

Etički dizajn sistema preporuka

Seminarski rad u okviru kursa
Metodologija stručnog i naučnog rada
Matematički fakultet

Andela Bašić, Jelena Bondžić, Ognjen Popović, Petar Nikić

mi231018@alas.matf.bg.ac.rs, mi231028@alas.matf.bg.ac.rs,

mi231027@alas.matf.bg.ac.rs, mi231012@alas.matf.bg.ac.rs

22. novembar 2023.

Sažetak

Sistemi preporuka su algoritmi koji generišu personalizovani sadržaj korisnicima, u svrhu efikasnijeg korišćenja servisa na internetu. Uz više podataka i implementaciju raznih tehnika ubeđivanja, oni postaju uspešniji u tom procesu. Međutim, postoji sukob interesa u vidu saržaja koji prikazuju, jer preporuke treba da budu korisne korisnicima, ali i proizvođačima. Razmatrajući njihov uticaj na pojedinca i društvo u kontekstu korišćenja tehnika ubeđivanja, ispravnosti informacija, i uticaja na demokratiju i diskriminaciju u društvu, predočavamo određene etičke zamke. Najzad, izdvajamo dizajnerske principe etičkih sistemima preporuka.

Sadržaj

1	Uvod	2
2	Sistemi preporuka	2
2.1	Koncepti sistema preporuka	3
2.2	Pristupi u sistemima preporuka	3
3	Ubeđivanje u sistemima preporuka	4
3.1	Sistemi preporuka i digitalno gurkanje	4
3.2	Primer Vitable	5
4	Posledice masovne izloženosti sistemima preporuka	6
4.1	Posledice po socijalne veze	6
4.2	Širenje dezinformacija	7
4.3	Posledice po demokratsko funkcionisanje	7
5	Diskriminacija u sistemima preporuka	8
5.1	Primeri diskriminacije	8
5.2	Da li algoritam namerno vrši diskriminacije?	9
6	Etički osvrt	9
6.1	Dizajn etičkog sistema preporuka	10
7	Zaključak	11
	Literatura	12

1 Uvod

Amazon nudi kupcima čak 353 miliona proizvoda [1], a Google pretraživač indeksira i pretražuje stotine milijardi veb strana [2]. Postoji potreba da se korisnicima efikasno pruži pregled sadržaja sa kojim rade, a posebno sadržaja koji su interesantni svakom korisniku. Kompanije rade intenzivno na razvijanju novih, boljih tehnika pretrage i preporučivanja sadržaja kako bi poboljšali iskustvo korisnika i smanjili troškove.

Pravljenje dobrih preporuka je direktno u vezi sa nekoliko biznis modela. Kompanije kao što su Google i Meta, koje deo prihoda ostvaruju na osnovu reklama, pružaju mnogo funkcionalnosti i izuzetnu uslugu da bi korisnici duže koristili njihove servise i samim tim bili izloženi većem broju reklama. Dodatno, praćenjem i analizom podataka o korišćenju servisa, dobijaju uvid u interesovanja korisnika, što koriste ne samo za prilagođavanje sadržaja, već i reklama. Na primer, Amazon pomaže kupcima da lakše pronađu proizvode, čime poboljšava uslugu, ali i predlaže dodatne proizvode, čime povećava zaradu.

Pored dobrih predikcija, na prihvatanje preporuka utiče i način komunikacije sistema sa korisnicima [5]. Korisnici percipiraju sistem kao sagovornika [48], pa se neki socijalni obrasci iz dinamike ubeđivanja u međuljudskim odnosima mogu se primeniti na relaciju sistema preporuka i čoveka. Dakle, sistem treba prvo da odredi preporuku, ali i da je korisniku prikaže na ubedljiv način, kako bi ona bila prihvaćena.

Na primer, korisnici mogu da posmatraju kompjuter u kontekstu pola i etničke pripadnosti i da shodno tome formiraju predrasude [33]. Naime, ekperiment iz 1997. testira mišljenja korisnika o sistemu koji prikazuje informacije muškim i ženskim glasom. Sistem “muškog pola” je ocenjen kao stručniji od sistema “ženskog pola” u oblasti tehnologije. Međutim, poredak je obrnut u oblasti ljubavi i odnosa. Takođe, sistem “muškog pola” je viđen kao saradljiviji i informativniji.

Uticao je uticaj ubedljivog sistema postoji na individualnom, ali i globalnom nivou. Internet koristi 5.18 milijardi ljudi (64.6% populacije) u proseku 6.5 sati dnevno, primarno radi pronalaženje informacija. U zavisnosti od starosne grupe, 30%-37% korisnika koristi društvene mreže najpre za čitanje vesti [25]. Stoga, internet i društvene mreže, preko svojih sistemima preporuka, utiču na mentalno zdravlje ljudi, ispravnost informacija, politiku i sistemsku agresiju [14]. Zbog uticaja koji sistemi preporuka imaju na pojedinca i društvo, razmatraju se načini funkcionisanja koji su u skladu sa etičkim načelima. Na primer, ubedljiv sistem preporuka lakše prelazi granicu između manipulacije i preporučivanja, jer postoji sukob interesa između dobre preporuke i profita [48]. Takođe, ako podaci nad kojima su algoritmi preporuke trenirani nisu reprezentativni, oni mogu doprineti diskriminaciji u društvu.

Etički sistem preporuka treba da zaobiđe negativni uticaj na društvo kroz bolje razumevanje potencijalnih problema i odgovorniji proces implementacije.

2 Sistemi preporuka

Razvoj veba i povećanje količine podataka za obradu direktno je motivisalo razvoj više povezanih naučnih disciplina koje obrađuju slične probleme, pri čemu su dve koje se izdvajaju **pronalaženje informacija** (eng. *information retrieval*) i **sistemi preporuka** (eng. *recommender*

systems, recommendation systems). U oba slučaja, cilj je pomoći korisniku da dobije željene rezultate.

Sistemi preporuka mogu se opisati kao kompleksni softverski sistemi čija je glavna uloga da svojim **korisnicima** (eng. *users*) daju personalizovan pregled **predmeta** (eng. *items*) registrovanih u sistemu za koje sistem proceni da im mogu biti interesantni.

Sistemi preporuka su opšteprisutni. Netflix ih koristi za preporuke filmova i serija, Amazon da bi predložio proizvode koje možete kupiti, Facebook, Twitter i LinkedIn da vam predlože interesantne objave sa kojima možda nemate direktnu vezu ili ljude sa kojima možete da se povežete, Google prilagođava preporuke za pretragu i same rezultate pretrage u zavisnosti od korisnika, a YouTube, Instagram i TikTok prikazuju veliki broj slika i video sadržaja, pri čemu su neki organizovani u neprekidne personalizovane nizove koje generiše sistem.

Tehnike pronalaženja informacija su izrazito prisutne u implementaciji raznih vrsta pretraživača, pri čemu korisnik parametrima pretrage opisuje željene rezultate. U slučaju modernih sistema preporuka, nije neophodno da korisnik pravi upite, već sistem može da pretpostavi šta je relevantno korisniku, i pruži mu personalizovan rezultat. Presek ove dve oblasti su moderni pretraživači, koji integrišu tehnike sistema preporuka za pružanje personalizovanih pretraga, gde na rezultate upita korisnika utiču ne samo podaci navedeni u upitu, već i kontekst koji čine interesovanja ili lokacija korisnika [3].

2.1 Koncepti sistema preporuka

U osnovi starih sistema preporuka stoji sama ideja **preporuke** koja nam je potrebna da donesemo odluku kada nemamo dovoljno znanja o svim opcijama. Oni su zamišljeni kao prostor gde bi korisnici delili svoje utiske, a sistem bi ih sakupljao, agregirao i učinio dostupnim ostalim korisnicima [41]. Kasniji sistemi preporuka pomeraju fokus sa koncepta preporuka korisnika na sam preporučeni predmet. Koncept preporuke se apstrahuje u meru koja se zove upotrebljivost (eng. *utility*) ili ocena (eng. *rating*) [8].

Unapređivanjem korisničkih interfejsa, korisnicima postaje lakše da ocenjuju sadržaj [3]. Ipak, analiziranje isključivo tih ocena nije uvek reprezentativno. Jedan uzrok je pristrasnost uzorka, jer pravi skup korisnika i skup ljudi koji ostavljaju ocenu ne moraju da se poklapaju [36], bilo zato što korisnici ne žele da daju ocene, ili zbog lažnih ocena. Zbog toga sistemi preporuka takođe počinju da prate, sakupljaju i analiziraju podatke o interakcijama korisnika sa različitim predmetima. Obradeni podaci i eksplicitne ocene korisnika sistem koristi u formiranju procene upotrebljivosti predmeta za tog korisnika.

2.2 Pristupi u sistemima preporuka

Postoje obrasci u ocenama predmeta, kako između različitih predmeta, tako i između različitih ljudi [3]. Dva glavna pristupa koja koriste ove obrasce i koja se koriste danas u sistemima preporuka su **filtriranje po sadržaju** (eng. *content-based filtering*) i **zajedničko filtriranje** (eng. *collaborative filtering*).

Tehnike zasnovane na sadržaju obuhvataju određivanje i analiziranje raznih karakteristika sadržaja predmeta koje je neki korisnik prethodno

razmatrao i kako te karakteristike utiču na ocenu korisnika. Dobijeno znanje se onda koristi u preporuci novih predmeta. Na primer, ukoliko neko često sluša određeni žanr muzike, veća je verovatnoća da će mu se svideti neka nova pesma upravo iz tog žanra.

Tehnike zajedničkog filtriranja koriste informacije o aktivnosti više korisnika u određivanju preporuka. Primer jednostavnog pristupa zajedničkog filtriranja je zasnovan na susedstvima, odnosno traženju sličnosti u ocenama između više korisnika ili više predmeta. Ukoliko dve osobe, na primer, imaju slične utiske za nekoliko predmeta, pretpostavlja se da će imati slične utiske i za neki novi predmet. Alternativno, mogu se koristiti razne tehnike istraživanja podataka ili mašinskog učenja u definisanju modela na osnovu koga će se vršiti preporuke [3].

3 Ubeđivanje u sistemima preporuka

Ubeđivanje je forma komunikacije u svrhu uticaja na rasuđivanje i ponašanje sagovornika [43]. Po definiciji, ono je različito od manipulacije, jer se kroz proces usmerenog predlaganja, a ne prinude, postiže promena stavova sagovornika u cilju njegovog i opšteg dobra.

Ukoliko softverski sistem ubeđuje, a ne čovek, radi se o tehnologiji ubeđivanja. Sistemi preporuka su namenjeni da preporučuje, a ne ubede korisnika [4]. Ipak, ukoliko preporuka menja stavove i ponašanja, ona je ubedljiva [48].

Određeni socijalni obrasci iz dinamike ubeđivanja u međuljudskim odnosima mogu se primeniti na relaciju sistema preporuka i čoveka [48]. Takođe, njihov uticaj je izraženiji ukoliko sistem preporuka komunicira sa korisnikom preko virtuelnog agenta.

Studija iz 2006. [20] posmatra uticaj virtuelnog agenta na onlajn kupovinu. Dodavanje virtuelnog agenta korisničkom interfejsu poboljšalo je potrošački stav kupaca. Takođe, slično međuljudskoj interakciji, agenti boljeg fizičkog izgleda imaju bolje ubeđivačke sposobnosti nad korisnicima koji nisu uključeni u kupovinu na visokom nivou. Međutim, u slučaju visokouključenih korisnika, stručni agenti imaju bolje ubeđivačke sposobnosti.

3.1 Sistemi preporuka i digitalno gurkanje

Gurkanje (eng. nudge) predstavlja mehanizme koji se, oslanjajući se na određene uvide iz psihologije donošenja odluka, a pre svega na inertnost, averziju prema gubitku i heuristike [44], primenjuju na kontekst prikaza izbora, da bi se osoba navela na određenu opciju. Gurkanje su 2008. uveli Thaler i Sunstein kao grupu mehanizama koji imaju plemenite ciljeve [22]. Na primer, korisniku se neke opcije, poput hrane za naručivanje, mogu prikazati u određenom poretku tako da neke zdravije opcije budu pri vrhu liste preporuka, iako istorijski nisu bile u skladu sa korisnikovim preferencama [22].

Digitalno gurkanje definišemo kao specifičan dizajn korisničkog interfejsa nekog sistema, takav da navodi korisnika na određenu promenu ponašanja ili stavova [47]. Namera je da se, bez ugrožavanja slobode izbora, korisnik "gurka" u pravcu gde je njegov izbor u najboljem interesu društva, a dugoročno i u njegovom najboljem interesu [23].

Studija iz 2013. [37] posmatra uticaj osnovnih podešavanja korisničkog interfejsa elektronskog zdravstvenog sistema na količinu odabranih

tretmana pri prijavljivanju. Rezultati pokazuju da ukoliko su svi tretmani bili preselektovani, a od korisnika se očekivalo da izbací one koje ne želi (eng. *opt-out*), naručeno je znatno više testova nego u slučaju dizajna gde svi tretmani nisu selektovani, a od korisnika se očekuje da odabere one koje želi (eng. *opt-in*).

Takođe, Amazon je do februara 2023. nudio opciju AmazonSmile, koja obezbeđuje donacije u ime kupaca preko standardne kupovine. Naime, ukoliko korisnici odaberu kupovinu preko AmazonSmile sistema, Amazon će donirati 0.5% od cene nekih proizvoda odabranoj humanitarnoj organizaciji [30].

Istraživanje iz 2020. razmatra uticaj korisničkog interfejsa sličnog sistema, gde je prvi pristup opt-in, a drugi prinudan izbor (eng. *forced-choice*). U prvom sličaju, pri prijavljivanju, korisnici mogu da označe ako žele da onlajn prodavnici pristupe preko donatorskog sistema. U drugom slučaju, pre nego što se otvori forma za prijavljivanje, korisnicima se prikazuje prozor koji ih informiše o donatorskom sistemu, i pita da li žele da ga koriste, ili da nastave bez njega. Rezultati pokazuju da korisnici 6.56 puta verovatnije biraju donatorski sistem ako korisnički interfejs koristi prinudan izbor prikaza. Dodatno, pokazuje se da u slučaju prinudnog izbora treba primeniti minimalistički dizajn. Takođe, istraživanje pokazuje da godine, primanja i pol ne utiču na ishod.

3.2 Primer Vitable

Vitable¹ je onlajn kompanija koja se bavi prodajom personalizovanih suplemenata. Pri prijavljivanju, korisnik prvo popunjava kviz o svom zdravstvenom stanju. Na osnovu rezultata sistem preporučuje osnovni skup suplemenata, koji korisnik može da proširi po potrebi. Grupe suplemenata, objašnjenje i prosečna cena suplementa po grupi dati su u tabeli 1.²

Tabela 1: Vitable, prosečne cene suplemenata po kategorijama

Kategorija	Opis	Prosečna cena
Vitamini	Jedinjenja potrebna telu u malim količinama	10.6
Minerali	Elementi esencijalni za funkcionisanje	12.5
Biljke	Terapeutska delovanja, deo tradicionalne medicine	16.5
Probiotici	Pravilno funkcionisanje digestivnog sistema	16
Specijaliteti	Dodatno poboljšanje zdravlja	16.58
Suplementi u prahu	Specifični zdravstveni ciljevi i problemi	32

Jedno od pitanja u kvizu je da li korisnik oseća vizuelnu vrtoglavicu nakon dugog korišćenja ekrana. Dugo korišćenje ekrana je realnost modernog čoveka, te je vizuelna vrtoglavica očekivana posledica. Zbog toga, pretpostavka je da će korisnik odgovoriti pozitivno na to pitanje. Na kraju testa, sistem preporučuje vitamin Astaxanthin iz kategorije specijaliteta (cena AUD 19) kao rešenje za vizuelnu vrtoglavicu.

Ako odaberemo po jedan proizvod iz prve tri kategorije i Astaxanthin iz četvrte kategorije, iz 1, prosečna cena paketa je AUD 74.6. Na slici

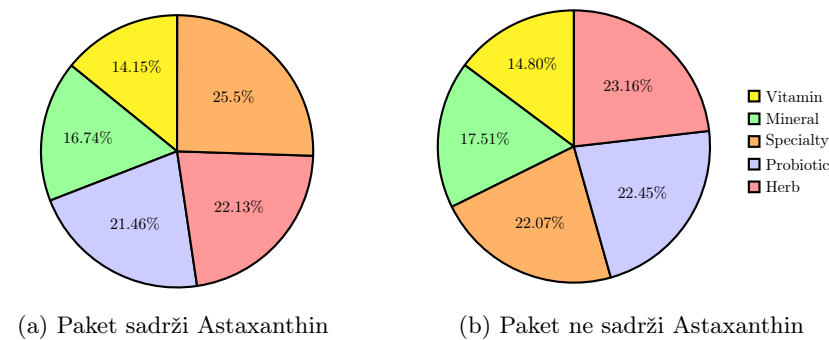
¹Dodatne informacije o Vitable pogledati na adresi <https://www.vitable.com.au>

²Podaci o cenama su dostupni na adresi <https://www.vitable.com.au/products>, a kod za izračunavanje statistika na adresi <https://github.com/pearpanda/mnsr-sistemi-preporuka/blob/main/VitableStatistics.ipynb>

1a prikazan je udeo cene svakog suplementa u takvom paketu. Izostavljamo kategoriju suplemenata u prahu, jer su oni namenjeni za konkretne zdravstvene probleme i specifičnije ciljeve.

Slično, ako odaberemo po jedan proizvod iz prve tri kategorije i proizvod koji nije Astaxanthin iz četvrte kategorije, prosečna cena paketa je AUD 71.7. Na Slici 1b prikazan je udeo suplementa u takvom paketu.

Prosečna cena paketa je veća ako je Astaxanthin izabran. Takođe Astaxanthin je u proseku najskuplji suplement u posmatranom paketu. Pretpostavimo da svi ljudi koriste računar ili mobilni telefon i da će oni odgovoriti pozitivno na postavljeno pitanje. Onda bi svi ljudi imali Astaxanthin u spisku osnovnih preporučenih proizvoda, u opt-in stilu. Takav vid digitalnog gurkanja se ne preporučuje etičkim dizajnerskim principima.



Slika 1: Prosečni udeo cene suplementa u paketu koji sadrži po jedan suplement iz svake kategorije

4 Posledice masovne izloženosti sistemima preporuka

Kada sistemi preporuka prikazuju korisniku sadržaj za koji je predviđeno da mu najbolje privuče pažnju, u sadržaju se nalaze i dezinformacije - netačne informacije koje imaju zadatak da zavaraju onog ko na njih naiđe, i lažne vesti - dezinformacije koje se predstavljaju kao cela ili deo vesti. Takođe, pored posledica sistema preporuka zbog sadržaja koji je preporučen korisnicima, postoje i problemi koji se javljaju zbog samog vremena koje ljudi provode na uređajima, a koje sistemi preporuka povećavaju.

4.1 Posledice po socijalne veze

Društvene mreže, a i mobilni uređaji često služe da bi se ostvarila i održala veza sa drugim ljudima [25]. Iako su mobilni telefoni i društvene mreže doneli mogućnost komunikacije i povezivanja na daljinu, u isto vreme oni sprečavaju povezivanje sa ljudima u bliskoj okolini. Posledica je da se ljudi osećaju istovremeno i povezani i socijalno usamljeni [14]. U eksperimentima je pokazano da samo prisustvo telefona ometa vezu i smanjuje empatiju, poverenje i osećaj bliskosti između dve osobe koje su fizički prisutne na istom mestu [38].

Pored toga što uređaji ometaju održavanje i produbljivanje veza između bliskih ljudi, oni takođe i smanjuju mogućnost sticanja novih poznanika

i prijatelja. Još jedna od posledica prisustva uređaja u socijalnim okupljanjima je smanjivanje sposobnosti za emotivno povezivanje sa strancima [14] [38], što dodatno može da produbi osećanje usamljenosti.

Količina vremena koje ljudi provode na uređajima takođe utiče i na porodične odnose. Korišćenje telefona od strane roditelja tokom igranja sa decom dovodi do značajnih nivoa neraspoloženja kod dece [14] [32], a pošto 50% roditelja priznaje da bar tri puta dnevno pravi prekide u igri sa decom zbog mobilnog telefona [14], dolazimo do situacije gde mnogo dece imaju povećan nivo negativnih osećanja. Takođe, tokom detinjstva, pažnja porodice je kritična za stimulaciju komunikacijskih veština deteta [19]. Međutim, ne provode samo roditelji dosta vremena na uređajima. Deca mlađa od 14 godina u proseku provedu dva puta više vremena na uređajima nego u komunikaciji sa porodicom [14] [10], što ne samo da dodatno utiče na porodične odnose već i otvara mogućnost ka budućim problemima u komunikaciji dece

4.2 Širenje dezinformacija

Preko sistema preporuka se lažne vesti šire šest puta brže nego istinite [46] [14]. Ovo se dešava jer će korisnik biti više zainteresovan senzacionalnim naslovima i sadržajem koji je karakterističan lažnim vestima. Takve vesti češće prouzrokuju bes kod korisnika [27], čime se dodatno ubrzava njihovo širenje [12] [14].

Kada veliki broj ljudi dođe u kontakt sa istim dezinformacijama i lažnim vestima, dolazi do loše informisanosti javnog mnjenja i samim tim nemogućnosti da se reaguje adekvatno u situacijama poput klimatskih promena ili epidemije. Takođe se šire i teorije zavere koje smanjuju verovanja ljudi u naučne činjenice i imaju loš uticaj na socijalno ponašanje [45]. Na primer, istraživanje sa Oxford-a pokazuje da je u 22 miliona ispitanih objava više bilo dezinformacija, lažnih vesti i teorija zavere nego tačnih informacija [14] [21].

Postojanje botova čiji je cilj širenje dezinformacija, lažnih vesti i teorija zavere dodatno otežava rešavanje ovog problema. 45% objava na Twitter-u o Covid-19 virusu je bilo postavljeno od strane botova [14] [34], čemu automatska i ručna provera činjenica (eng. *fact checking*) nije u mogućnosti da se suprotstavi [35].

4.3 Posledice po demokratsko funkcionisanje

Uticaj na politiku postoji preko širenja propagande i polarizujućeg sadržaja. Propaganda je informacija (često dezinformacija) čiji je cilj da utiče na javnost ili manipuliše njom radi ostvarenja nekog političkog cilja. Polarizujući sadržaji su namenjeni da stvore razdor i segregaciju u narodu [14]. Sadržaj sa političkim protivnicima ima 67% više šanse da bude podeljen [39], pa će ga sistemi preporuka češće preporučiti korisnicima. Na primer, 2017. godine vojske Mjanmara je zloupotrebila sistem preporuka Facebook-a započevši propagandnu kampanju protiv manjine u Mjanmaru. To je dovelo do genocida i dolazak generala vojske Mjanmara na vlast [40].

Za osetljivost ljudi na političke vesti govori podatak da preko 20% neodlučenih glasača menja svoje mišljenje na osnovu redosleda rezultata na pretraživaču [14] [11] i da lažna politička vest može da promeni sećanje ljudi tako da su ubeđeni da se ona zapravo desila [31].

Pored lakog i brzog širenja ovakvih vesti, problem uvećava njihova količina i dugotrajnost. Par nedelja pred napad na Kapitol, 5 miliona političkih dezinformacija je objavljeno na Facebook-u [14] [28]. Takođe, 3 meseca pred izbore u Sjedinjenim Američkim Državama 2016. godine, objavljeno je više lažnih političkih naslova, tri puta više ljudi ih je pročitao, i čak i posle dve godine takve lažne vesti su bile u prvih 10 priča na Twitter-u [14] [42] [16].

Esencijalni preduslov za demokratiju čini autonomnost individue [17] koja se gubi zbog širenja polarizujućih vesti i propagande. Posledica je viđenje da su razlike u stavovima sa političkim protivnicima veće nego što zapravo jesu, što vremenom dovodi do segregacije društva. Kao indirektna posledica, demokratija postaje disfunkcionalna jer neistomišljenici i politički protivnici teže rešavaju probleme zajedno, zbog toga što nemaju ništa zajedničko [15].

Manipulacija stanovništva od strane legitimno demokratski izabrane vlasti može se u nekim situacijama smatrati opravdanom. Obično su to situacije gde se narodom manipuliše zarad dobra stanovništva ili države [17]. U ovakvim slučajevima sistemi preporuka bi olakšali širenje poruke i dopiranje do velikog dela društva.

5 Diskriminacija u sistemima preporuka

Glavni cilj sistema preporuka je da pruži personalizovane sugestije svojim korisnicima. Međutim, zbog određenih nepravilnosti prilikom davanja takvih sugestija, otvara se pitanje diskriminacije u sistemima preporuka. Postoji niz problema koji proizlaze iz pristrasnih algoritama, koji često mogu za ishod da imaju nejednakost u pristupu informacijama i resursima [26]. Razumevanje i rešavanje ovog problema je jedan od ključnih koraka ka razvijanju pravičnih i nediskriminišućih personalizacija.

5.1 Primeri diskriminacije

Algoritmi zaduženi za davanje preporuka poslova mogu stvoriti etičku diskriminaciju u oblasti zapošljavanja kada sistem favorizuje gore kvalifikovanu osobu druge etničke pripadnosti [18]. Algoritam koji u sebi ima naznake diskriminacije ne rangira određene korisnike pravilno. Zbog toga, oni neće biti razmatrani za određene preporuke. U slučaju zaposlenja, takav propust direktno utiče na poslovnu sudbinu osobe. Takođe, u situaciji kada algoritam ima ulogu u odlučivanju davanja kredita nekoj osobi, može doći do diskriminacije tako što će algoritam favorizovati određeni profil osobe, dok će neke druge grupe ljudi ostati bez takve finansijske prilike [18].

Primer rodne diskriminacije se može dešavati u oblasti zapošljavanja. Naime, ako bismo posmatrali mušku i žensku osobu koje imaju identične kvalifikacije za određeni posao, oni će dobijati različite preporuke poslova. Žene će češće nego muškarci vidati poslove sa niskim platama koji zahtevaju manje radnog iskustva, koji zahtevaju bolje socijalne i administrativne veštine i koji sadrže reči koje se češće pripisuju ženama. Poslovi koje vide samo muškarci sadrže veći broj reči koje opisuju karakteristike kao što su inženjer, liderstvo i prekovremeni rad, dok je veća verovatnoća da će žene biti izložene oglasima koji uključuju reči kao što su pomoć, administrativne sposobnosti, strpljivost i temperamentnost, i ostale koje možemo videti u tabeli 2 [49].

	Ženske reči	Muške reči
Veštine	pomaganje, podaci, administracija, razgovor, komunikacioni alati, dokumentacija	donošenje odluka, dizajn, saradnja, timski rad, inženjering, menadžment, nezavisnost, vođstvo
Radni odnos	fleksibilan, fiksno radno vreme	putovanje do posla, prekovremen rad, noćne smene
Pogodnosti	osnovna plata, odmor za venčanje, porodijsko odsustvo, penzija, roditeljsko odsustvo	prevoz, zdravstveno osiguranje, odmor, obrok, bonus, akcije
Firma	učenje	berza
Zahtevi	tačna, pažljiva, komunikativna, predstavnic, strpljiva, zdrava, temperamentna, skoro diplomirala, mlada, bez iskustva, verodostojna	samomotivisan, pritisak, inovativan, iskusan

Tabela 2: Rodne razlike u rečima i rodni stereotipi [49]

Gledano sa pozitivne strane, ako bi cilj sistema preporuka bio podrška lokalnim zajednicama, moguće je koristiti pozitivnu diskriminaciju u korist lokalnih proizvoda, usluga ili stvaralaca kako bi se podržala lokalna ekonomija i kultura.

5.2 Da li algoritam namerno vrši diskriminacije?

Sistemi preporuke su algoritmi koji rade na osnovu podataka. Stoga, pristrasnost proizilazi iz podataka korišćenih u njihovom treniranju, odnosno internih nejednakosti prisutnih u podacima koje algoritam koristi. Ipak, odsustvo zlonamernosti u sistemima preporuka ne umanjuje njihov uticaj na reprodukciju i pojačanje diskriminacije u društvu.

Sa druge strane, algoritam može da radi na određeni način da bi davao doprinos pojedincima. Na primer, neovlašćeni pristup preko HTTP saobraćaja tako što se željeni preporučeni sadržaj ubrizgava u sadržaj veb strane (eng. *Web Injection*) koju korisnik posmatra. To će kasnije naterati korisnika da poseti ciljani proizvod koji je preporučen, i na taj način će se odigravati manipulacija sistema preporuka kroz sesije pregledača [50].

6 Etički osvrt

Iako namenjeni da preporuče sadržaj koji je najrelevantniji korisniku, postoji sukob interesa između različitih strana u definisanju koliko je preporuka korisna. Na primer, prodavcu se možda više isplati da preporuči predmet gde je njegova zarada veća, umesto da preporuči predmet koji je najkorisniji kupcu [8]. U primeru odabira zdravstvenih tretmana iz 3.1, korisnik očekuje da ga sistem navodi na optimalan izbor tretmana, a ne onaj koji generiše najveći profit. Manjak transparentnosti i eksploatacija heuristika u donošenju odluka korisnika [7] čine takav sistem manipulativnim.

U slučaju reklama, proizvodi koji se preporučuju mogu biti nekvalitetni ili tvrdnje prikazane u reklamama lažne. Dodatno, neke reklame, iako profitabilne za kompanije, mogu biti veoma štetne za posebne korisnike.

Na primer, reklame za kockarnice mogu katastrofalno uticati na zavisnike od kockanja.

Takođe, prikupljanje podataka od strane sistema ističe rizike po privatnost korisnika. Radi boljeg efekta reklama, postaje bitnije formirati profile korisnika. Profili mogu sadržati, na primer, interesovanja, starost, pol, politička opredeljenja, mesto stanovanja, bogatstvo (eng. *net worth*). Istraživanja pokazuju kako praćenje aktivnosti nije samo u vezi sa pojedinačnim sajtovima, već preko reklama i usluga analize saobraćaja povezuje veliki broj naizgled nezavisnih sajtova [9], pri čemu način upotrebe prikupljenih podataka često nije jasno definisan.

Da bi se obezbedio što pozitivniji uticaj sistema preporuka i zaobišle određene etičke zamke, važno je ustanoviti dizajnerske principe koji vode ka etički svesnom dizajnu sistema preporuka.

6.1 Dizajn etičkog sistema preporuka

Izdvajamo dve metode koje vode ka etički prihvatljivijim dizajnom sistema preporuka - etički dizajn ubeđivanja (eng. *ethical persuasive design*) [6] i sistemi podrške u promeni stavova (eng. *behavior change support system*) [24]. Naime, prihvatanje preporuke treba da bude dobro po korisnika ili društvo, a proces ubeđivanja maksimalno transparentan tako da je i efikasan. Dizajn treba da odgovara potrebama korisnika, ne remeti njegove ostale aktivnosti i obezbedi odgovarajući prostor izbora u opt-in stilu. Takođe, treba postići ravnotežu između zaštite privatnosti i personalizacije. Dizajn treba da omogući personalizovane opcije podešavanja privatnosti, koje će pružiti korisnicima veću kontrolu nad svojim podacima.

Postoji rizik da bi transparentan sistem bio manje efikasan. Stoga, izučava se uticaj otkrivanja psihološkog delovanja podrazumevanih podešavanja na efikasnost sistema, kao i uticaj informisanja korisnika o cilju njihove primene. Bruns et. al. [7] su sprovedli istraživanje na primeru donacija ustanovama koje se bave ublažavanjem posledica klimatskih promena. Rezultati pokazuju da efikasnost sistema u slučaju ma koje transparentnosti, kao i kombinovane, nije umanjena, te savetuju da bi ih trebalo implementirati.

Takođe, u kontekstu širenje objava, informacija i vesti (deo 4), sistem preporuka bi imao pozitivan uticaj na stanovništvo ako bi se iskoristio za propagiranje činjenica i otkrića. Da bi došlo do toga potrebno je rešiti problem dezinformacija, koji današnja provera činjenica (eng. *fact checking*) većinski nije u stanju da reši. Jedan od predloga za rešavanje ovog problema je uključivanje strategija za upravljanje dezinformacijama iz istraživanja socijalnih nauka u modele sistema preporuka [13].

Uz implementaciju etički prihvatljivijeg dizajna, važna je i edukacija korisnika i dizajnera. Korisnici treba da imaju uvid u to kako algoritmi generišu preporuke, što može povećati svest o važnosti etičkog korišćenja ovakvih sistema. Takođe, veća edukacija programera na temu etičkih smerica koje treba da koriste prilikom implementacije algoritama preporuka bi doprinela razvoju fer algoritama [24].

U kontekstu implementacije algoritama mašinskog učenja, pravičnost se može postići pravilnom primenom 3 vrste metoda. Prve su metode koje se odnose na ispravljanje podataka pre samog učenja, gde bi se već u podacima uklonila pristrasnost, mada to može naškoditi tačnosti algoritma. Druga vrsta metoda su one koje se primenjuju na sam proces učenja i koje treba da balansiraju između tačnosti i pravičnosti. Treće vrste metoda

su one koje mogu da transformišu izlaz iz modela tako da otklone delove koji utiču na to da algoritam ne bude fer [26]. Svaka od ovih metoda ima svoje prednosti i nedostatke, a izbor metoda nije samo tehničko pitanje već zahteva razmatranje socijalnih i pravnih konteksta.

Najzad, trebalo bi podstaći korisnike na što češće davanje povratne informacije o preporukama, jer ti podaci mogu doprineti poboljšanju sistema [29]. Uz to, veća saradnja između tehnoloških kompanija, stručnjaka za etiku i relevantnih regulatornih tela može dovesti do uspostavljanja boljih standarda i smernica koji promovišu odgovorno korišćenje sistema preporuka [26].

7 Zaključak

Sistemi preporuka su široko rasprostranjeni, i imaju uticaj na pojedinca i društvo preko sadržaja koji prikazuju. Unapređivanje algoritama preporučivanja se sprovodi u smislu poboljšanja kvaliteta preporuka, ali i upotrebi socijalnih faktora koji utiču na njihovo prihvatanje. Na primer, implementacija metoda ubeđivanja, heuristika u donošenju odluka i virtuelnih agenata poboljšava stopu prihvatanja preporuka od strane korisnika.

Međutim, sukob interesa između proizvođača i korisnika, uticaj preporuka na demokratiju i razdvajanje u društvu, i diskriminacija kao posledica korišćenja nedovoljno reprezentativnih podataka su situacije u kojima implementacija etičkih dizajnerskih principa obezbeđuje smanjen negativni uticaj, i nalazi balans između ubeđivanja i manipulacije.

Metode etičkog dizajna ubeđivanja i sistema podrške u promeni stavova naglašavaju važnost svesnog dizajniranja sistema, odnosno ohrabruju dizajnere da razmišljaju o potencijalnim etičkim problemima pre implementacije. Etički svestan sistem je što transparentniji tako da je i efikasan, odgovara potrebama korisnika, koristi opt-in dizajn, reguliše pitanja privatnosti kroz ponuđene konfiguracije, i koristi metode provera činjenica za suzbijanje lažnih vesti.

Iako postoji pomak u definisanju etičkih dizajnerskih principa, i postavljanju teorijskih okvira u toj sferi, potrebno je i implementirati takav sistem da bi se dokazalo da je on delotvoran u praksi. Potencijalne teme budućih istraživanja mogu upravo biti implementacija, testiranje i evaluacija etičkih sistema preporuka, odnosno potvrđivanje i unapređivanje postojećih teorijskih modela.

Literatura

- [1] 57 amazon statistics to know in 2023.
- [2] Organizing information – how google search works.
- [3] Charu C. Aggarwal. *Recommender Systems*. Springer International Publishing.
- [4] Alaa Alslaity and Thomas Tran. Towards persuasive recommender systems. *2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT)*, 2019.
- [5] Alaa Alslaity and Thomas Tran. Users’ responsiveness to persuasive techniques in recommender systems. *Frontiers in Artificial Intelligence*, 4, 2021.
- [6] Dennis Benner, Sofia Schobel, Andreas Janson, and Jan Marco Leimeister. How to achieve ethical persuasive design: A review and theoretical propositions for information systems. *AIS Transactions on Human-Computer Interaction*, 14(4):548–577, 2022.
- [7] Hendrik Bruns, Elena Kantorowicz-Reznichenko, Katharina Klement, Marijane Luistro Jonsson, and Bilel Rahali. Can nudges be transparent and yet effective? *Journal of Economic Psychology*, 65:41–45, 2018.
- [8] Robin Burke, Alexander Felfernig, and Mehmet H. Göker. Recommender systems: An overview. 32(3):13–18.
- [9] Abdelberi Chaabane, Mohamed Ali Kaafar, and Roksana Boreli. Big friend is watching you: analyzing online social networks tracking capabilities. In *Proceedings of the 2012 ACM workshop on Workshop on online social networks*, WOSN ’12, pages 7–12. Association for Computing Machinery.
- [10] Laura Donnelly. Children spend twice as long on smartphones as talking to parents, Feb 2019.
- [11] Robert Epstein and Ronald E. Robertson. The search engine manipulation effect (seme) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, 112(33), 2015.
- [12] Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. Anger is more influential than joy: Sentiment correlation in weibo. *PLoS ONE*, 9(10), 2014.
- [13] Miriam Fernandez and Alejandro Bellogín. Recommender systems and misinformation: the problem or the solution? 2020.
- [14] Center for Humane Technology.
- [15] Center for Humane Technology. Democratic functioning.
- [16] Knight Foundation, Oct 2018.
- [17] Sergio Genovesi, Katharina Kaesling, and Scott Robbins. *6.4 Manipulation*. Springer Nature Switzerland AG, 2023.
- [18] Sergio Genovesi, Katharina Kaesling, and Scott Robbins, editors. *Recommender Systems: Legal and Ethical Issues*, volume 40 of *The International Library of Ethics, Law and Technology*. Springer International Publishing, Cham, 2023.

- [19] Caroline Gooden and Jacqui Kearns. The importance of communication skills in young children. research brief. summer 2013. *Human Development Institute*, 2013.
- [20] Martin Holzwarth, Chris Janiszewski, and Marcus M Neumann. The influence of avatars on online consumer shopping behavior. *Journal of Marketing*, 70(4):19–36, 2006.
- [21] Philip Howard, Bence Kollanyi, Samantha Bradshaw, and Lisa-Maria Neudert. Social media, news and political information during the us election: Was polarizing content concentrated in swing states? 02 2018.
- [22] Mathias Jesse and Dietmar Jannach. Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports*, 3:1–3, 2021.
- [23] Randi Karlsen and Anders Andersen. Recommendations with a nudge. *Technologies*, 7(2):45–47, 2019.
- [24] Pasi Karppinen and Harri Oinas-Kukkonen. Three approaches to ethical considerations in the design of behavior change support systems. *Persuasive Technology*, page 87–98, 2013.
- [25] Simon Kemp. Digital 2023 april global statshot report - datareportal – global digital insights, Apr 2023.
- [26] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, Juntao Tan, Shuchang Liu, and Yongfeng Zhang. Fairness in Recommendation: Foundations, Methods, and Applications. *ACM Trans. Intell. Syst. Technol.*, 14(5):1–19, October 2023.
- [27] Donna Lu. Fake news gets shared more when it is angry and anxiety-inducing, May 2020.
- [28] Ryan Mac and Craig Silverman. After the us election, key people are leaving facebook and torching the company in departure notes, Dec 2020.
- [29] Dionisis Margaritis, Costas Vassilakis, and Dimitris Spiliotopoulos. What makes a review a reliable rating in recommender systems? *Information Processing & Management*, 57(6):102304, November 2020.
- [30] Christian Meske, Ireti Amojó, and Peter Mohr. Digital nudging to increase usage of charity features on e-commerce platforms. *WI2020 Zentrale Tracks*, page 1203–1218, 2020.
- [31] Gillian Murphy, Elizabeth F. Loftus, Rebecca Hofstein Grady, Linda J. Levine, and Ciara M. Greene. False memories for fake news during ireland’s abortion referendum. *Psychological Science*, 30(10):1449–1459, 2019.
- [32] Sarah Myruski, Olga Gulyayeva, Samantha Birk, Koraly Pérez-Edgar, Kristin A. Buss, and Tracy A. Dennis-Tiwary. Digital disruption? maternal mobile device use is related to infant social-emotional functioning. *Developmental Science*, 21(4), 2017.
- [33] Clifford Nass, Youngme Moon, and Nancy Green. Are machines gender neutral? gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, 27(10):864–876, 1997.
- [34] Bobby Allyn / NPR. Researchers: Nearly half of accounts tweeting about coronavirus are likely bots, May 2020.

- [35] Jon-Patrick Allem; Assistant Professor of Research. Social media fuels wave of coronavirus misinformation as users focus on popularity, not accuracy, Nov 2023.
- [36] Stefano Panzeri, Cesare Magri, and Ludovico Carraro. Sampling bias. 3(9):4258.
- [37] C. Adam Probst, Victoria A. Shaffer, and Y. Raymond Chan. The effect of defaults in an electronic health record on laboratory test ordering practices for pediatric patients. *Health Psychology*, 32(9):995–1002, 2013.
- [38] Andrew K. Przybylski and Netta Weinstein. Can you connect with me now? how the presence of mobile communication technology influences face-to-face conversation quality. *Journal of Social and Personal Relationships*, 30(3):237–246, 2012.
- [39] Steve Rathje, Jay J. Van Bavel, and Sander van der Linden. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), 2021.
- [40] Special Reports. Why facebook is losing the war on hate speech in myanmar, Aug 2018.
- [41] Paul Resnick and Hal R. Varian. Recommender systems. 40(3):56–58.
- [42] Craig Silverman. This analysis shows how viral fake election news stories outperformed real news on facebook, Nov 2016.
- [43] Herbert W. Simons and Jean G. Jones. *Persuasion in society*. Taylor & Francis, second edition, 2011.
- [44] Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, first edition, 2008.
- [45] Sander van der Linden. The conspiracy-effect: Exposure to conspiracy theories (about global warming) decreases pro-social behavior and science acceptance. *Personality and Individual Differences*, 87:171–173, 2015.
- [46] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.
- [47] Markus Weinmann, Christoph Schneider, and Jan vom Brocke. Digital nudging. *Business & Information Systems Engineering*, 58(6):433–434, 2016.
- [48] Kyung H. Yoo, Ulrike Gretzel, and Markus Zanker. *Persuasive recommender systems: Conceptual background and implications*. Springer, 2013.
- [49] Shuo Zhang. Measuring Algorithmic Bias in Job Recommender Systems: An Audit Study Approach.
- [50] Yubao Zhang, Jidong Xiao, Shuai Hao, Haining Wang, Sencun Zhu, and Sushil Jajodia. Understanding the Manipulation on Recommender Systems through Web Injection. *IEEE Transactions on Information Forensics and Security*, 15:1 – 12, 2020. Conference Name: IEEE Transactions on Information Forensics and Security.