UDK 323.266:004.738.5
Biblid: 0025-8555, 75(2023)
Vol. LXXV, br. 4, str. 685–710
DOI: https://doi.org/10.2298/MEDJP2304685K

Pregledni rad
Primljen 31. 8. 2023.
Odobren 5. 10. 2023.

Veštačka inteligencija za otkrivanje lažnih vesti na socijalnim medijima – ispitivanje stavova

Ana KOVAČEVIĆ[1], Emir DEMIĆ[2]

Apstrakt: U današnjem digitalnom dobu, socijalni mediji (društvene mreže) imaju značajnu ulogu u širenju informacija. Međutim, pojave kao što su lažne vesti mogu značajno narušiti poverenje korisnika. U isto vreme, veštačka inteligencija otvara nove izvanredne mogućnosti, dok donosi i značajne izazove. Veštačka inteligencija se već primenjuje u otkrivanju lažnih vesti na društvenim mrežama, čineći ovo značajnim područjem istraživanja zbog velikog uticaja na društvo i politiku. Ovaj rad proučava primenu veštačke inteligencije na društvenim mrežama, s posebnim osvrtom na otkrivanje lažnih vesti, i to kroz analizu stavova studentske populacije. Pored toga, ispituju se razlike između starijih i mlađih studenata u njihovim odgovorima. Rezultati ukazuju na ograničenu upućenost studenata u ovu vrstu programa, čime se nameće potreba za obrađivanjem ovih tema u budućim univerzitetskim nastavnim planovima. Veći broj studenata smatra da su algoritmi veštačke inteligencije za otkrivanje lažnih vesti korisni za društvo u odnosu na one koji smatraju suprotno. Rezultati ukazuju da postoji zabrinutost u pogledu moguće zloupotrebe ovih tehnologija i njihovog uticaja na slobodu govora, stoga se ističe potreba za nepristrasnim nadzorom i regulativom. Istraživanje je pokazalo da su studenti svesni prednosti i ograničenja primene veštačke inteligencije na društvenim mrežama, a stariji studenti pokazuju veću otvorenost prema tim programima.

Ključne reči: rizici, mašinsko učenje, statistika, upitnik, algoritmi, Srbija.

---

[1] Fakultet bezbednosti, Univerzitet u Beogradu, vanredni profesor, kana@fb.bg.ac.rs, ORCID: 0000-0003-4928-9848

Rad je rezultat naučog projekta koji finansira Fond za nauku Republike Srbije u okviru Programa "IDEJE" – Management of New Security Risks Research and Simulation Development, NEWSIMR&D, #7749151.

[2] s.dg.emir@gmail.com, ORCID: 0000-0001-7825-026X.

Uvod

Socijalni mediji (društvene mreže) su postali integralni deo savremenog društva pružajući obilje podataka i vesti. Ne postoji ništa slično platformama društvenih mreža, a da je tako brzo preplavilo svet (Singer & Brookings 2018). U istraživanju sprovedenom 2023. godine zabeleženo je 4,88 milijarde aktivnih korisničkih naloga na socijalnim medijima, što čini 60,6% svetske populacije, dok ima 5,18 milijardi Internet korisnika, što čini 64,6% svetske populacije (Kemp 2023b). Koliki je rast korišćenja ovih platformi najbolje govori podatak da je 2015. godine, samo pre 8 godina, 30% svetske populacije koristilo socijalne medije, te da taj broj i dalje raste (Kemp 2023b). Najpopularnije društvene mreže prema izveštaju iz aprila 2023. godine, rangirano prema broju aktivnih korisnika, su: Facebook (Fejsbuk, 2,96 milijarde korisnika), YouTube (Jutjub, 2,52 milijarde), WhatsApp (Vatsap) i Instagram (2 milijarde korisnika) (Kemp 2023a). Istraživanje pokazuje da korisnici provedu prosečno u toku jednog dana 2 sata i 26 minuta na socijalnim medijima (Kemp 2023b). Prema istraživanju Republičkog zavoda za statistiku Srbije, internet koristi 83,5% ispitanika (ili 100% ispitanih studenata), a od tog broja preko 81% internet populacije ima nalog na socijalnim društvenim platformama (Kovačević i dr. 2022).

Pored upotrebe navedenih medija radi ostvarivanja kontakta, korisnici neretko takve aplikacije koriste i radi informisanja. Prema istraživanju sprovedenom 2020. godine izneto je da 36% ispitanika redovno dobija vesti preko Fejsbuka i 23% preko Jutjuba, a broj onih na koje vesti sa ovih platformi imaju indirektno uticaja na razne aspekte života je verovatno mnogo veći (Shearer & Mitchell, 2021). Međutim, važno je napomenuti da prisustvo informacija na socijalnim medijima može dovesti do izazova u vezi sa verodostojnošću, tačnošću i interpretacijom istih. Povećanje broja korisnika, kao i eksplozivni rast količine objavljenih informacija, uz primenu generativne veštačke inteligencije (poput GPT3), uticali su da lažne vesti postaju značajan izazov sa potencijalno ozbiljnim društenim i političkim implikacijama. Otkrivanje tih lažnih vesti postaje sve teže, na šta ukazuju i Solejman i saradnici (Solaiman et al 2023).

Količina podataka koja se svakodnevno generiše prevazilazi mogućnosti ljudske analize. Jedno od mogućih rešenja ovog problema bi bila primena veštačke inteligencije, koja bi imali dvojaku ulogu: klasifikaciju vesti prema verodostojnosti, kao i prikazivanje vesti odgovarajućim korisnicima. Na društvenim platformama se sve više koristi veštačka inteligencija da bi se upravljalo mnogim aspektima online okruženja (Oremus et al. 2021; Rainie et al. 2022). To stvara brojne prednosti, ali i nove rizike.

Programi veštačke inteligencije za uklanjanje lažnih vesti imaju za cilj identifikaciju, filtriranje i ograničenje njihovog širenja. Prilikom korišćenja ovih

algoritama nameću se brojna pitanja: Kolika je mogućnost greške? Postoji li rizik od netačne klasifikacije što dovodi do cenzure ili uklanjanja tačnih informacija? Da li su, i ukoliko da, koliko su transparentni algoritmi koje koriste velike kompanije da omogućavaju uvid u klasifikaciju vesti i/ili zaključivanja? Pored toga nameću se i pitanja ko ima kontrolu nad programom i kako se odlučuje koji će sadržaj biti uklonjen? Da li komercijalni i/ili politički interesi i/ili pritisak mogu dovesti do toga da se neki sadržaj ukloni? Da li se ovim narušava sloboda govora? Implementacija nosi sa sobom izazove u vezi sa definisanjem lažnih vesti, ocenom konteksta i slobodom izražavanja, a pitanje transparentnosti i odgovornosti u vezi sa programom postaje ključno.

Veštačka inteligencija – primena na društvenim mrežama

Ne postoji opšte prihvaćena definicija veštačke inteligencije, a brojne definicije su navedene kod Skuta (Schuett 2023). U radu ćemo predstaviti definiciju koju je dao Minski (Minsky 1968), jedan od pionira veštačke inteligencije, kao nauku činjenica gde mašine rade stvari koje bi zahtevale inteligenciju ukoliko bi ih radio čovek. Mašinsko učenje predstavlja oblast veštačke inteligencije koja ima sposobnost samostalnog učenja zakonitosti iz podataka (Kovačević 2023). Prema formalnoj definiciji mašinskog učenja smatra se da kompjuterski program uči iz iskustva (I), vezanog za zadatak (Z) i meru performansi (P), ukoliko se njegove performanse na zadatku (P), merene metrikama (M), unapređuju sa iskustvom (I) (Mitchell 1997). Podela mašinskog učenja na osnovu tipa odlučivanja može biti nadgledano (supervised learning), nenadgledano učenje (unsupervised learning) i učenje uz podsticaj (reinforcement learning).

Primena veštačke inteligencije na socijalnim medijima ima za cilj rast poslovanja i povećanje zadovoljstva korisnika (Kenyon 2021). Idetifikuju se obrasci ponašanja i interesovanja korisnika i na osnovu toga im se isporučuje sadržaj. Isporučeni sadržaj utiče na stvaranje takozvane *eho komore* (eho sobe, balon filtera), gde korisnici uglavnom dobijaju sadržaj koji potvrđuje njihove postojeće stavove i interesovanja.

Rezultat algoritama veštačke inteligencije umnogome zavisi od ulaznih podataka. Zato je izuzetno je važno koji ulazni podaci se koriste, jer nedostatak raznolikosti u ulaznim podacima može uticati na pojavu pristrasnosti u zaključivanju. Na primer, ako su algoritmi trenirani na podacima o belim muškarcima, rezultat će pokazivati pristrasnost protiv drugih (manjina ili žena). U literaturi se ukazuje na probleme pristrasnosti koji se javljaju kod algoritama veštačke inteligencije, poput rasizma (Fong 2021; Dwoskin et al. 2021) ili cenzure od strane socijalnih medija (Belli 2021). Čak i naizgled neutralni skupovi podataka, ukoliko se ukrste sa drugim podacima, mogu proizvesti rezultate koje diskriminuši po rasi, polu i uzrastu (Lohr

2021). S druge strane, pristalice alogoritama tvrde da su automatski sistemi manje podložni diskriminaciji (Burke et al. 2021).

Društvene platforme povećavaju vidljivost određenih postova (tzv. boosting ili busting), time što ih prikazuju većem broju ljudi nego što bi bilo bez bustinga. Cilj ovoga je povećanje angažovanja korisnika kroz interakciju: lajkove, deljenje, komentare i klikove. Na ovaj način se generišu kontroverzne i senzacionalističke teme koje uzrokuju emocionalne reakcije, a zanemaruju se tačnost ili relevantnost. Zbog prethodno navedenog, izuzetno je važno adaptirati algoritme da dele autentičan i odgovarajući sadržaj, dok je važno ograničiti uticaj lažnih vesti i provokativnog sadržaja. Na osnovu kombinovanja internih dokumenata Fejsbuka, javno dostupnih informacija i izjava insajdera dobija se uvid u to kako različiti pristupi algoritmu mogu značajno da promene kategorije sadržaja koje su vidljive korisniku (Oremus et al. 2021). Pobornici algoritama smatraju da bi uklanjanjem algoritama bilo više, a ne manje govora mržnje i dezinformacija **(**Vakil 2021). U odsustvu algoritma za filtriranje sadržaja, dominantnu poziciju bi zauzeli pojedinci ili organizacije koje objavljuju najčešće i koje imaju najširu publiku, ali to ne bi garantovalo kvalitet sadržaja. U takvom scenariju, objave sa manjim brojem pratilaca suočavale bi se sa smanjenim mogućnostima da dopru do šire publike.

Zlonamerni korisnici mogu koristiti generativnu veštačku inteligenciju (poput Chat GPT) za kreiranje sadržaja za podršku kampanjama, političkim agendama ili širenje mržnje. Generisanje lažnih informacija može imati cikličan efekat, rezultujući sve većim obimom dezinformacija. Odnosno, generativni sistemi koji se budu obučavali nad lažnim podacima putem metode učenja uz podsticaj generisnće nove netačne izlaze, što može da ubrza širenje lažnih vesti, utiče na javno mnjenje, uznemirava pojedince ili utiče na politiku i izbore (Ferguson et al. 2023). U ubedljivom tekstu koji je generisan pomoću veštačke inteligencije jedan profesor prava se našao na listi pravnika koji su seksualno uznemiravali, iako takva tvrdnja nije postojala (Verma i Oremus, 2023). Takođe, su bili navedeni njegovi nepostojeći radovi, koji su izgledali ubedljivo. Generativni sistemi veštačke inteligencije mogu da proizvedu štetne informacije i bez namere.

Sofisticirani generativini sistemi veštačke inteligencije mogu imati izuzetno uverljive rezultate, a nivo rizika zavisi od konkretnog slučaja upotrebe, od lažnog predstavljanja do kampanja dezinformisanja (Solaiman et al 2023), posledice pokušaja opasnih tretmana tokom Covida-19 i podsticanja nasilja (Merrill et al. 2022). Businka i saradnici (Bucinca et al. 2021) su zaključili da su pojedinci skloni preuveličavanju i višem stepenu poverenja prema sadržaju koji je generisan, naročito ako je rezultat autoritativan ili se nalaze u situaciji koja je vremenski osetljiva. Indirektna posledica sofisticirane generativne veštačke inteligencije poslednjih godina je proširila mogućnost kampanja lažnih vesti i čini da ljudi teže

veruju pravim vestima ili onome što čuju/vide (Buchanan et al 2021; Guess et al. 2021). Kad sadržaj generisan veštačkom inteligencijom bude uobičajen, nećemo više znati šta je istina, a šta je laž.

Koliko je važno da se uspostavi pravni okvir za primenu veštačke inteligencije pokazuje Evropski parlament koji je u junu 2023. godine sa velikom većinom glasova izglasao Pregovaračku poziciju o nacrtu Zakona o veštačkoj inteligenciji (Marcin 2023). Prema navedenom zakonu, obavezno je da se za sisteme veštačke inteligencije opšte namene, pre njihovog korišćenja na tržištu Evropske unije, sprovede procena i ublažavanje potencijalnih rizika, te da se njihovi modeli registruju u bazi podataka Evropske unije. Dodatno, od suštinskog je značaja da sadržaj kreiran pomoću generativnih sistema veštačke inteligencije nosi odgovarajuću oznaku o generisanju, da uključuje informacije o autorstvu korišćenih podataka, kao i da izbegava generisanje nelegalnog sadržaja (MEP 2023).

Lažne vesti na društvenim platformama

Postojanje lažnih vesti nije novi fenomen, no sa socijalnim medijima doživljava ekspanziju. Društveni mediji imaju veliku prednost u širenju vesti: jeftini su i dostupni za brzo širenje informacija. Na primer, tokom pandemije COVID-19 su se širile razne glasine, poput toga da tadašnji predsednik Sjedinjenih Američkih Država (SAD) razmatra nacionalnu blokadu granica ili lažna lista medicinskih saveta Univerziteta Stenford (Statt 2020). Koliko je značajan problem masovne digitalne dezinformisanosti govori i činjenica da ga je još 2013. godine Svetski ekonomski forum označio kao značajan tehnološki rizik (Howell 2013).

Vosagi i saradnici (Vosoughi et al. 2018) su sproveli opsežno istraživanje o širenju lažnih i istinitih vesti na Twitteru (Tviter) u periodu od 2006. do 2017, gde je analizirano oko 126,000 tvitova koje je oko 3 miliona ljudi podelilo više od 4,5 miliona puta. Autori su zaključili da se lažne vesti šire brže i imaju širu publiku nego istinite, pri čemu se pokazalo da lažne političke vesti imaju veći uticaj nego lažne vesti o terorizmu, prirodnim katastrofama, nauci, urbanim legendama ili finansijskim informacijama (Vosoughi et al. 2018).

Istraživanje kojim je obuhvaćeno 171 miliona tvitova iz perioda poslednjih pet meseci pred izbore u SAD 2016. godine je pokazalo da je od toga 30 miliona tvitova (2,3 miliona korisnika) sadržalo linkove do sajta sa vestima, od čega je četvrtina (7,5 miliona tvitova) delilo informacije koje su neistinite ili ekstremno pristrasne (Bovet & Makse 2019). Ovakav obim lažnih vesti je imao uticaja na izbore. Tokom poslednja tri meseca predsedničkih izbora u SAD 2016. godine, dvadeset najuspešnijih lažnih vesti u vezi sa izborima je izazvalo više deljenja, komentara i

reakcija na Fejsbuku nego dvadeset najuspešnijih vesti novinarskih izvora (Hughes & Waismel-Manor, 2020). U skladu sa tim istraživanje Tvitera pokazuje da je 70% veća verovatnoća da se lažna vest ponovo podeli nego tačna vest, kao i da lažna vest dostiže niz od deset ponovnih deljenja dvadeset puta brže nego tačna (Dizikes 2018). Slično se dešava i na drugim platformama, pa Silverman (2016) navodi da se tokom predsedničkih izbora u SAD 2016 godine lažna vest brže širila na Fejsbuku nego najpopularnija autentična vest.

Fenomen lažnih vesti predstavlja predmet istraživanja u različitim oblastima, pri čemu su brojni autori sugerisali niz definicija za ovaj termin. Šu i saradnici (Shu et al. 2017) definišu lažne vesti u užem smislu, kao vest koja je kreirana s lošom namerom i da je proverljivo netačna. Preciznije, lažnu vest karakteriše to da je informacija lažna i da je moguće proveriti njenu netačnost, te da je takva informacija kreirana sa nepoštenom namerom kako bi se konzument informacije doveo u zabunu. Drugi autori zagovaraju širu definiciju lažnih vesti, koja se fokusira ili na verodostojnost ili na nameru, kao što su sledeće kategorije (Shu et al. 2017):

- dezinformacije – koje su nenamerno kreirane (Balmas, 2012);
- teorije zavere – teško je proveriti njihovu tačnost (Ash, 1951);
- glasine – ne potiču od vesti, tvrdnje o činjenicama koje se nisu pokazale kao istinite, ali se prenose od jedne do druge osobe i njihov kredibilitet se zasniva na tome što drugi ljudi veruju u njih (Allcott & Gentzkow 2017; Sunstein 2007);
- satirične vesti – imaju lažan sadržaj, ali često i bez namere da obmanu korisnika, no u njih se može poverovati ako sadržaj satire nije poznat ili jasan (Balmas 2012; Rubin 2016, Jin 2016);
- obmane – imaju nestinit sadržaj sa elementima prevare ili zabave.

Drugi autori su naveli i drugačije podele. Na primer, Tandok i saradnici (Tandoc et al. 2018) lažne vesti dele u sledeće kategorije: satire, parodije, fabrikovanje, manipulacije, oglašavanje i propagandu. S druge strane, Mišra i saradnici (Mishra et al., 2001) pod lažnim vestima smatraju obmane, fabrikovane i satirične vesti. U daljem tekstu pod lažnim vestima smatraćemo lažne vesti u širem obliku kao kod Šua i saradnika (Shu et al., 2017).

Lažne vesti imaju tendenciju bržeg deljenja na socijalnim mrežama nego prave vesti. Dizajks (Dizikes, 2018) smatra da se lažne vesti brže šire od stvarnih, pošto ljudi vole novitete. Uz to lažne vesti izazivaju jače reakcije i to je razlog njihovog češćeg deljenja. Neki ljudi imaju više sklonosti ka deljenju lažnih vesti od istinitih. Pojedinci su dodatno podložni lažnim informacijama ukoliko su im izloženi više puta, pogotovo ako informacije dolaze od osoba iz njihove socijalne mreže. Dodatni razlog za širenje lažnih vesti može biti podložnost kognitivnim pristrasnostima, odnosno uklapanje novih informacija sa postojećim znanjem, što je naročito izraženo kod

preopterećenosti informacijama. Ljudi po prirodi nisu uspešni u razlikovanju stvarnih i lažnih vesti, pa se iskorišćavaju ranjivosti pojedinaca poput (Shu et al. 2017):

- naivnog realizma – korisnik pokušava da shvati da je njegova/njena percepcija realnosti jedino ispravna, dok su drugi koji ne misle tako neiformisani, iracionalni ili pristrasni (Ward et al. 1997);
- pristrasnosti prihvatanja - korisnik preferira da prima informacije koje potvrđuju njegove/njene postojeće poglede (Nickerson 1998).

Pored toga, na prihvatanje lažnih vesti imaju uticaja i sledeći psihološki faktori (Shu et al. 2017, Paul & Matthew's 2016):

- socijalni kredibilitet: pojedinac smatra izvor verodostojnim, ako drugi smatraju da je izvor kredibilan (pogotov ako nema dovoljno informacija da se proveri);
- frekvencija heuristike**:** pojedinac može prirodno davati prioritet informacijama koje često čuje, čak iako su lažne; odnosno često ponovljena laž postaje istina.

Ova pristrasnost se dodatno povećava tako što korisnici socijalnih medija dobijaju informacije koje odgovaraju njihovim uverenjima, što dodatno učvršćuje njihove stavove, a to vodi do dalje polarizacije i nastanka eho komora (Petković 2022). Na ovaj način manifestuje se tendencija sve dubljeg zatvaranje u sopstveno iskrivljeno uverenje putem deljenja lažnih informacija u dezinformisanoj eho komori, što dovodi do homogenih društava koja postaju primarni pokretač širenja informacija koje dalje jačaju polarizaciju (Menczer & Hills 2020; Del Vicario et al. 2016). Jednom formirano pogrešno shvatanje je teško ispraviti: studija pokazuje da pokušaj ispravljanja lažnih informacija predstavljanjem tačnih činjenica nekada može čak povećati uverenje u pogrešno shvatanje (Nyhan & Reier 2010). Dodatni problem sa lažnim vestima je da one mogu činiti da ljudi osećaju nepoverenje i zbunjenost, ne mogu da razdvoje istinu od laži i to utiče i na njihovo poimanje istinitih vesti (Lynch 2016).

Jednostavno i jeftino je kreirati nalog na socijalnim mrežama, a iza naloga ne mora da bude čovek. Pored regularnih korisnika, sa stanovišta širenja lažnih vesti, na društvenim medijima se mogu indentifikovati i ostali korisnici (Shu et al. 2017):

- Socijalni botovi – nalog na društvenim medijima koji kontroliše računarski algoritam za automatsko kreiranje sadržaja i komuniciranje sa ostalima. Može specifično biti kreiran za širenje lažnih vesti. Npr. tokom predsedničkih izbora u SAD socijalni botovi su imali veliki uticaj.
- Kiborzi – mogu širiti lažne vesti na način koji se automatski kombinuje sa ljudskim doprinosom. Jednostavno prebacivanje funkcionalnosti između čoveka i bota, omogućava jednostavno širenje lažnih vesti (Chu et al., 2012), pri čemu se najčešće čovek registruje a potom se postavljaju automatski programi za aktivnosti na socijalnim medijima.

- Trolovi – su realni korisnici koji hoće da poremete online okruženje putem namernog širenja emocionalno nabijenih poruka usmerenih ka izazivanju reakcija kod korisnika.

Istraživački ciljevi

U pregledu literature na temu ispitivanja stavova studenata o programima veštačke inteligencije koji se koriste u socijalnim medijima za detektovanje lažnih vesti, te o njihovim mišljenjima o regulativi takvih algoritama i ulogama koju bi različite zainteresovane strane trebalo da igraju u takvoj regulativi, autori ovog rada nisu našli istraživanje na uzorku iz Srbije. S obzirom na to da je kod nas regulativa algoritama veštačke inteligencije tek u razvoju, te da i samo polje veštačke inteligencije uživa veće interesovanje studenata i šire populacije, neophodno je ispitati stavove o korišćenju takvih algoritama, kako bi se na adekvatan način uticalo na javne politike i poboljšali univerzitetski nastavni planovi. Stoga, je cilj ovog istraživanja dvojak. Prvi cilj istraživanja se odnosi na ispitivanje stavova studenata o programima veštačke inteligencije koji se koriste u aplikacijama socijalnih medija u generalne svrhe i za detekciju lažnih vesti. Drugi istraživački cilj se odnosi na ispitivanje razlika u takvim stavovima između mlađih (studenata I godine) i starijih studenata (studenata IV godine), pod pretpostavkom da stariji studenti imaju razvijenije znanje i informisanost o takvim algoritmima od mlađih studenata.

Istraživanje opisano u ovom radu je deo šireg istraživanja koje se tiče stavova prema korišćenju algoritama veštačke inteligencije u oblastima izvan socijalnih medija. Međutim, u ovom radu su opisani rezultati na relevantnim stavkama koje obrađuju temu programa veštačke inteligencije za detekciju lažnih vesti u socijalnim medijima.

Kako bismo odgovorili na postavljene istraživačke ciljeve, istraživanje smo sproveli kroz online upitnik koji se sastojao od stavki prevedenih iz prethodnog istraživanja sprovedenog od strane Pew Research Center-a (Rainie et al. 2022). Podaci su prikupljeni na odgovarajućem uzorku studenata.

Metodologija

Uzorak se sastojao od 403 studenta, od kojih je 315 (78,2%) ženskog, a 88 (21,8%) muškog pola. Istraživanje je sprovedeno na Fakultetu bezbednosti Univerziteta u Beogradu. Svi studenti su u trenutku popunjavanja upitnika pohađali

kurs *Informatika* (318 studenata) na prvoj godini osnovnih studija, odnosno *Bezbednost informacija* (83 studenta) na četrvrtoj godini osnovnih studija.

Za uzorak smo izabrali studente pod pretpstavkom da je ta populacija dobro upoznata sa socijalnim medijima (Kovačević i dr. 2022). Pored toga izabrali smo da istraživanje sprovedemo nad studentima Fakulteta bezbednosti, Univerziteta u Beogradu, pošto ti studenti izučavaju različite aspekte bezbednosti (nacionalnu, stratešku, korporativnu i/ili ekološku) i uz pretpostavku da imaju viši nivo svesti o bezbednosti od studenata drugih fakulteta (Kovačević et al. 2020), pa i u širenju lažnih vesti na socijalnim mrežama. Pored toga, studenti Fakulteta bezbednosti postaju zainteresovane strane u donošenju javnih politika iz ovog domena, što predstavlja još jedan razlog prikupljanja podataka na takvom uzorku.

Instrument korišćen u ovom istraživanju sastoji se od pitanja koja su prevedena iz upitnika korišćenog u istraživanju Pew Research Center-a (Rainie et al. 2022). Celokupan upitnik se sastojao iz tri dela, dok su za ovaj rad relevantna njegova dva dela. Prvi deo se sastojao iz socio-demografskih pitanja koja su se odnosila na pol, uzrast ispitanika i godinu osnovnih studija koju ispitanici pohađaju. Drugi deo upitnika se sastoji iz stavki koje se odnose na stavove ispitanika prema korišćenju algoritama veštačke inteligencije i sastoji se iz četiri supskale - stavovi prema generalnoj primeni takvih algoritama, stavovi prema primeni takvih algoritama u aplikacijama socijalnih medija, stavovi prema primeni takvih algoritama za prepoznavanje lica, stavovi prema primeni takvih algoritama u samo-vozećim automobilima. U ovom radu su predstavljeni rezultati relevantnih stavki sa supskale kojom se ispituju stavovi prema primeni algoritama veštačke inteligencije u aplikacijama socijalnih medija, a koja su navedena u Prilogu 1.

Istraživanje je sprovedeno putem online platforme Moodle u periodu od 25. maja do 2. juna 2023 godine. Učešće u istraživanju je bilo dobrovoljno, a ispitanici su mogli da odustanu od učešća u istraživanju u bilo kom momentu. Ispitanici su najpre ukratko upoznati sa temom istraživanja, a potom su popunili upitnik. Za popunjavanje celokupnog upitnika je trebalo oko 20 minuta, prilikom čega vreme za popunjavanje nije bilo ograničeno.

Rezultati

Podaci su obrađeni uz pomoć softvera Statistical Package for Social Sciences (SPSS) v29, kao i uz pomoć Python programskog jezika. Deskriptivne mere su ekstrahovane za sve varijable, prilikom čega smo koristili prosek i standardnu devijaciju kao deskriptivne mere kontinualnih varijabli, a proseke i frekvencije kao

deskriptivne mere kategoričkih varijabli. Za testiranje razlika između različitih demografskih grupa koristili smo hi-kvadrat test i korigovane standardizovane reziduale za interpretaciju rezultata, prilikom čega je korišćena apsolutna vrednost korigovanih standardizovanih reziduala od 2 za beleženje odstupanja odgovora od očekivane distribucije. Poređenja sa rezultatima istraživanja sprovedenog od strane Pew Research Center-a (Rainie et al. 2022) su navedena, ali zbog neujednačenosti veličine uzoraka i drugačijeg načina uzorkovanja, nisu sprovedena statistička poređenja ovih rezultata.

Na podacima korišćenog uzorka, čak 25,3% ispitanika je izvestilo da nisu čitali niti čuli o računarskim programima koje kompanije socijalnih medija koriste za pronalaženje lažnih informacija na svojim sajtovima. S druge strane, 64% uzorka je izvestilo da su čuli o takvim programima u maloj meri, dok je 10,7% uzorka izvestilo da su puno čuli ili čitali o navedenim programima. Poređenje studenata završne i studenata početne godine osnovnih studija ukazuje na to da su studenti završne godine studije u nešto većoj meri izvestili da su o takvim programima čitali u maloj meri ($\chi^2(2) = 7{,}422$, $p < {,}05$; Tabela 1).

Tabela 1. Poređenje starijih i mlađih studenata u prijavljenoj meri čitanja o programima veštačke inteligencije u kompanijama socijalnih medija

Koliko ste čuli ili čitali o računarskim programima koje kompanije socijalnih medija koriste za pronalaženje lažnih informacija na svojim sajtovima?	Mlađi studenti	Stariji studenti
Nimalo	27,4% (1,8)	17,6% (-1,8)
Malo	60,7% (-2,7)	76,5% (2,7)
Puno	11,9% (1,6)	10,7% (-1,6)

Napomena: U zagradama se nalaze korigovani standardizovani reziduali.

20,1% ispitivanog uzorka je izvestilo da misle da je široko rasprostranjena upotreba računarskih programa od strane kompanija socijalnih medija za pronalaženje lažnih informacija na njihovim sajtovima loša ideja za društvo, 46,1% smatra da je dobra ideja za društvo, dok ostalih 33,5% uzorka nije sigurno. Kada se ovi rezultati uporede sa rezultatima Rejneja i saradnika (Rainei et al. 2022), naš uzorak prijavljuje nešto pozitivnije gledište na široku rasprostranjenost upotrebe takvih računarskih programa od strane kompanija socijalnih medija. Dok 38% uzorka u istraživanju Rejneja i saradnika (Rainei et al. 2022), smatra da je takva upotreba dobra ideja za društvo, 31% smatra da je to loša ideja za društvo. Takođe,

stariji studenti nešto češće prijavljuju da smatraju da je takva upotreba dobra ideja za društvo u poređenju sa mlađim studentima ($\chi^2(2) = 14{,}518$, $p < ,01$; Tabela 2).

Tabela 2. Poređenje starijih i mlađih studenata u njihovom stavu da li bi upotreba računarskih programa za pronalaženje lažnih informacija bila dobra ili loša ideja za društvo

Da li mislite da je široko rasprostranjena upotreba računarskih programa od strane kompanija socijalnih medija za pronalaženje lažnih informacija na njihovim sajtovima …	Mlađi studenti	Stariji studenti
Dobra ideja za društvo	43,1% (-2,6)	58,8% (2,6)
Loša ideja za društvo	23,9% (3,7)	5,9% (-3,7)
Nisam siguran/na	33% (-.4)	35,3% (.4)

Napomena: U zagradama se nalaze korigovani standardizovani reziduali.

Na pitanje da li su ikada videli informaciju na sajtovima socijalnih medija koja je označena kao lažna, 64,5% ispitanika je izvestilo da jesu, dok je 26,6% ispitanika izvestilo da nisu videli takve informacije (Tabela 3). S druge strane, 74% ispitanika je u istraživanju Rejneja i saradnika (Rainei et al. 2022) izvestilo da jeste videlo oznaku za lažnu vest, dok 26% ispitanika nije videlo takvu oznaku. Razlike između starijih i mlađih studenata na ovom pitanju nisu zabeležene ($\chi^2(2) = 2{,}378$, $p > ,05$).

Tabela 3. Da li su ispitanici videli oznaku za lažnu vest

Da li ste ikada videli informaciju na sajtovima socijalnih medija da je označena kao lažna?	Srbija
Da, jesam	64,5%
Ne, nisam	26,6%
Ne koristim sajtove socijalnih medija	8,9%

S druge strane, 88,1% ispitanika je izvestilo da se pogrešno uklanjanje vesti i informacija verovatno ili definitivno dešava, dok 74,6% ispitanika smatra da isto važi i za cenzurisanje političkih gledišta. S druge strane, 83,4% ispitanika smatra da je uz korišćenje računarskih programa za otkrivanje lažnih informacija lakše nalaženje pouzdanih informacija definitivno ili verovatno slučaj, dok 68,5% ispitanika smatra da isto važi i za omogućavanje ljudima vođenja smislenijih razgovora (Tabela 4).

Tabela 4. Prikaz stavova o prednostima i manama računarskih programa za otkrivanje lažnih vesti na socijalnim medijima

Da li smatrate da korišćenje računarskih programa za otkrivanje lažnih informacija od strane kompanija socijalnih medija čine da se sledeće dešava na njihovim sajtovima…	Pogrešno uklanjanje vesti i informacija	Cenzurisanje političkih gledišta	Lakše nalaženje pouzdanih informacija	Omogućava ljudima vođenje smislenijih razgovora
Definitivno se dešava	26,1%	36,2%	31,5%	18,1%
Verovatno se dešava	62%	48,4%	51,9%	50,4%
Verovatno se ne dešava	10,7%	12,2%	13,6%	24,6%
Definitivno se ne dešava	1,2%	3,2%	3%	6,9%

Kada su upitani o tome koliku kontrolu misle da korisnici imaju nad stvarima koje vide na sajtovima socijalnih medija, ispitanici daju odgovore koji približno odgovaraju distribuciji odgovora u istraživanju sprovedenom od strane *Pew Research Center*-a (Rainei et al. 2022). Tačnije, tek 8,2% ispitanika je izvestilo da imaju puno kontrole (u poređenju sa 10% u SAD), 53,6% ispitanika je izvestilo da imaju malo kontrole (48% u SAD), dok je 27,5% ispitanika izvestilo da nemaju kontrole (33% u SAD). Rezultati hi-kvadrat analize ukazuju na to da razlike između starijih i mlađih studenata u distribuciji odgovora na ovom pitanju nisu zabeležene ($\chi^2(3) = 7{,}605$, $p > {,}05$; Tabela 5).

Tabela 5. Percepcija korisnika koliku kontrolu imaju nad prikazom sadržaja na socijalnim medijima

Koliku kontrolu mislite da korisnici imaju nad stvarima koje vide na sajtovima socijalnih medija?	Srbija
Puno kontrole	8,2%
Malo kontrole	53,6%
Nemaju kontrole	27,5%
Nisam siguran/na	10,7%

Kako bismo proverili povezanost percepcije kontrole koju korisnici imaju nad stvarima koje vide na sajtovima socijalnih medija i njihove percepcije da li je široka rasprostranjenost upotrebe računarskih programa za detekciju lažnih informacija dobra ideja za društvo, sproveli smo Hi-kvadrat test na relevantnim varijablama,

prilikom čega rezultati testa ukazuju na značajnu povezanost nivoa ovih varijabli ($\chi^2(6) = 29{,}408$, $p < {,}001$; Tabela 6). Tačnije, rezultati ukazuju na to da ispitanici koji nisu sigurni da je upotreba takvih računarskih programa dobra ideja za društvo u manjoj meri smatraju da korisnici imaju kontrolu nad sadržajem koji vide (2,2%), ali u znatno većoj meri od drugih ispitanika nisu sigurni u stepen kontrole krajnjih korisnika (20%, Tabela 6).

Tabela 6. Procenjena dobrobit upotrebe računarskih programa s obzirom na percipiranu kontrolu nad sadržajem

		Percipirana dobrobit upotrebe računarskih programa za detekciju lažnih informacija		
		Loša ideja za društvo	Nisam siguran/na	Dobra ideja za društvo
Percipirani stepen kontrole korisnika nad sadržajem na socijalnim medijima	Puno kontrole	12,3% (1,5)	2,2% (-3,1)	10,7% (1,7)
	Malo kontrole	53,1% (-.1)	51,9% (-.5)	55,1% (.6)
	Nemaju kontrole	33,3% (1,3)	25,9% (-.5)	26,2% (-.6)
	Nisam siguran	1,2% (-3,1)	20% (4,3)	8% (-1,6)

Napomena: U zagradama se nalaze korigovani standardizovani reziduali.

S druge strane, 33,7% ispitanika smatra da bi kompanije socijalnih medija trebalo da daju prioritet brzim odlukama, čak iako su neke tačne informacije uklonjene greškom, dok 66,3% ispitanika smatra da bi prioritet trebalo dati tačnim odlukama, čak iako su neke lažne informacije ostale na sajtovima tokom dužeg vremenskog perioda. Razlike u distribuciji odgovara između starijih i mlađih studenata nisu zabeležene ($\chi^2(1) = .357$, $p > {,}05$).

Kada su upitani o tome koliko poverenja imaju da će kompanije socijalnih medija koristiti računarske programe na odgovarajući način kako bi utvrdili koje informacije na njihovim sajtovima predstavljaju lažne informacije, čak 64,5% ispitanika je naglasilo da nemaju previše poverenja, 8,9% da nemaju nimalo poverenja, 22,1% da imaju prilično poverenja, dok tek 4,5% ispitanika prijavljuje da imaju veliko poverenje. Razlike u odgovorima između starijih i mlađih studenata nisu zabeležene ni na ovom pitanju ($\chi^2(3) = 5{,}36$, $p > {,}05$).

Na pitanja o ulogama vlade, kompanija socijalnih medija i krajnjih korisnika u regulisanju korišćenja računarskih programa za detekciju lažnih informacija, distribucija odgovora ispitanog uzorka nije ekstremno zakrivljena. Tačnije, 40,2%

ispitanog uzorka smatra da će vlada otići predaleko u regulisanju široke rasprostranjenosti upotrebe računarskih programa od strane kompanija socijalnih medija, dok 59,8% ispitanika smatra da vlada neće otići dovoljno daleko u takvog regulaciji. Hi-kvadrat test ($\chi^2(1) = 4{,}139$, $p < {,}05$) ukazuje na to da stariji studenti smatraju da vlada neće otići dovoljno daleko u regulaciji (69,4%) nešto češće od mlađih studenata (57,2%).

S druge strane, većina ispitanog uzorka smatra da bi kompanije socijalnih medija trebalo da imaju glavnu ulogu u postavljanju standarda za način na koji takve kompanije koriste računarske programe za pronalaženje lažnih informacija (Tabela 7). Rezultati Hi-kvardat testa ($\chi^2(2) = 8{,}467$, $p < {,}05$) ukazuju na to da stariji studenti nešto češće smatraju da bi agencije vlade trebalo da imaju manju ulogu (58,8%) u poređenju sa mlađim studentima (45,6%), mlađi studenti nešto češće prijavljuju da agencije vlade ne bi trebalo da imaju nikakvu ulogu (13,5%) u poređenju sa starijim studentima (3,5%), dok su i stariji (37,6%) i mlađi studenti (40,9%) u približnoj meri prijavili da bi agencije vlade trebalo da imaju glavnu ulogu u postavljanju adekvatnih standarda. S druge strane, razlike u distribuciji odgovora između starijih i mlađih studenata nisu registrovane kada je u pitanju uloga kompanija socijalnih medija ($\chi^2(2) = 2{,}563$, $p > {,}05$) i uloga kranjih korisnika socijalnih medija ($\chi^2(2) = 2{,}529$, $p > {,}05$).

Tabela 7. Percepcija korisnika o važnosti grupa za postavljanje standarda za kompanije socijalnih medija

Koliku ulogu svaka od sledećih grupa bi trebalo da ima u postavljanju standarda za način na koji kompanije socijalnih medija koriste računarske programe za pronalaženje lažne informacije na svojim sajtovima?	Agencija vlade	Kompanije socijalnih medija	Krajnji korisnici
Glavnu ulogu	40,2%	57,3%	30,3%
Manju ulogu	48,4%	33,7%	50,1%
Nikakvu ulogu	11,4%	8,9%	19,6%

Dok 47,9% ispitanog uzorka smatra da bi odluka o tome koja informacija detektovana kao lažna trebalo da dođe i od ljudi i računarskih programa, 30% ispitanika smatra da bi takva odluka trebalo da bude pretežno donesena od ljudi, 10,2% da bi trebalo da bude donesena pretežno od računarskog programa, a 11,9% ispitanog uzorka nije sigurno u svoj odgovor. Rezultati Hi-kvadrat testa ukazuju na to da se stariji i mlađi studenti ne razlikuju značajno u svojim odgovorima na ovom

pitanju ($\chi^2(3) = 4{,}865$, $p > {,}05$). S druge strane, kada su upitani o tome da li računarski programi za detekciju lažnih informacija taj posao obavljaju bolje od ljudi, tek 19,6% ispitanika je odgovorilo potvrdno, 25,5% je odgovorilo negativno, 15,1% ispitanika smatra da računarski programi obavljaju isti posao kao i ljudi, dok 39,7% ispitanika nije sigurno u svoj odgovor. Razlike između starijih i mlađih studenata nisu registrovane ni na ovom pitanju ($\chi^2(3) = 2{,}922$, $p > {,}05$).

Kako bismo proverili povezanost frekventnijeg čitanja o računarskim programima za detekciju lažnih informacija i mišljenja o tome da li takvi programi obavljaju navedeni posao bolje od ljudi, sproveden je Hi-kvadrat test na relevantnim pitanjima. Rezultati Hi-kvadrat testa ukazuju na to da ispitanici u različitoj meri procenjuju da li računarski programi bolje detektuju lažne informacije od ljudi u zavisnosti od frekventnosti čitanja o takvim programima ($\chi^2(6) = 15{,}401$, $p < {,}05$; Tabela 8).

Tabela 8. Interesovanje za računarske programe s obzirom na percepciju njihove uspešnosti u detekciji lažnih vesti

		Koliko ste čuli ili čitali o računarskim programima za pronalaženje lažnih informacija na sajtovima socijalnih medija		
		Nimalo	Malo	Puno
Prilikom nalaženja lažnih informacija na njihovim sajtovima, da li smatrate da računarski programi koje koriste kompanije socijalnih medija rade:	Bolji posao od ljudi	15,7% (-1,2)	21,3% (1,2)	18,6% (-.2)
	Lošiji posao od ljudi	22,5% (-.8)	23,6% (-1,2)	44,2% (3)
	Isti posao kao i ljudi	11,8% (-1,1)	16,3% (.9)	16,3% (.2)
	Nisam siguran/na	50% (2,5)	38,8% (-.5)	20,9% (-2,7)

Napomena: U zagradama se nalaze korigovani standardizovani reziduali.

Diskusija

Cilj sprovedenog istraživanja je bio dvojak. S jedne strane, ispitani su stavovi studenata prema algoritmima veštačke inteligencije koji se koriste u socijalnim medijima u generalne svrhe i u svrhe detekcije lažnih vesti. S druge strane, ispitane su razlike između starijih i mlađih studenata u njihovim odgovorima.

Rezultati istraživanja ukazuju na to da je većina studenata malo čula ili čitala o računarskim programima koji se koriste u socijalnim medijima, prilikom čega stariji studenti tvrde da su više čitali o takvim programima od mlađih studenata. Takođe, stariji studenti u većoj meri smatraju da je korišćenje takvih računarskih programa dobra ideja za društvo, dok mlađi ispitanici u većoj meri smatraju da su takvi programi loša ideja za društvo. Ovi nalazi ukazuju na veću otvorenost starijih studenata ka navedenim računarskim programima, što može biti posledica razvijenijeg znanja o takvim programima. Buduća istraživanja bi mogla eksplicitno da testiraju vezu između teorijskog i praktičnog znanja o algoritmima veštačke inteligencije i otvorenosti ka korišćenju takvih algoritama u različitim aplikacijama i softverskim rešenjima. Takođe, navedeni rezultati ukazuju i na značajnost obrade ovakvih tema u okviru relevantnih kurseva na univerzitetima. Jedan takav primer predstavlja kurs na Univerzitetu Indijana koji je kreiran 2022. godine, a koji nosi naslov „Manipulacija na socijalnim medijima 101" (Indiana, 2022).

Međutim, studenti su takođe svesni i prednosti i mana korišćenja takvih algoritama. Tačnije, velika većina ispitanih studenata smatra da dolazi do pogrešnog uklanjanja vesti i informacija i cenzurisanja političkih gledišta. S druge strane, većina studenata istovremeno smatra da korišćenje takvih računarskih programa omogućava lakše nalaženje pouzdanih informacija, te i da takvi programi olakšavaju komunikaciju među ljudima. Ovakav nalaz ukazuje na odsustvo ekstremnog polarizovanja stavova na ispitanom uzorku, te da studenti percipiraju i pozitivne i negativne strane korišćenja takvih računarskih programa. Ovakav nalaz upućuje na to da, iako krajnji korisnici socijalnih medija nemaju pristup performansama modela veštačke inteligencije koji su razvijeni s ciljem detekcije lažnih vesti, oni su svesni postojanja mogućnosti cenzure i pogrešnih detekcija. Performanse modela bi stoga bilo korisno podeliti sa krajnjim korisnicima, naročito uzevši u obzir da se krajnji korisnici najčešće i informišu putem socijalnih medija u vanrednim situacijama (Santoni & Rufat, 2021). Važnost tačnog detektovanja lažnih vesti je naglašena i većinskim stavom ispitanika da bi takvi algoritmi trebalo da daju prednost tačnim informacijama nasuprot brzom donošenju odluka.

Rezultati ovog istraživanja repliciraju obrasce prethodnih istraživanja kada se radi o poverenju koje ispitanici imaju u to da se algoritmi za detekciju lažnih vesti koriste na odgovarajuće načine. Na primer, Flečer i Nilsen (Fletcher & Nielsen, 2019) su zabeležili da ljudi primaju vesti na socijalnim medijima sa "generalizovanim skepticizmom", odnosno da ljudi imaju nisko poverenje u način na koji su vesti selektovane. Navedeni obrazac su uočili u čak četiri zemlje - Velika Britanija, SAD, Nemačka i Španija. Pored toga, autori su zabeležili i da mlađi ispitanici imaju veće poverenje u način na koji algoritam bira vesti, što je rezultat koji bi trebalo proveriti u budućim istraživanjima.

Rezultati ovog istraživanja ukazuju i na to da studenti opažaju potrebu da se u detekciji lažnih vesti koriste i algoritmi veštačke inteligencije i ljudski kapaciteti. Takva procedura bi, na primer, mogla da uključi klasifikaciju vesti na one koje su lažne i one koje to nisu od strane algoritma, a potom bi stručnjaci iz oblasti potvrdili ili opovrgli takvu klasifikaciju (Somer 2018). Takvo simultano korišćenje i ljudskih kapaciteta i algoritama za detekciju lažnih vesti neki autori označavaju kao najbolji pristup u detekciji lažnih vesti (Cohen 2020). Međutim, takvo kombinovanje se može koristiti u trenutku obučavanja modela za detekciju lažnih vesti, dok nije jasno kako bi se ljudski potencijali mogli iskoristiti za detekciju lažnih vesti u produkciji usled ogromnog broja generisanih informacija. Jedan od načina može predstavljati uključenje krajnjih korisnika u potvrđivanje ili opovrgavanje odluke algoritma, ali bi takva povratna informacija bila nepouzdana. Takođe, Li i Fang (Lee & Fung 2021) ističu da bi se algoritmi za detekciju lažnih vesti mogli zloupotrebiti, te da mogu imati negativnih implikacija po slobodu govora. Stoga je neophodno uključiti i neutralnu treću stranu koja bi vršila nadzor nad korišćenjem takvih algoritama. Na primer, poslanici Evropskog parlamenta su u visoko-rizične aplikacije uvrstili sisteme veštačke inteligencije koji mogu naneti štetu ljudskom zdravlju, sigurnosti, osnovnim pravima i okruženju (MEP, 2023). Pored toga, visoko-rizičnim sistemima dodati su sistemi veštačke inteligencije koji utiču na glasače na izborima i u platformama socijalnih medija u sistemima preporuke (MEP, 2023). Rezultati ovog istraživanja potvrđuju navedeno zapažanje - većina studenata smatra da bi ulogu u postavljanju standarda za korišćenje takvih algoritama trebalo da imaju ili agencije vlade ili kompanije koje koriste takve algoritme, dok najmanje poverenja daju upravo krajnjim korisnicima takvih aplikacija.

Na kraju, ovo istraživanje je imalo nekoliko poteškoća i nedostataka koje bi trebalo prevazići u budućim istraživanjima. Iako je prigodan uzorak studenata Fakulteta bezbednosti odabran sa namerom, ovakvi rezultati se ne mogu generalizovati na populaciju Srbije. Kako bi se na adekvatniji način uticalo na javne politike neophodno je sprovesti ovakvo ili slično istraživanje na reprezentativnom uzorku, poput istraživanja Rejneja i saradnika (Rainie et al., 2022). S druge strane, iako su stavke koje su korišćene u ovom istraživanju prevedene iz istraživanja sprovedenog od strane Rejneja i saradnika (Rainie et al., 2022), tip stavki ne omogućava sprovođenje sofisticiranijih analiza koje bi utvrdile latentne faktore u osnovi stavova prema korišćenju algoritama za detekciju lažnih vesti ili potencijalno klasterizovanje ispitanika s obzirom na njihove odgovore. Stoga bi se buduća istraživanja mogla fokusirati na prilagođavanje stavki na intervalni tip radi davanja odgovora na složenije istraživačke hipoteze.

Zaključak

Opšti princip je da je rezultat determinisan ulaznim podacima, a politički procesi u svojoj složenosti i osetljivosti tu nisu izuzetak. Odnosno, pogrešne ili lažne informacije za rezultat mogu da imaju izuzetno loše, pa i tragične posledice po pojedinca i društvo, poput podsticanja nasilja, uticaja na izbore, a sekundarno i gubitak poverenja u prave vesti. Zbog obimne produkcije informacija postaje nužno automatizovati proces njihove selekcije. Kompanije koje stoje iza platformi socijalnih medija koriste algoritme veštačke inteligencije kako bi ostvarile različite funkcije na svojim mrežama. Primenom algoritama korisnici dobijaju sadržaj prilagođen njihovim preferencama i stavovima, i istovremeno otkrivaju i sprečavaju širenje lažnih vesti. U okviru istraživanja predstavljeni su rezultati analize sprovedene među studentima koji se smatraju tehnološki najprogresivnijim delom društva i budućim donosiocima odluka.

Ovo istraživanje sugeriše da većina studenata ima ograničeno poznavanje algoritama, ali da većina studenata smatra da je upotreba takvih algoritama dobra ideja za društvo. Ovaj rezultat je naročito izražen kod starijih studenata, što je potencijalno rezultat većeg znanja o ovim programima. Ispitanici su u stanju i da prepoznaju potencijalne prednosti i nedostatke upotrebe ovih algoritama, poput pogrešnog uklanjanja informacija i cenzure.

Nalazi ovog istraživanja ukazuju na potrebu razvoja ili daljeg produbljivanja univerzitetskih kurseva koji se bave poljem veštačke inteligencije i njenom regulativom, ali i da transparentnost performansi modela veštačke inteligencije može biti od koristi krajnjim korisnicima. Takođe, rezultati ukazuju na nužnost kombinovanja algoritama veštačke inteligencije i ljudske ekspertize u detekciji lažnih vesti i postavljanje adekvatnih standarda.

U budućim istraživanjima mogla bi se detaljnije ispitati povezanost između teorijskog i praktičnog razumevanja algoritama veštačke inteligencije, kao i sklonosti prema primeni ovakvih algoritama u različitim aplikativnim i softverskim okvirima. Uprkos izvesenim ograničenjima ovog istraživanja, treba istaći da ono predstavlja prvu analizu stavova u Srbiji prema programima veštačke inteligencije korišćenim na društvenim mrežama radi prikazivanja sadržaja i suzbijanja lažnih vesti.

Bibliografija

Allcott, Hunt, and Matthew Gentzkow. 2017. ‘Social Media and Fake News in the 2016 Election’. *Journal of Economic Perspectives* 31 (2): 211–36.

Asch. 1951. 'Effects of Group Pressure on the Modification and Distortion of Judgments.' In *Groups, Leadership and Men, Pittsburgh*, 117–90. Pittsburgh, PA: Carnegie Press.

Balmas, Meital. 2012. 'When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation and Cynicism'. *Communication Research* 41 (July). https://doi.org/10.1177/0093650212453600.

Belli, Luca. n.d. 'Examining Algorithmic Amplification of Political Content on Twitter'. Accessed 20 August 2023. https://blog.twitter.com/en_us/topics/company/2021/rml-politicalcontent.

Bovet, Alexandre, and Hernán A. Makse. 2019. 'Influence of Fake News in Twitter during the 2016 US Presidential Election'. *Nature Communications* 10 (1): 7. https://doi.org/10.1038/s41467-018-07761-2.

Brandom, Russell. 2020. 'Facebook's Misinformation Problem Goes Deeper than You Think'. The Verge. 17 March 2020. https://www.theverge.com/2020/3/17/21183341/facebook-misinformation-report-nathalie-marechal.

Buchanan, Ben, Andrew Lohn, Micah Musser, and Katrina Sedova. 2021. 'Truth, Lies, and Automation'. *Center for Security and Emerging Technology* (blog). 2021. https://cset.georgetown.edu/publication/truth-lies-and-automation/.

Buçinca, Zana, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. 'To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making'. *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1): 188:1-188:21. https://doi.org/10.1145/3449287.

Burke, Garance, Martha Mendoza, Juliet Linderman, and Michael Tarm. 2022. 'How AI-Powered Tech Landed Man in Jail with Scant Evidence'. AP News. 5 March 2022. https://apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220.

Cohen, Ira. 2020. 'Can AI Analytics Stop Fake News?' Anodot. 26 January 2020. https://www.anodot.com/blog/ai-analytics-stops-fake-news/.

Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012. 'Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?' *IEEE Transactions on Dependable and Secure Computing* 9 (6): 811–24.

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2016. 'The Spreading of Misinformation Online'. *Proceedings of the National Academy of Sciences* 113 (3): 554–59. https://doi.org/10.1073/pnas.1517441113.

Dizikes, Peter. 2018. 'Study: On Twitter, False News Travels Faster than True Stories'. *MIT News* 8: 2018.

Dwoskin, Elizabeth, Nitasha Tiku, and Craig Timber. 2021. 'Facebook's Race-Blind Practices around Hate Speech Came at the Expense of Black Users, New Documents Show'. Washington Post. 21 November 2021. https://www.washingtonpost.com/technology/2021/11/21/facebook-algorithm-biased-race/.

Fergusson, Grant et al. 2023. 'Generating Harms Generative AI's Impact & Paths Forward'. Epic. https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf.

Fletcher, Richard, and Rasmus Kleis Nielsen. 2019. 'Generalised Scepticism: How People Navigate News on Social Media'. *Information, Communication & Society* 22 (12): 1751–69. https://doi.org/10.1080/1369118X.2018.1450887.

Fong, Joss. 2021. 'Are We Automating Racism?' Vox. 31 March 2021. https://www.vox.com/videos/2021/3/31/22348722/ai-bias-racial-machine-learning.

Sze-Fung Lee and Fung, Benjamin C. M. 2021. 'Artificial Intelligence May Not Actually Be the Solution for Stopping the Spread of Fake News'. The Conversation. 28 November 2021. http://theconversation.com/artificial-intelligence-may-not-actually-be-the-solution-for-stopping-the-spread-of-fake-news-172001.

Guess, Andrew M., Pablo Barberá, Simon Munzert, and JungHwan Yang. 2021. 'The Consequences of Online Partisan Media'. *Proceedings of the National Academy of Sciences* 118 (14): e2013464118. https://doi.org/10.1073/pnas.2013464118.

Howell, Lee. 2013. 'Digital Wildfires in a Hyperconnected World'. World Economic Forum. https://www.weforum.org/reports/world-economic-forum-global-risks-2013-eighth-edition/.

Hughes, Heather C., and Israel Waismel-Manor. 2021. 'The Macedonian Fake News Industry and the 2016 US Election'. *PS: Political Science & Politics* 54 (1): 19–23. https://doi.org/10.1017/S1049096520000992.

Indiana. n.d. 'Click Here for an Easy A: Social Media Manipulation 101'. Accessed 20 August 2023. https://osome.iu.edu/education/social-media-manipulation-101.

Jin, Zhiwei, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. 'News Verification by Exploiting Conflicting Social Viewpoints in Microblogs'. *Proceedings of the AAAI Conference on Artificial Intelligence* 30 (1). https://doi.org/10.1609/aaai.v30i1.10382.

Kemp, Simon. 2023a. 'Digital 2023 April Global Statshot Report'. DataReportal – Global Digital Insights. 27 April 2023. https://datareportal.com/reports/digital-2023-april-global-statshot.

Kemp, Simon. 2023b. 'Digital 2023 July Global Statshot Report'. DataReportal – Global Digital Insights. 20 July 2023. https://datareportal.com/reports/digital-2023-july-global-statshot.

Kenyon, Tilly. 2021. 'How Are Social Media Platforms Using AI?' 26 June 2021. https://aimagazine.com/ai-strategy/how-are-social-media-platforms-using-ai.

Kovačević, Ana. 2023. 'Mašinsko Učenje i Sajber Bezbednost'. In *Zbornik Radova Sa Konferencije Strateški i Normativni Okvir Republike Srbije Za Reagovanje Na Savremene Bezbednosne Rizike*. Beograd: Univerzitet u Beogradu - Fakultet bezbednosti.

Kovačević, Ana, Nenad Putnik, and Oliver Tošković. 2020. 'Factors Related to Cyber Security Behavior'. *IEEE Access* 8: 125140–48.

Kramer, Melody. 2017. 'Do Facebook and Google Have Control of Their Algorithms Anymore? A Sobering Assessment and a Warning'. *Poynter* (blog). 14 November 2017. https://www.poynter.org/business-work/2017/do-facebook-and-google-have-control-of-their-algorithms-anymore-a-sobering-assessment-and-a-warning/.

Lohr, Steve. 2021. 'Group Backed by Top Companies Moves to Combat A.I. Bias in Hiring'. *The New York Times*, 8 December 2021, sec. Technology. https://www.nytimes.com/2021/12/08/technology/data-trust-alliance-ai-hiring-bias.html.

Lynch, Michael P. 2016. 'Opinion | Fake News and the Internet Shell Game'. *The New York Times*, 28 November 2016, sec. Opinion. https://www.nytimes.com/2016/11/28/opinion/fake-news-and-the-internet-shell-game.html.

Marcin, Cesluk-Grajewski. 2023. 'Artificial Intelligence [What Think Tanks Are Thinking]', June.https://policycommons.net/artifacts/4315682/artificial-intelligence-what-think-tanks-are-thinking/5125253/.

Menczer, Filippo, and Thomas Hills. 2020. 'Information Overload Helps Fake News Spread, and Social Media Knows It'. *Scientific American* 323 (6): 54–61.

MEP 'MEPs Ready to Negotiate First-Ever Rules for Safe and Transparent AI | News | European Parliament'. n.d. Accessed 21 August 2023. https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai.

Merrill, Craig Silverman, Craig Timberg,Jeff Kao,Jeremy B. 2022. 'Facebook Hosted Surge of Misinformation and Insurrection Threats in Months Leading Up to Jan.

6 Attack, Records Show'. ProPublica. 4 January 2022. https://www.propublica.org/article/facebook-hosted-surge-of-misinformation-and-insurrection-threats-in-months-leading-up-to-jan-6-attack-records-show.

Minsky, Marvin. 1968. *Semantic Information Processing-The MIT Press*. The MIT Press.

Nickerson, Raymond S. 1998. 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises'. *Review of General Psychology* 2 (2): 175–220.

Nyhan, Brendan, and Jason Reifler. 2010. 'When Corrections Fail: The Persistence of Political Misperceptions'. *Political Behavior* 32 (2): 303–30.

Oremus, Will, Chris Alcantara, Jeremy B. Merrill, and Artur Galocha. n.d. 'How Facebook Shapes Your Feed'. Washington Post. Accessed 10 August 2023. https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works/.

Paul, Christopher, and Miriam Matthews. 2016. 'The Russian "Firehose of Falsehood" Propaganda Model'. *Rand Corporation* 2 (7): 1–10.

Peters, Jay. 2020. 'Facebook Was Marking Legitimate News Articles about the Coronavirus as Spam Due to a Software Bug'. The Verge. 17 March 2020. https://www.theverge.com/2020/3/17/21184445/facebook-marking-coronavirus-posts-spam-misinformation-covid-19.

Petković, Mateja. 2022. 'Širenje Lažnih Vesti Na Društvenim Mrežama'. Diplomski rad, Beograd: Univerzitet u Beogradu, Fakultet bezbednosti.

Preston, Ashlee Marie. n.d. 'Taking On Tech: Dr. Timnit Gebru Exposes The Underbelly Of Performative Diversity In The Tech Industry'. Forbes. Accessed 10 August 2023. https://www.forbes.com/sites/forbestheculture/2021/07/30/taking-on-tech-dr-timnit-gebru-exposes-the-underbelly-of-performative-diversity-in-the-tech-industry/.

Rainie, Lee, Cary Funk, Monica Anderson, and Alec Tyson. 2022. 'AI and Human Enhancement: Americans' Openness Is Tempered by a Range of Concerns'. *Pew Research Center: Internet, Science & Tech* (blog). 17 March 2022. https://www.pewresearch.org/internet/2022/03/17/ai-and-human-enhancement-americans-openness-is-tempered-by-a-range-of-concerns/.

Rubin, Victoria L., Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. 'Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News'. In *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, 7–17.

Santoni, Victor, and Samuel Rufat. 2021. 'How Fast Is Fast Enough? Twitter Usability during Emergencies'. *Geoforum* 124: 20–35.

Schuett, Jonas. 2023. 'Defining the Scope of AI Regulations'. *Law, Innovation and Technology* 15 (1): 60–82.

.Shearer, Elisa, and Amy Mitchell. 2021. 'News Use Across Social Media Platforms in 2020'. *Pew Research Center's Journalism Project* (blog). 12 January 2021. https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/.

Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. 'Fake News Detection on Social Media: A Data Mining Perspective'. *ACM SIGKDD Explorations Newsletter* 19 (1): 22–36.

Silverman, Craig. 2016. 'This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook'. BuzzFeed News. 16 November 2016. https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.

Singer, Peter Warren, and Emerson T. Brooking. 2018. *LikeWar: The Weaponization of Social Media*. Eamon Dolan Books.

Solaiman, Irene, Zeerak Talat, William Agnew, Lama Ahmad, Dylan Baker, Su Lin Blodgett, Hal Daumé III, et al. 2023. 'Evaluating the Social Impact of Generative AI Systems in Systems and Society'. arXiv. https://doi.org/10.48550/arXiv.2306.05949.

Somer, Iryna. 2018. 'Lithuanians Create Artificial Intelligence with Ability to Identify Fake News in 2 Minutes'. *Kyiv Post*.

Sunstein, Cass. 2007. 'Republic. Com 2.0. Princeton University Press'.

Statt, Nick. 2020. 'Major Tech Platforms Say They're "Jointly Combating Fraud and Misinformation" about COVID-19'. The Verge. 17 March 2020. https://www.theverge.com/2020/3/16/21182726/coronavirus-covid-19-facebook-google-twitter-youtube-joint-effort-misinformation-fraud.

Tandoc, Edson C., Zheng Wei Lim, and Richard Ling. 2018. 'Defining "Fake News"'. *Digital Journalism* 6 (2): 137–53. https://doi.org/10.1080/21670811.2017.1360143.

Vakil, Caroline. 2021. 'More Hate Speech, Misinformation Possible If Algorithms Are Removed, Facebook VP Says'. Text. *The Hill* (blog). 10 October 2021. https://thehill.com/homenews/sunday-talk-shows/576118-more-hate-speech-misinformation-possible-if-algorithm-is-removed/.

Verma, Pranshu, and Will Oremus. 2023. 'ChatGPT Invented a Sexual Harassment Scandal and Named a Real Law Prof as the Accused'. *Washington Post*, 14 April 2023. https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. 'The Spread of True and False News Online'. *Science* 359 (6380): 1146–51.

Ward, Andrew, L. Ross, E. Reed, E. Turiel, and T. Brown. 1997. 'Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding'. *Values and Knowledge*, 103–35.

Ковачевић, Миладин, Владимир Шутић, Урош Рајчевић, and Ивана Минаева. 2022. 'Употреба Информационо-Комуникационих Технологија у Републици Србији, 2022.' Републички завод за статистику, Београд. https://www.stat.gov.rs/publikacije/publication/?p=14856.

Ana KOVAČEVIĆ, Emir DEMIĆ

ARTIFICIAL INTELLIGENCE FOR DETECTING FAKE NEWS ON SOCIAL MEDIA – ATTITUDE SURVEY

Abstract: In the modern digital era, social media are playing an important role in in the dissemination of information. Nevertheless, phenomena such as fake news have the potential to considerably compromise user trust. Concurrently, artificial intelligence unveils remarkable opportunities while also introducing great challenges. Artificial intelligence is already employed in identifying fake news on social media platforms, rendering this a crucial area of exploration due to its profound implications on society and politics. This paper examines the application of artificial intelligence on social media, with a particular focus on detecting fake news, through an analysis of the attitudes of the student population. The findings suggest a limited awareness among students about these kinds of programs, highlighting a necessity for the inclusion of these topics in future university curricula. A majority of students believe that artificial intelligence algorithms for detecting fake news are good for society, compared to those who believe the contrary. The results indicate a prevailing concern regarding the potential misuse of these technologies and their impact on freedom of speech, thereby emphasizing the need for impartial oversight and regulation. The research has demonstrated that students are aware of the advantages and limitations of applying artificial intelligence on social media, with older students exhibiting greater receptiveness towards these programs.

Keywords: risks, machine learning, statistics, questionnaire, algorithms, Serbia.