV-a (Slika 5).

[32] Usp. Rajan. Amol. Do digital echo chambers exist?, 4. 3. 2019. URL: https://www.bbc.com/news/entertainment-arts-47447633
[33] Usp. Dubois, Elizabeth; Blank, Grant. The myth of echo chambers, 9. 3. 2018. URL: https://www.oii.ox.ac.uk/news-events/news/the-myth-of-the-echo-chamber/
[34] Usp. About goverment and state-affiliated media account labels on Twitter. URL: https://help.twitter.com/en/rules-and-policies/state-affiliated

Slika 3. Primjer Twitter računa koji pripada vladi[35]

Slika 4. Primjer članka koji vodi na izvor koji je u posjedu države[36]

[35] Isto
[36] Isto

Slika 5. Primjer članka na Twitteru koji razotkriva istinu o lažnim informacijama[37]

Kako u ovom radu razmatramo utjecaj tehnologije na informacijski poremećaj, potrebno je pojasniti na koji način algoritmi, te teorijski koncepti filter mjehurić i eho komore, mogu doprinositi njegovu jačanju. U slučaju Amazona, istraživanja su pokazala da mnogi online trgovci koriste algoritamske vođene preporuke kako bi ih usmjerili prema proizvodima koji odgovaraju njihovim interesima. Taj je koncept sam po sebi bezopasan, ali mogu se javiti velike implikacije ako se u obzir uzme kontekst teorija zavjere i uznemiravajućeg sadržaja. Za većinu korisnika sekcije kako što su “Korisnici koji su vidjeli ovaj proizvod vidjeli su i sljedeće” najčešće su koristan način da pronađu određeni proizvod, ali za teoretičare zavjera i osobe s ekstremnim stajalištima, te sekcije mogu poslužiti za daljnju radikalizaciju njihovih stajališta. Ako osoba pogleda jednu knjigu o teorijama zavjere, njoj će se prikazati ne samo knjige o teoriji zavjere koju su zatražili, nego i ostale teorije zavjere koje mogu odvesti korisnika dublje u zavjernički sadržaj. Također, problem se javlja kod tražilica koje samostalno nadopunjavaju korisnikov upit, i za koje je vrijedno napomenuti, nisu problem samo kod Amazona nego i kod ostalih velikih kompanija kako što su Google i drugi. Automatsko nadopunjavanje može odvesti korisnike na krivu stranu te im umjesto korisnih proizvoda, može prikazati proizvode koji šire

[37] Snimak zaslona Twittera s korisničkog računa autora rada

ekstremistička stajališta i teorije zavjera bez da su korisnici toga svjesni. Slični slučaj se događa i kod autora, gdje na svakoj stranici zasebnog autora postoji "Korisnici su također su kupili porizvod od". Korisnici koji kupe knjigu od jednog ektremističkog autora skloni su kupovanju sličnih knjiga od nekog drugog autora, a algoritam se time samostalno trenira da prikazuje više takvih autora novim korisnicima.[38]

4.2. Algoritamske pristranosti

Već spomenuta automatizacija pomoću algoritama koji se koriste u poslovanju zahtjeva samostalnu adaptaciju algoritama kako bi bili što prikladniji za poslovanje te kako bi se u potpunosti mogle izbaciti ljudske pogreške. Tako se barem govori u teoriji, ali u praksi je slučaj nešto drugačiji. Algoritmi strojnog učenja oslanjaju se na pristupačnosti velike količine podataka koju koriste za treniranje i nadogradnju, a s obzirom na količinu podataka online, omogućen je velik broj takvih algoritama. Ako algoritam krene analizirati grupu podataka koja je nepotpuna ili ako sadrži greške, dolazi do toga da algoritam krivo uči te neće biti u stanju pravilno izvoditi ulogu za koju je namijenjen. Također postoji problem ako se algoritmi treniraju na skupu podataka koji sadrži određene pristranosti što, na primjer, može dovesti do segregacije određene društvene skupine.[39] Kao primjer može se uzeti Amazon koji je morao ukloniti algoritam koji je prilikom zapošljavanja bio pristran prema muškom spolu. Do tog je slučaja došlo tako što su Amazonovi algoritmi proučavali uzorke u životopisima kandidata kroz zadnjih 10 godina. S obzirom na dominaciju muškog spola u tehnološkoj industriji, većina tih kandidata bila je muškog spola, te je došlo do toga da su ženski kandidati dobivali manje bodova samo na temelju spola.[40]

Za primjere algoritamskih pristranosti također se može navesti razne slučajeve vezane za tražilicu Google. 2009. godine Safiya Noble ukazala je na činjenicu da tražilica Google na upit „black girls", najprije prikazuje hiperseksualizirane rezultate koji su uključivali taj pojam.[41]

[38] Usp. Thomas, Elise. Amazon's Algorithms, Conspiracy Theories and Extremist Literature, 4. 5. 2021. URL: https://www.isdglobal.org/digital_dispatches/amazons-algorithms-conspiracy-theories-and-extremist-literature/
[39] Usp. Listeš, Daniela. Algoritamska pristanost. URL: https://issuu.com/foi.stak/docs/stak22/s/11886533
[40] Usp. Dastin, Jeffrey. Amazon scraps secret AI recruiting tool that showed bias against women, 11. 10. 2018. URL: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G
[41] Usp. Noble, Safiya. Google Has a Striking History of Bias Against Black Girls, 26. 3. 2018. URL: https://time.com/5209144/google-search-engine-algorithm-bias-racism/

Google je ponovno morao prilagođavati svoje algoritme nakon pojavljivanja govora mržnje u ponuđenim rezultatima nakon upisivanja određenih upita u tražilicu kao što je „are Jews...“. Također, nakon što bi neki korisnik u tražilicu upisao „did Holocaust happen“ prvi rezultat vodio bi na bjelačke supermacijsku stranicu, a slične bi se stvari događale i s pretraživanjem sličnih upita vezanih uz ostale etničke manjine.[42]

5. Uloga umjetne inteligencije u stabilnosti informacijskog okoliša

Društvene mreže danas se oslanjaju na algoritme koji samostalno uklanjaju bot profile, „trolove“ i druge oblike osjetljivog sadržaja. Prema podacima iz Facebooka, 99.5% objava povezanih uz terorizam, 98.5% lažnih profila, 96% sadržaja seksualnog karaktera i 86% sadržaja koji sadrži nasilje uklonjeno je od strane algoritama koji su istrenirani od strane grupe ljudi koji im selektivno sabiru podatke za obradu kako bi što efektivnije odrađivali svoju zadaću. U Facebooku se nadaju da će ovakav način rada moći primijeniti za uklanjanje lažnih informacija i onih informacija za koje je dokazano da su raskrinkane.[43]

Za uklanjanje neprikladnog sadržaja na društvenim mrežama zaduženi su također i moderatori sadržaja. Trenutno ta zadaća još uvijek zahtijeva ljudski nadzor i ne može u potpunosti prepustiti strojevima. Tijekom vremena izolacije prvih nekoliko mjeseci pandemije bolesti COVID-a 19, algoritmi su sami otkrivali i uklanjali neprikladni sadržaj, ali ubrzo je bilo otkriveno kako su bez ljudskog nadzora uklanjali previše sadržaja, uključujući i onaj čije uklanjanje nije bilo potrebno.

Daljnji je problem količina sadržaja. Na dnevnoj bazi Facebook, Twitter i YouTube dosegnu više od milijardu objava, a svaku od njih je potrebno pregledati. Samo na Facebooku je prijavljeno 3 milijuna objava dnevno.[44] U prva 3 mjeseca 2020. godine na Facebooku je uklonjeno i obilježeno kao neprimjereno više od 3 milijarde objava, od čega je većinski prijavljeni sadržaj bio spam i lažni profili. U isto vremenskom razdoblju, algoritmi su samostalno

[42] Usp. Williams, Weston. Google updates algorithm to filter out Holocaust denial and hate sites, 21. 12. 2016. URL: https://www.csmonitor.com/Technology/2016/1221/Google-updates-algorithm-to-filter-out-Holocaust-denial-and-hate-sites

[43] Usp. Kertysova, Katarina. Nav.dj. Str. 59

[44] Usp. Barret, Paul M. Who Moderates the Social Media Giants?: A Call to End Outsourcing. NYU Stern Center for Business and Human Rights, 2020. Str. 3-4.
URL: https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5ed9854bf618c710cb55be98/1591313740497/NYU+Content+Moderation+Report_June+8+2020.pdf

uklonili, bez da je stvarna osoba prijavila, 99.5% dječje golotinje i ostalog seksualnog sadržaja vezanog uz djecu, 99% nasilnog sadržaja, 97.7% objava vezanih uz samoubojstvo, 88.8% govora mržnje i tek 15.6% objava vezanih uz nasilničko ponašanje i uznemiravanje. Istraživanje je također pokazalo da je na YouTube-u tijekom prva 3 mjeseca 2019. godine uklonjeno 5.8 milijuna videa, 540 milijuna komentara i nešto više od 2 milijuna kanala, dok je na Twitteru u prvoj polovici 2019. godine uklonjeno 1.2 milijuna profila. Za usporedbu, dok Facebook i YouTube imaju najmanje uklonjenog sadržaja vezanog uz nasilničko ponašanje i uznemiravanje, na Twitteru je to glavni razlog zbog čega je neki profil uklonjen, a 50% tih tweetova je prepoznato od strane algoritama.[45]

Danas je u svijetu oko 15 000 moderatora sadržaja koji rade za Facebook, Instagram, Youtube, Twitter i ostale kompanije koji ručno, bez pomoći tehnologije, moderiraju sadržaj na društvenim mrežama.[46] Kao jeftina radna snaga koja zarađuje nešto manje od 29 000 dolara godišnje, dok prosječni zaposlenik Facebooka zarađuje 240 000 dolara godišnje, alternativni su izbor za hibridni pristup moderiranja sadržaja, a zapošljavaju se kako bi detaljnije pregledavali neželjeni sadržaj koji je prijavljen od strane korisnika. Taj sadržaj uključuje ekstreme interneta kao što su ubojstva, govori mrženje, grafički prikazi pornografije i sl. U Facebooku je zato prije zapošljavanja potrebno najprije proći 3 godine treninga kako bi se osobe mentalno pripremile za izazove koje ih čekaju. Zbog velike količine traumatičnog sadržaja, općeg osjećaja straha i stresa i vremenskog pritiska, to je posao koji ostavlja velike posljedice na one koji su odluče njime baviti koje mogu toliko ekstremne da zaposlenici počinju razvijati simptome PTSP-a i ostalih mentalnih poremećaja, dok cijelo vrijeme u Facebooku negiraju pojavu takvih posljedica navođenjem kako je zaposlenicima uvijek na raspolaganju sva potrebna pomoć.[47]

S obzirom da će algoritmima trebati još vremena da dođu na razinu gdje samostalno mogu pregledavati i moderirati sadržaj te da taj isti posao ostavlja velike posljedice na ljude, postavlja se pitanje koji je način moderiranja sadržaja na društvenim mrežama najučinkovitiji, kako doći do tog cilja bez ostavljanja trajnih posljedica na zaposlenike te kako stvoriti okruženje u kojemu se korisnicima društvenih mreža uopće ne prikazuje uznemirujući, pogrešan ili lažan sadržaj.

---

[45] Usp. Isto. Str. 10-11.
[46] Usp. Stanarević Katavić, Snježana. Tehnološki aspekti informacijskog poremećaja. Kritička informacijska pismenost. Sveučilište J. J. Strossmayer, Filozofski fakultet, Odsjek za informacijske znanosti. Osijek, 11. 4. 2022. [Predavanje] (2022-08-05)
[47] Usp. Newton, Casey. The Trauma Floor: The Secret lives of Facebook moderators in America, 25. 2. 2019. URL: https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

6. Zaključak

Znanstvenici su prepoznali da su društveni mediji postali ključne platforme na kojima se narušava stabilnosti informacijskog okoliša jer je uočeno koliko je lako stvoriti informaciju koja je netočna, a koju će korisnici medija širiti bez provjeravanja njezine točnosti dokle god je ona korisna za stranu koju podržavaju. Znanstvenici su svjesni da postoji potreba za podizanjem svijesti o razlikovanju točnih informacija od lažnih, ali isto tako i za stvaranjem osjećaja odgovornosti na društvenim mrežama koje nude veliku slobodu za objavljivanje sadržaja. Unatoč ozbiljnim posljedicama do kojih širenje lažnih i netočnih informacija može dovesti kao što su gubitak povjerenja u institucije i medije, ugrožavanje javnog zdravlja, ekonomske dobiti, demokratskih procesa itd., stručnjaci smatraju da se protiv informacijskog poremećaja ne treba boriti koristeći rigidnu cenzuru, nego korištenjem raznih protokola i smjernica kod kojih se očekuje da korisnici sami shvate najprije opasnosti informacijskog poremećaja, a zatim i što poduzeti u njegovu suzbijanju te u prepoznavanju njegovih kreatora.

Iako su korisnici kao kreatori, primatelji i širitelji lažnih i netočnih informacija ključni akteri u širenju i jačanju informacijskog poremećaja, automatizacijske tehnologije kreirane s ciljem unapređenja obrade informacija, također imaju nezanemarivu ulogu, kako je i opisano u brojnim primjerima u ovome radu. Daljnjim razvojem i pojavom novih tehnologija, elementi informacijskog poremećaja ostat će prisutni u digitalnom okruženju, ali je za očekivati da će i tehnološki odgovori biti snažniji i učinkovitiji. Iako se rad fokusira na moguće opasnosti koje mogu proizaći iz uporabe umjetne inteligencije u upravljanju i obradi informacija, umjetna je inteligencija i dalje jedna od najboljih opcija za suzbijanje informacijskog poremećaja. Znanstvenici će morati pronaći način kako trenirati algoritme koji će efektivno uklanjati nepoželjni sadržaj na internetu, ali do tog cilja bi moglo proći još neko vrijeme. Umjetna je inteligencija nepristrano i isplativo rješenje za provjeravanje informacija na društvenim mrežama, no puna automatizacija još je uvijek daleki cilj, te se za sada ostaje na hibridnom modelu korištenja i tehnologije i ljudi. Jedan od razloga zašto je ljudski nadzor potreban je taj što su trenutačni algoritmi još uvijek limitirani u smislu identificiranja izjava što znači da ne mogu prepoznati kontekst u kojem se neka riječ nalazi, a proći će još dosta vremena dok se algoritme

ne nauče koncepti poput ironije i sarkazma, te posebni kulturološki i politički konteksti, a također se javlja i prijašnje spomenuti problem algoritamske pristranosti.[48]

7. Literatura

1. About automated account labels. URL: https://help.twitter.com/en/using-twitter/automated-account-labels (2022-07-11)

2. About goverment and state-affiliated media account labels on Twitter. URL: https://help.twitter.com/en/rules-and-policies/state-affiliated (2022-07-12)

3. Albahar, Marwan; Almalki, Jameel. Deepfakes: threats and countermeasures: systematic review // Journal of Theoretical and Applied Information Technology 97, 22 (2019), str. 3243. URL: http://www.jatit.org/volumes/Vol97No22/7Vol97No22.pdf (2022-07-11)

4. Arakerimath, Anjana R; Gupta, Pramod Kumar. Digital Footprint: Pros, Cons and Future // International Journal of Latest Technology in Engineering, Managment & Applied Science 4, 10 (2015), str. 52-53. URL: https://d1wqtxts1xzle7.cloudfront.net/39569967/52-56-with-cover-page-v2.pdf?

[48] Usp. Kertysova, Katarina. Nav. dj., str. 60-61.

Expires=1659999043&Signature=gzkBPmSddBC~DWGKe8823Zotqa3NhQG26vAYnOBcQZbNNv38-XPCdtklVDm7B2cPveG-QdL1FSgELX0b6kBznY4mpTKsYmFLxDlHBKsrZJHebCQXgZLZSkLnLgFU9ZvjNZX~6SKkjbIdqvHHLxOTzQN41ZbLhlW3P4k3Xot4rs-IXX0ItcNotBfMPoYzOOu-Zh4IlpXmALd3pqTp17yYAO5rL5rHBNQyY8KKjTGfQlv82MfXOvCvfcK~xAPOqTwB6Dj4K3Q18EAQJIQoJgFxX1tflZnIulUgrjtsh7ZAfD3fGzofWqk81wEecpkW4RU3Te-~xUNmKINFO5JVR28CzA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA (2022-07-12)

5. Araújo, Ana Cristina The Lisbon Earthquake of 1755, 2006. URL: https://www.brown.edu/Departments/Portuguese_Brazilian_Studies/ejph/html/issue7/html/aaraujo_main.html (2022-08-30)

6. Barbu, Oana. Advertising, Microtargeting and Social Media // Procedia – Social and Behavioral Sciences, 163, 2014, str. 44-48. URL: https://www.sciencedirect.com/science/article/pii/S187704281406385X (2022-07-12)

7. Barret, Paul M. Who Moderates the Social Media Giants?: A Call to End Outsourcing. NYU Stern Center for Business and Human Rights, 2020. Str. 3-4, 10-11. URL: https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5ed9854bf618c710cb55be98/1591313740497/NYU+Content+Moderation+Report_June+8+2020.pdf (2022-08-11)

8. Cherry, Kendra. What Is Cognitive Bias?, 19. 7. 2020. URL: https://www.verywellmind.com/what-is-a-cognitive-bias-2794963 (2022-07-12)

9. Dastin, Jeffrey. Amazon scraps secret AI recruiting tool that showed bias against women, 11. 10. 2018. URL: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (2022-08-08)

10. Dezinformacije i propaganda na internetu, CERT.hr-PUBDOC-2019-3-375, 2019. URL: https://www.cert.hr/wp-content/uploads/2019/03/dezinformacije_i_propaganda_na_Internetu.pdf (2022-07-11)

11. Dokler, Ana. Ako koristite Google, Facebook ili Twitter, morate znati kako djeluju algoritmi, 12. 11. 2019. URL: https://www.medijskapismenost.hr/ako-koristite-google-facebook-twitter-trebate-znati-kako-djeluju-algoritmi/ (2022-07-12)

12. Dubois, Elizabeth; Blank, Grant. The myth of echo chambers, 9. 3. 2018. URL: https://www.oii.ox.ac.uk/news-events/news/the-myth-of-the-echo-chamber/ (2022-07-12)

13. Dumbrava, Costica. Key social media risks to democracy: Risks from surveillance, personalisation, disinformation, moderation and microtargeting. Brussels: European Parliamentary Research Service, 2021. URL: https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/698845/EPRS_IDA(2021)698845_EN.pdf (2022-07-11)

14. Fletcher, Richard. The truth behind filter bubbles: Bursting some myths. // Reuters Institute, 24. 1. 2020. URL: https://reutersinstitute.politics.ox.ac.uk/news/truth-behind-filter-bubbles-bursting-some-myths (2022-07-12)

15. Frau-Meigs, Divina. Information Disorders: Risks and Opportunities for Digital Media and Information Literacy. // Medijske studije 10, 19 (2019), str. 16-18. URL: https://hrcak.srce.hr/file/330016 (2022-07-10)

16. Gesenhues, Amy. Survey: 90% Of Customers Say Buying Decisions Are Influenced By Online Reviews, 9. 4. 2013. URL: https://martech.org/survey-customers-more-frustrated-

by-how-long-it-takes-to-resolve-a-customer-service-issue-than-the-resolution/ (2022-08-08)

17. Kanoje, Sumitkumar; Girase, Sheetal; Mukhopadhyay, Debajyoti. User Profiling Trends, Techniques and Applications // International Journal of Advance Foundation and Reserch in Computer 1, 1 (2014), str. 2-4. URL: https://arxiv.org/ftp/arxiv/papers/1503/1503.07474.pdf (2022-07-12)

18. Kandel, Nirmal: Information Disorder Syndrome and Its Management. // Journal of Nepal Medical Association 58, 224 (2020), str. 281. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7580464/pdf/JNMA-58-224-280.pdf

19. Kertysova, Katarina. Artificial Intelligence and Disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered // Security and Human Rights 29, 1-4 (2018), str. 59-67. URL: https://brill.com/downloadpdf/journals/shrs/29/1-4/article-p55_55.xml (2022-07-11)

20. Listeš, Daniela. Algoritamska pristanost. URL: https://issuu.com/foi.stak/docs/stak22/s/11886533 (2022-07-12)

21. Lokot, Tetyana; Diakopoulos, Nicholas. News Bots: Automating news and information dissemination on Twitter. // Digital Journalism 4, 6 (2016), str. 2-7. URL: http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/newsbots_final.pdf (2022-07-11)

22. Nenadić, Iva; Vučković, Milica. Dezinformacije: Edukativna brošura i vježbe za razumijevanje problema dezinformacije. // Agencija za elektroničke medije i UNICEF, Zagreb, 2021. Str. 9-10. URL:

https://www.medijskapismenost.hr/wp-content/uploads/2021/04/brosura-Dezinformacije.pdf (2022-08-05)

23. Newton, Casey. The Trauma Floor: The Secret lives of Facebook moderators in America, 25. 2. 2019. URL: https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

24. Noble, Safiya. Google Has a Striking History of Bias Against Black Girls, 26. 3. 2018. URL: https://time.com/5209144/google-search-engine-algorithm-bias-racism/ (2022-08-08)

25. Powers, Shawn; Kounalakis, Markos. Can Public Diplomacy Survive the Internet?: Bots, Echo Chambers and Disinformation, 2017. Str. 18. URL: https://www.state.gov/wp-content/uploads/2019/05/2017-ACPD-Internet.pdf (2022-08-08)

26. Rajan. Amol. Do digital echo chambers exist?, 4. 3. 2019. URL: https://www.bbc.com/news/entertainment-arts-47447633 (2022-07-12)

27. Stanarević Katavić, Snježana. Informacijski poremećaj – osnovni pojmovi. Kritička informacijska pismenost. Sveučilište J. J. Storssmayer, Filozofski fakultet, Odsjek za informacijske znanosti. Osijek, 7. 3. 2022. [Predavanje] (2022-08-05)

28. Stanarević Katavić, Snježana. Tehnološki aspekti informacijskog poremećaja. Kritička informacijska pismenost. Sveučilište J. J. Strossmayer, Filozofski fakultet, Odsjek za informacijske znanosti. Osijek, 11. 4. 2022. [Predavanje] (2022-08-05)

29. Thomas, Elise. Amazon's Algorithms, Conspiracy Theories and Extremist Literature, 4. 5. 2021. URL: https://www.isdglobal.org/digital_dispatches/amazons-algorithms-conspiracy-theories-and-extremist-literature/ (2022-08-31)

30. Vote Leave's targeted Brexit ads released by Facebook, 2018. URL: https://www.bbc.com/news/uk-politics-44966969

31. Wardle, Claire; Derakhshan, Hossein. Information disorder: Toward an interdisciplinary framework for research and policy making, 2017. URL: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c (2022-08-19)

32. What are bots? – Definiton and Explanation. URL: https://www.kaspersky.com/resource-center/definitions/what-are-bots (2022-07-11)

33. Williams, Weston. Google updates algorithm to filter out Holocaust denial and hate sites, 21. 12. 2016. URL: https://www.csmonitor.com/Technology/2016/1221/Google-updates-algorithm-to-filter-out-Holocaust-denial-and-hate-sites (2022-08-08)