

Odgovorna umjetna inteligencija



Pregled uticaja umjetne inteligencije na ljudska prava i perspektive
medijske pismenosti u kontekstu Bosne i Hercegovine

www.coe.int/freedomofexpression

Engleska verzija:

Responsible Artificial Intelligence

Ova studija je provedena uz podršku Vijeća Evrope i projekta pod nazivom „Medijska i informacijska pismenost: za ljudska prava i demokratiju“ u okviru Akcionog plana Vijeća Evrope za Bosnu i Hercegovinu za period 2018-2021. Finansiranje na nivou Akcionog plana obezbijedili su Luksemburg i Norveška.

Mišljenja iznesena u ovoj publikaciji su isključivo odgovornost autora i ne predstavljaju nužno službenu politiku Vijeća Evrope.

Reproduciranje odlomaka (do 500 riječi) je dozvoljeno za nekomercijalne svrhe, pod uslovom da je očuvan integritet teksta, da se odlomak ne koristi izvan konteksta, da ne pruža nepotpune informacije, odnosno da na neki drugi način ne dovodi čitaoca u zabludu u pogledu prirode, opsega ili sadržaja teksta. Izvorni tekst se uvijek mora citirati kako slijedi: „© Vijeće Evrope, godina objavljivanja“. Svi ostali zahtjevi u vezi sa reprodukcijom/prevodom cjelokupnog dokumenta ili nekog njegovog dijela trebaju se uputiti Direktoratu za komunikacije Vijeća Evrope (F-67075 Strasbourg Cedex ili na adresu publishing@coe.int).

Dizajn i grafička priprema: Odjel za informaciono društvo, Vijeće Evrope

Fotografije: ©Shutterstock

Jedinica za lektorisanje SPDP nije lektorirala ovu publikaciju radi ispravke štamparskih i gramatičkih grešaka.

© Vijeće Evrope, juni 2022.

Odgovorna umjetna inteligencija

Pregled uticaja umjetne inteligencije na ljudska prava
i perspektive medijske pismenosti u kontekstu
Bosne i Hercegovine

februar 2022.

Autorice:

Bojana Kostić

je dugogodišnja zagovornica slobode medija sa velikim iskustvom u oblasti istraživanja koja provodi zajednica i inkluzivnih istraživanja. U svojim istraživanjima bavi se problematikom ljudskih prava i tehnologije, uz poseban osvrt na slobodu izražavanja na internetu, online zlostavljanje, sigurnost i ćutanje.

Caroline Sindere

je kritički dizajner i umjetnik. Posljednjih nekoliko godina proučava međusobni uticaj umjetne inteligencije, intersekcionalne pravde, dizajna sistema, negativnih uticaja i politike u digitalnom konverzionom prostoru i na tehnološkim platformama.

Vijeće Evrope

Sadržaj

I. Uvod	3
II. Osnovni rječnik UI	4
III. UI i svakodnevni život: ključni koncepti	8
3.1. Digitalna moć	8
3.2. Uticaj UI na donošenje odluka ljudi	9
3.3. Prenamjena podataka	10
3.4. Moć kontrole i djelovanja korisnika	12
3.5. Štetni uticaji algoritamskih sistema	12
IV. Pregled: UI i sloboda izražavanja	13
4.1. Moderacija sadržaja	14
4.2. Priređivanje i dijeljenje sadržaja	15
4.3. Govor mržnje i UI	16
4.4. Uvjerljiv krivotvoreni sadržaj (eng. deepfakes)	17
4.5. Perspektive slobode medija	18
4.6. Preliminarni zaključak	20
V. Reglativa i politike: kratki pregled	21
5.1. Medijska i informacijska pismenost (MIL) i UI: perspektive u zemlji	23
VI. Zaključak	24

Skraćenice

AAAI	Udruženje za unapređenje umjetne inteligencije
AARP	Američko udruženje penzionera
ADS	Automatizovani sistem odlučivanja
UI	Umjetna inteligencija
AVM	Audio-vizuelni mediji
BiH	Bosna i Hercegovina
CAHAI	Ad-hoc odbor Vijeća Evrope za umjetnu inteligenciju
VE	Vijeće Evrope
CoM	Odbor ministara
RAK	Regulatorna agencija za komunikacije
Dekl.	Deklaracija
DMA	Akt o digitalnim tržištima
DSA	Akt o digitalnim uslugama
EK	Evropska komisija
ECI	Evropska građanska inicijativa
EDRi	Evropska digitalna prava
EFF	Fondacija za elektronske granice
EU	Evropska unija
FSLN	Sandinistički front nacionalnog oslobođenja
GDPR	Opšta uredba o zaštiti podataka
GIFCT	Globalni internetski forum za borbu protiv terorizma
HITL	Čovjek u petlji (eng. human-in-the-loop)
IBM	International Business Machines Corporation
IEEE	Institut inženjera elektrotehnike i elektronike
KRIK	Mreža za istraživanje kriminala i korupcije
MIL	Medijska i informacijska pismenost
OSCE	Organizacija za sigurnost i saradnju u Evropi
PSVE	Parlamentarna skupština Vijeća Evrope
Rec	Preporuka
RFoM	Predstavnik OSCE-a za slobodu medija
SAIFE	Vještačka inteligencija i sloboda izražavanja u fokusu
UNHRC	Vijeće Ujedinjenih nacija za ljudska prava
UNISYS	United Information Systems
UvA	Univerzitet u Amsterdamu
ZB	Zapadni Balkan

I. Uvod

U javnom i privatnom sektoru, kako online tako i offline, umjetna inteligencija (UI) i algoritmi donose odluke koje su značajne za život ljudi. Stoga je UI postala predmet razmatranja globalne politike u cilju rješavanja izazova umjetne inteligencije u domenu ljudskih prava. Ova studija o odgovornosti, umjetnoj inteligenciji i odgovornoj UI bavi se višestrukim uticajem UI na ljudska prava te taj uticaj konkretno ispituje kroz široku prizmu slobode izražavanja i slobode medija.

Velike tehnološke kompanije (platforme društvenih medija, platforme za streaming i dijeljenje, a ponajviše Facebook, Google, Twitter, Tik-Tok, Netflix, Amazon, Apple, Microsoft, IBM, Samsung itd.) stižu do sada nezabilježenu ekonomsku i političku moć. Rastući razvoj i primjena UI i algoritama povećava njihov društveni uticaj i direktno se odražava na ljudska prava pojedinaca, demokratiju i vladavinu prava (CM Decl (13/02/2019)¹, paragraf 9). Analiziranje, objašnjavanje i situiranje uticaja UI je neophodno za stvaranje novih oblika okvira odgovornosti i novih intervencija medijske i informacijske pismenosti (MIL). S obzirom na dati kontekst ova studija ima dva cilja: i. upoznati aktere i stručnu zajednicu sa praktičnim implikacijama društvenog uticaja UI; ii. predložiti resurse za MIL u kontekstu konkretnog uticaja UI.

Studija se sastoji od dva konceptualna elementa u okviru kojih se analizira:

i) Moć automatizovanih sistema u svakodnevnom životu ljudi u digitalnom ambijentu, uz poseban osvrt na sljedeće koncepte:

- uticaj UI na donošenje odluka ljudi
- digitalna moć
- moć djelovanja korisnika
- prenamjena podataka
- uvjerljivi krivotvoreni sadržaji (eng. deepfakes)
- štetni uticaji algoritamskih sistema.¹

ii) Uticaj UI na slobodu izražavanja i slobodu medija u vezi sa popularnim i globalnim platformama, softverima i aplikacijama.

Zahvaljujući kombinaciji detaljne analize akademske literature i opsežne kompilacije resursa MIL ova studija daje kritički uvid u postojeće stanje u oblasti UI i izazove koje ona postavlja za ljudska prava. Treba napomenuti da ovom studijom nisu obuhvaćeni svi relevantni aspekti UI i ljudskih prava budući da brojne teme vezane za UI ne spadaju u predmet ove studije - kao što su primjena UI u određenim sektorima,² problematika zaštite podataka nevezano za UI i slobode izražavanja, kao i standardi praćenja i ocjene poštovanja ljudskih prava. Neki od ovih aspekata spominju se kao ilustrativni primjeri ili prateći argumenti, ali je fokus na artikulaciji i mapiranju postojećih stručnih i regulatornih diskusija kroz prizmu UI i slobode izražavanja.

Utemeljena na razumijevanju MIL i aktivnog građanstva, ova studija se namjerno fokusira na velike tehnološke kompanije poput Facebooka zbog pažnje koju su nedavno privukle u javnosti i akademskim krugovima (Newton 2021). Pored toga, Facebook je najkorištenija platforma

¹ Napomena: ne postoji opšteprihvaćena ili mjerodavna definicija ovih pojmova. Oni zavise od konteksta, otvorenog su tipa i fluidni.

²Kao što su dobra uprava i vladavina prava, finansijski sistem itd.

društvenih medija u Bosni i Hercegovini (BiH). U zemlji u kojoj 96% stanovništva ima pristup internetu, više od tri četvrtine (78%, VE 2021, str.5) ima naloge na društvenim medijima, od čega 71% (ibid, str.6) čine korisnici Facebooka. Premda nisu tako popularne kao Facebook, platforme za dijeljenje video zapisa koristi 42% korisnika interneta u BiH (ibid, str. 5). U poređenju sa odraslim korisnicima interneta, mladi provode više vremena online – u prosjeku četiri sata (ibid, str. 5) – i češće pasivno konzumiraju sadržaj koji je algoritamski preporučen (Hodžić 2019, str. 32, 34). Međutim, mlađa populacija nije jedina grupa koja je izložena rizicima automatizovanog sadržaja. Korisnici interneta načelno češće čitaju nego što stvaraju ili dijele novi sadržaj i komentare (VE 2021, str. 6) pa je javna rasprava na internetu relativno malo zastupljena (46%, ibid), iako se društveni mediji uglavnom koriste kao forumi za javnu diskusiju (ibid.). S obzirom na to da se radi o zemlji u kojoj je povezanost internetom na visokom nivou, platforme društvenih medija (52%, ibid) i online portali sa vijestima (45%, ibid) predstavljaju važne izvore javnog informisanja i rasprave, ali kao što se vidi iz nalaza Studije o medijskim navikama i stavovima odraslih u Bosni i Hercegovini (VE 2021), većina korisnika pasivno konzumira ovaj sadržaj.

Važno je napomenuti da su informacije o odnosima državnih i nedržavnih aktera u zemlji sa ovim velikim tehnološkim kompanijama oskudne ili nepostojeće (Kostić (u pripremi)). Rastući broj problema koji se javljaju na društvenim medijima, kao što su govor mržnje, polarizacija, sužavanje online građanskih prostora, politička i međunacionalna mržnja, ostaje neriješen (Turčilo et al. (u pripremi); Cvjetičanin 2019, str. 7, 22, 40, 63; Sokol 2020, str. 20). U ovoj zemlji koja prolazi kroz poslijeratni period ta neriješena pitanja imaju drugačiji uticaj i utiču na javni diskurs i izgled za očuvanje mira i postizanje društvene kohezije. Iz tog razloga, pored mnogih drugih intervencija, one u oblasti medijske i informacijske pismenosti, koje građanima nude načine kritičkog korištenja UI i algoritama i razumijevanja njihovog značaja i uticaja na slobodu izražavanja na internetu, su od ključnog značaja za razvijanje demokratske kulture i društvenog prosperiteta u zemlji.

Da bi se razumjeli izazovi koje UI i algoritamski sistemi predstavljaju za ljudska prava, a posebno za slobodu izražavanja i slobodu medija, ova studija daje kritičku procjenu postojećeg digitalnog okruženja i okvira odgovornosti za transparentnost i odgovornost društveno-tehnoloških procesa. U poglavlju II definisani su i objašnjeni osnovni koncepti UI kojima se ova studija bavi. Zajedno sa ilustrativnim primjerima, u narednom poglavlju III data je analiza ključnih koncepata UI. U poglavlju IV dat je pregled izazova koje UI i algoritamski sistemi predstavljaju za slobodu izražavanja i tenzija koje se pojavljuju između slobode medija i velikih tehnoloških kompanija. Kratak pregled regulative dat je u poglavlju V uz poseban osvrt na intervencije u oblasti medijske i informacijske pismenosti. Na samom kraju studije ponovo se ističe značaj uticaja UI i algoritamskih sistema na našu budućnost i predstavlja komplet korisnih resursa u oblasti medijske i informacijske pismenosti (Prilog A), kao i rječnik pojmova (Prilog B) koji bi trebali biti od koristi u planiranju lokalnih intervencija u oblasti medijske i informacijske pismenosti i UI.

II. Osnovni rječnik UI

AI i algoritamsko odlučivanje pojavili su se kao krovni pojmovi stoga ne postoji konsenzus oko njihovih definicija (Zuiderveen Borgesius 2018, str.11) niti jedinstvena klasifikacija koncepata koji se pojavljuju u ovoj oblasti. Pojmovi UI, algoritmi i mašinsko učenje se koriste naizmjenično u struci, akademskoj zajednici, novinarstvu i civilnom društvu. U tehničkom smislu, svaki od ovih izraza ima različito značenje, ali se oni kolokvijalno koriste kao sinonimi.

Zajednički nazivnik koji ove tehnološke izraze okuplja pod zajednički terminološki okvir je njihova funkcionalnost usmjerena na zadatak, veliki skupovi podataka,³ njihove tehničke strukture i njihovi elementi. Radi boljeg razumijevanja terminologije korištene u ovoj studiji u ovom poglavlju je dato objašnjenje sljedećih osnovnih i ključnih koncepata UI:

Algoritmi. Algoritam se može opisati kao „apstraktni, formalizovani računski proces i [...] uopšteno govoreći, algoritam bi se mogao smatrati računarskim programom“ (Zuiderveen Borgesius 2018, str.11) ili „nizom koraka (ili skupom pravila) za rješavanje ili izvršavanje nekog zadatka“ (Onuoha i Nucera 2018, str. 8). Pojmovnik umjetne inteligencije Vijeća Evrope definiše algoritme kao „konačan skup formalnih pravila (logičkih operacija, instrukcija) koja omogućavaju dobijanje rezultata iz ulaznih elemenata. Ovaj skup pravila može biti predmet automatizovanog procesa izvršavanja i oslanjati se na modele dizajnirane putem mašinskog učenja“ (Pojmovnik UI VE). Dok neki algoritmi izvršavaju zadatke automatski, kao u slučaju filtera za uklanjanje neželjene pošte ili hashtagova koji uklanjaju govor mržnje i ekstremistički govor na internetu (GIFCT - Share Industry Database), drugi pomažu moderatorima u donošenju odluka o uklanjanju sadržaja. Na primjer, algoritmi označavaju problematičan sadržaj, ali moderatori su ti koji donose konačnu odluku o tome da li će označeni sadržaj biti zadržan ili uklonjen (Kostić 2021, str. 23).

Umjetna inteligencija (UI) i mašinsko učenje. Umjetna inteligencija je, grubo rečeno, „nauka stvaranja pametnih mašina“ (Zuiderveen Borgesius 2018, str. 12). Pojmovnik Vijeća Evrope definiše UI kao „skup nauka, teorija i tehnika koje imaju za cilj da reprodukuju, pomoću mašine, kognitivne sposobnosti ljudskog bića kako bi se mašini mogli povjeriti složeni zadaci koji su ranije dodjeljivani čovjeku“ (Pojmovnik UI VE). Međunarodne organizacije, kao što je Organizacija za sigurnost i saradnju u Evropi (OSCE), definišu i objašnjavaju UI kao „zasnovanu na algoritmima, a to su skupovi instrukcija koje su pripremili ljudi, sa programiranim procedurama za pretvaranje ulaznih podataka u željene izlazne rezultate na osnovu posebnih izračuna“ (OSCE SAIFE 2020, str. 27). Tehnološka kompanija IBM, koja postoji više od 100 godina, definiše UI kao „umjetnu inteligenciju [koja] koristi računare i mašine za oponašanje sposobnosti ljudskog uma radi rješavanja problema i donošenja odluka“ (IBM Cloud Education 2020).

Ukratko, algoritmi su gradivni elementi i samim tim sastavni dio UI. Neki algoritmi imaju mogućnost tzv. računarskog ili mašinskog učenja. Ovaj vid (polu)autonomnog učenja može se opisati kao razvojna tehnika umjetne inteligencije osmišljena radi unapređenja kvaliteta automatizovanog odlučivanja na osnovu prepoznavanja obrazaca i „pravilnosti“ kako bi se određeni zadaci obavljali bez intervencije čovjeka - drugim riječima, mogućnost „učenja bez eksplicitnog programiranja“ (Privacy International i Article 19 2018, str. 7). Ovaj proces je također poznat pod nazivom mašinsko učenje ili „grana umjetne inteligencije u kojoj računar generiše pravila i predviđanja na osnovu neobrađenih podataka koji se u njega unesu“ (Onuoha i Nucera 2018, str. 8).

Međutim, izraz *učenje* je pogrešan pošto „računar uspijeva pronaći sličnosti i razlike u podacima zahvaljujući repetitivnom podešavanju svojih parametara“ (Leslie et al. 2021, str. 8). Mašinsko učenje je zasnovano na strukturama klasifikatora koji mašinama omogućavaju donošenje skupa pretpostavki zbog čega su podložene greškama. Algoritmi za mašinsko učenje su sposobni za izvršavanje repetitivnih računskih zadataka, ali ne mogu ispravljati greške u

³Izraz „veliki skupovi podataka“ podrazumijeva skupove podataka koje tehnološke kompanije prikupljaju i kreiraju za potrebe sistema UI. Sistemima UI su za trening i generisanje modela UI neophodne stotine hiljada tačaka podataka u skupovima podataka.

svojim pretpostavkama bez ljudske intervencije. Njihova repetitivna priroda, dakle, odražava sposobnost mehaničkog ponavljanja, a ne razumijevanja informacija (Leslie et al. 2021, str. 8-10).

Ova autonomna i pogrešna percepcija inteligencije nalik ljudskoj izazvala je ogromnu pažnju javnosti i stručnjaka što je ove računske procese dodatno mistifikovalo. Bez obzira na ovaj veo mističnosti, UI nije neutralan nego aktivan društveni činilac (EK 2020, str. 24). UI su napisali, stvorili i programirali ljudi. Računski kod i baze podataka koje pokreću algoritme, a koji su gradivni elementi UI, odraz su ljudskih vrijednosti, subjektivnosti, potreba i želja. Inženjeri koji stvaraju te sisteme su akteri situirani u mrežama bogatih i moćnih korporacija i giganata društvenih medija koji te sisteme iskorištavaju radi vlastite koristi nauštrb demokratije u zemlji. Dakle, sama srž umjetne inteligencije je protkana ljudskim vrijednostima, subjektivnošću, potrebama i željama – na nivou pojedinca (inženjera) i na organizacionom i institucionalnom nivou (Mazzoli i Tambini 2020, str. 30).

Proces razvoja UI odraz je hiljada pojedinačnih subjektivnih mišljenja situiranih u okviru inicijativa kompanija koje ih zapošljavaju. Ova višedimenzionalna isprepletenost (na mikro i makro nivou) mora biti prepoznata. Pojmovi kao što su UI i algoritmi se u ovoj studiji koriste zasebno pošto je od ključnog značaja da se napravi etimološka, epistemološka i praktična razlika.

UI kroz ilustracije

Algoritmi su dio svakog računarskog sistema, oni su osnovni elementi gotovo svakog digitalnog artefakta – od sistema preporučivanja na društvenim medijima, audiovizuelnih medijskih usluga (AVM) do mnoštva aplikacija za zdravlje, fitnes i psihološku podršku. Na primjer, UI ponekad iziskuje sofisticiranije i složene algoritamske metode koje se trenutno koriste u tehnologijama kao što su UI botovi za ćaskanje (eng. chatbots) – „konverzacioni agent koji vodi dijalog sa svojim korisnikom“ (Pojmovnik UI VE), kao što je Microsoftov Tay koji tweetuje kao tinejdžer (The Verge 2016). Obrada jezika i složene operacije prevođenja – Google Perspective, koji može otkriti toksične komentare, također iziskuju sofisticiranije algoritme. To je također slučaj i kod brojnih tehnologija nadzora (pogledati: Reclaim your face, ECI 2021), kao i biotehnologija koje se trenutno koriste za kontrolu i suzbijanje priliva izbjeglica na granicama Evropske unije (Molnar 2020).

Automatizovani sistemi odlučivanja. Institut Alan Turing je 2021. godine objavio rad na temu UI, ljudskih prava, demokratije i vladavine prava u kojem su automatizovani sistemi odlučivanja (ADS) definisani kao tehnološki procesi u kojima se ljudski procesi odlučivanja unapređuju i zamjenjuju računarskim procesorima koji daju odgovore na različita pitanja poput „diskretne klasifikacije“ (npr. ženski-muški-nebinarni pol) ili procjene kao u slučaju procjene kreditne sposobnosti ili rizika od kriminala (Leslie et al. 2021, str. 36). Ovi sistemi se uglavnom oslanjaju na korištenje „treniranih“ skupova podataka koji su prethodno programirani da traže sličnu korelaciju i tačke podataka (Ackerman 2021). Često kod složenijih ADS ljudska intervencija nije neophodna, iako oni mogu automatizovati odluke i izbore koji utiču na nečiji život. Zbog nepostojanja značajnije transparentnosti, javne kontrole i odgovornosti aktera koji stvaraju ove tehnologije, osnovni problemi ADS često se mogu podvesti pod jednu od dvije dimenzije: internu i vanjsku.

Kada je riječ o *internim* problemima, kompanije često razvijaju tehnološku infrastrukturu i primjenjuju ADS ne uzimajući u obzir njihova ograničenja ili društvene ili političke implikacije. Prilikom stupanja u interakciju sa ovim sistemima koji ne prolaze nikakvu provjeru

javnost se bori da razumije tehnološke nijanse sa kojima se susreće što otežava bilo kakvu svrsishodnu intervenciju. To se može vidjeti kod sistema UI kreiranih za otkrivanje prevara koji su, umjesto da pojednostave dobijanje državne pomoći, „greškom“ to onemogućili ljudima kojima je ta pomoć najpotrebnija (Gilman 2020). Uzrok ovih internih problema leži u skupovima podataka koji se koriste za „treniranje“ ovih sistema. Skupovi podataka na osnovu kojih ADS donosi odluke su puni inherentnih „anomalija“ koje dovode do neadekvatnih ili iskrivljenih rezultata. Problem se svodi na: i. nedostatak podataka; ii. nedostatak diverzifikovanih podataka (npr. podaci o polu dati samo u binarnom obliku); i iii. netačni i nekvalitetni podaci (Kostić 2021, str. 14; za detaljnije obrazloženje pogledati: EK 2020, str. 27-30). U konačnici ovi interni problemi objašnjavaju zašto sistemi velikih tehnoloških kompanija za prepoznavanje lica imaju poteškoća sa tačnim prepoznavanjem rase i pola (Leufer 2021).

Ključno je razumjeti da su svi skupovi podataka subjektivni i da je proces razvoja ovih sistema inherentno zasnovan na vrijednostima i ciljevima ljudi koji ih razvijaju i koriste. Da bismo bili sigurni da se pri projektovanju i razvoju ovih tehnologija u obzir uzimaju i njihovi potencijalni štetni uticaji i rizici te da bismo ljudska prava i ličnu autonomiju vratili u prvi plan, moramo se zapitati: *koga ovi sistemi iznevjeravaju?*

Prema tome, za rješavanje *vanjskih* problema automatizovanog donošenja odluka neophodni su jasni, pristupačni, djelotvorni i dostupni mehanizmi pravne zaštite pomoću kojih pojedinci mogu prijaviti „probleme“ (VE 2019, str. 13, VE MSI 2018, paragraf 4.4. i 4.5). U tom smislu, koncept *čovjeka u petlji*⁴ (HITL) (EK 2020, str. 32, pogledati također: VE 2019, str. 9-10) odnosi se na potrebu za ljudskom intervencijom u ovim automatizovanim procesima koja će imati nadzornu i korektivnu funkciju kako bi se garantovalo stvarno poštovanje i zaštita ljudskih prava pojedinca. Međutim, ljudi nisu homogeni i različiti ljudi također imaju različite potrebe i iskustva.

Stoga drugi važan koncept, a to je *društvo u petlji*⁵ (Rahwan 2017, str. 3; VE 2019, str. 11), predlaže naglašavanje vrijednosti i interesa različitih društvenih grupa, etičkih principa i principa ljudskih prava, kao i mogućnosti učešća u projektovanju, razvoju (interna dimenzija), uvođenju i dostupnosti mehanizama pravne zaštite (vanjska dimenzija) kod automatizovanih sistema donošenja odluka (Zaključci sa konferencije VE 2021, paragraf 8). Ako su pravilno razvijeni i projektovani, sistemi automatizovanog odlučivanja mogu imati pozitivan uticaj. Na primjer, uvođenje sistema preporučivanja koji pomno ističu sadržaj koji promovira mir i različitost.

Odgovorna UI. Rješavanje problema štetnih uticaja na pojedince i grupe u suštini podrazumijeva ispitivanje „vrijednosti ugrađenih u algoritme koje treba propitivati, kritikovati i osporavati. U stvari, algoritmi sami po sebi nisu ti koje treba pažljivo ispitivati u smislu uticaja na ljudska prava nego su to procesi odlučivanja na osnovu tih algoritama“ (Wagner et al. 2018, str. 8). Sve je veći broj inicijativa i aktera koji učestvuju u procesu razumijevanja i suzbijanja ovih štetnih uticaja što je dovelo do pojave jedne nove oblasti: odgovorne UI. Ova oblast nudi

⁴Napomena: „čovjek u petlji“ (eng. human in the loop) (HITL) odnosi se na mogućnost ljudske intervencije u svakom ciklusu donošenja odluka u sistemu, što u većini slučajeva nije ni moguće ni poželjno. HOTL se odnosi na mogućnost ljudske intervencije tokom projektovanja sistema i praćenja njegovog funkcionisanja. HIC se odnosi na mogućnost nadzora ukupne aktivnosti sistema UI (uključujući njegove šire ekonomske, društvene, pravne i etičke uticaje), kao i na mogućnost odlučivanja o tome kada i kako će se ovaj sistem koristiti u svakom pojedinačnom slučaju“ (EK 2020, str. 32).

⁵Društvo u petlji je pojam koji se nadovezuje na uobičajeni pojam u oblasti UI „čovjek u petlji“ koji podrazumijeva da je čovjek integrisan u sistem UI kao njegov dio sa funkcijom treninga UI, analize integriteta podataka ili pomoći u usavršavanju modela UI (Appen 2019).

smjernice za projektovanje, razvoj i korištenje UI u skladu sa etičkim standardima i standardima ljudskih prava (npr. Udruženje za unapređenje umjetne inteligencije (AAAI)). Ovaj okvir obuhvata skup principa i zahtjeva koji u prvi plan stavljaju ljudsku odgovornost za sve odluke koje utiču na prava i slobode pojedinaca u predloženom okviru društvene i pravne odgovornosti. Međutim, to je lakše reći nego uraditi. Projektovanje i primjena odgovorne UI podrazumijeva „suočavanje sa nesavršenošću i prihvatanje tenzija i dilema koje se mogu javiti u slučaju kada ne postoji očigledan i široko prihvaćen stav o tome šta je ispravno“ (Rakova et al. 2021, str. 3).

U suštini, integracija pojmova odgovornosti u UI iziskuje učešće različitih aktera i njihovo djelovanje, a ne samo tehnološka rješenja. U kontekstu regulatornog okvira u oblasti UI svjedoci smo multidisciplinarnu saradnju i intervencije različitih aktera, o čemu se ukratko govori u poglavlju V. Regulatorna i politika: kratki pregled . Da bi se razumjela potreba za ovakvom vrstom saradnje i intervencije, u narednom poglavlju objašnjeni su ključni koncepti koji se dotiču ljudskih prava i moći djelovanja pojedinca.

III. UI i svakodnevni život: ključni koncepti

Umjetna inteligencija i algoritmi prisutni su u gotovo svakom aspektu našeg života – oni utiču na način na koji pretražujemo informacije na internetu, vrstu sadržaja koji vidimo na društvenim medijima, medijske i zabavne sadržaje koje konzumiramo te tržišta koja koristimo. Utkana u potku naše stvarnosti, UI je jedan od ključnih pokretača promjene i „faktor koji određuje budućnost čovječanstva budući da će znatno promijeniti život ljudi i uticati na ljudske zajednice“ (PACE 2020, paragraf 1). U nastavku su obrađeni ključni koncepti da bi se razumjelo kako njihova sveprisutnost i naša sve veća zavisnost od ovih sistema utiče i oblikuje ličnu autonomiju pojedinca.

3.1. Digitalna moć

Dok se takmiče za naš besplatan rad u vidu korisničkih podataka – interakcija, angažman, vrijeme i obrasci ponašanja (poznato i kao „ekonomija pažnje“ ili „ekonomija lajka“) (Gerlitz et al. 2013) – sistemi koje ove kompanije stvaraju i oblikuju donose važne odluke o tome koje informacije će biti vidljive i kome. Strukturirajući vidljivost naše digitalne stvarnosti algoritmi i UI postaju „čuvari informacija“ i transformišu velike tehnološke kompanije koje su postale „nadzornici interneta“ (Gillespie 2018). Dajući prednost profitu u odnosu na ljudska prava i moć djelovanja pojedinca (PACE 2020, paragraf 6; Maréchal i Biddle 2020) ove kompanije zarađuju na iskorištavanju i izvlačenju ličnih podataka i obrazaca ponašanja. Oni prisvajaju „ljudski život kroz rad sa podacima kako bi uredili ekonomski i društveni život u cjelini“ postajući tako pokretačka sila digitalne moći (Couldry i Mejias 2019, str. 33).

U ovoj studiji digitalna moć se odnosi na koncentraciju podataka, informacija i uticaja kojim raspolaže mali broj tehnoloških kompanija u kombinaciji sa rastućom zavisnošću pojedinaca, društava i institucija o njihovim uslugama i nedostatkom djelotvorne demokratske kontrole (PACE 2020, paragraf 14.6; Yeung 2019, str. 38). Digitalna moć koju posjeduju omogućava kompanijama da steknu ne samo ekonomsku nego i političku moć pošto su „platforme društvenih medija ugrađene u složene strukture upravljanja i odnose odgovornosti u koje su uključeni brojni različiti akteri: ne samo državni organi nego i proaktivni korisnici, akteri iz civilnog društva i komercijalni partneri mogu ih motivisati da utiču na tok sadržaja“ (Leerssen 2020, str. 10).

Drugim riječima, digitalna moć velikih tehnoloških kompanija nije samo stvar ekonomije pažnje nego i „politike pažnje“ (ibid., str. 11). Ova ekonomska i politička moć je neravnomjerno raspoređena što dodatno marginalizuje digitalne potrebe zemalja poput Bosne i Hercegovine. Zbog malog broja stanovnika, neodgovarajuće informaciono-tehnološke infrastrukture i/ili male kupovne moći, Bosna i Hercegovina nije imala priliku da aktivno učestvuje u diskusiji o globalizaciji interneta. Stoga, bez obzira na široku upotrebu i oslanjanje na usluge društvenih medija, kompanije koje proizvode te tehnologije nemaju svoje predstavnike u ovoj zemlji (ili na Zapadnom Balkanu) pa ih pojedinci, novinari ili drugi akteri koji se suoče sa problemima vezanim za UI i algoritamske sisteme ne mogu kontaktirati kao bi „riješili“ problem (Kostić (u pripremi)).

Digitalna moć i diskriminacija

U narednim primjerima pokazana je koncentracija digitalne moći nad online komunikacijom i interakcijama radi ilustracije njihovog diskriminatornog postupanja prema:

- određenom sadržaju u odnosu na druge (huškački govor je prisutniji pošto privlači više pažnje i veći angažman korisnika, The Markup 2021);
- određenim vidovima govora u odnosu na druge (u skladu sa izuzetkom koji se odnosi na sadržaj vrijedan objavljivanja, a radi se o standardu koji je uveo Facebook kako bi se političkim akterima (i njihovom govoru) omogućilo da ostanu na internetu protivno njihovim internim uslovima pružanja usluga (Ohlheiser 2019; Klonick 2017, str. 1665);
- određenim jezicima (postoji sve više dokaza da je uklanjanje govora mržnje „efikasnije“ ako se radi o engleskom jeziku, Mozilla 2021, str. 19);
- određenim društvima i zemljama u odnosu na druge (odlučivanjem i regulatornim pitanjima diktiraju Sjedinjene Američke Države i EU).

Iz prethodnih primjera i kratkog obrazloženja digitalne moći UI može se vidjeti kako – kroz takozvani „mrežni efekat“ (Evropska inicijativa za digitalna prava 2020) – UI prebacuje moć sa državnih predstavnika i građana na privatne i neizabrane aktore (Kalluri 2020, str. 169). U Bosni i Hercegovini ovu promjenu u moći treba posmatrati kroz prizmu složenog medijskog ambijenta (EK 2020, str. 5, 17), naročito imajući u vidu opštu pasivnost ljudi kada je riječ o konzumiranju digitalnog sadržaja. Da bi se ova dinamika moći promijenila i oslobodio neograničeni potencijal javnog učešća u digitalnoj sferi, okviri odgovorne UI, uključujući intervencije medijske i informacijske pismenosti, moraju zaštititi i garantovati autonomiju pojedinca i slobodu izražavanja.

3.2. Uticaj UI na donošenje odluka ljudi

Kao što je prethodno rečeno, struktura našeg digitalnog društva zavisi od UI i algoritama. Ovi sistemi obrađuju naše lične podatke i obrasce ponašanja na internetu radi davanja procjena i generisanja rezultata. Od planiranja nečijeg režima mršavljenja preko pronalaženja jeftinijih proizvoda pa do složenog softvera za prepoznavanje lica, UI i procesi zasnovani na podacima uveliko utiču na naše odluke i izbore, a samim tim i na našu ličnu autonomiju - stepen aktivnog učešća u donošenju vlastitih odluka. Iako je prostor za ostvarivanje lične autonomije oduvijek bio pod uticajem različitih političkih, ekonomskih i društvenih faktora, danas je taj „prostor“ zasićen „automatizovanim činocima koji imaju moć da oblikuju kontekste u kojima ljudski činoci donose odluke“ (Couldry i Mejias 2019, str. 183). Jednostavno rečeno, kada odluke UI utiču na proces odlučivanja pojedinca, autonomija tog pojedinca biva „prenesena na“ velike tehnološke kompanije (ibid, str. 214).

Velike tehnološke kompanije i sloboda izražavanja

Kada je riječ o slobodi izražavanja, velike tehnološke kompanije bitno utiču na to šta će se reći, a šta neće; ko će biti viđen, a ko neće; kao i kojim informacijama neko može, a kojim ne može biti izložen. Kao što se vidi iz sljedećih primjera, moć kontrole koju ove kompanije imaju ograničava slobodu izražavanja:

- slobodni i pravični izbori (npr. ometanje i uplitanje u izborne procese (na primjer, pogledati slučaj firme Cambridge Analytica, The Guardian 2019);
- izgledi za mir (npr. povećanje društvene polarizacije - na primjer, analiza primjera Mijanmara, UN HRC 2018);
- lična autonomija i donošenje odluka (npr. oglašavanje i uticaj na oblikovanje mišljenja kroz masovno profiliranje, The Mark Up 2021);
- ponašanje pojedinca (npr. mračni obrasci namjerno pokušavaju izazvati određeno ponašanje krajnjih korisnika, Sinderson 2021);
- društvena kohezija (npr. uvjerljivi krivotvoreni sadržaj, diskriminacija ljudi drugačije boje kože ili autohtonih grupa, pogledati Noble 2018; PACE 2020, str. 21-23).

Među raznim rizicima koji proizlaze iz ovih sistema, profiliranje je notorno problematično. Profiliranjem tehnologije umjetne inteligencije redovno izvode lične podatke i obrasce ponašanja iz naizgled nezanimljivih podataka kako bi proizvele digitalne profile koji mogu, ali i ne moraju biti tačni. Ovaj proces koji se naziva „persuazivno profiliranje“ zasnovano na podacima (Couldry i Mejias 2019, str. 140) omogućava kompanijama da mapiraju kognitivne preferencije i ponašanja osobe kako bi informisali buduće odluke koje oni donose. Dakle, posljedične odluke (Privacy International 2017) oblikuju profili koje ovi sistemi izmišljaju radi „ciljanog, personalizovanog i često neprimjetnog uticaja na pojedince i društvene grupe što bi različite političke aktere moglo dovesti u iskušenje da to iskoriste u vlastitu korist“ (PACE 2020, paragraf 4). Iako je profiliranje relativno nov koncept u evropskom pravu o zaštiti podataka, ono je sada izričito definisano članom 4(4) Opšte uredbe EU o zaštiti podataka (GDPR) i odnosi se na automatizovanu obradu ličnih podataka radi izvođenja, dedukovanja, predviđanja ili procjene informacija o pojedincu (ili grupi) koje se koriste za analizu ili predviđanje identiteta, osobina, interesovanja i ponašanja pojedinca (ili grupe) (GDPR, čl. 4(4) i rec.71).

Netačno profiliranje i rizici koji iz toga proizilaze

Forbesov dopisnik Kalev Leetaru je 2017. godine, u svojim 30-im, zatražio informacije o sebi od Oraclea, velikog brokera podataka, nakon što je dobio ponudu za članstvo u Američkom udruženju penzionera (AARP), organizaciji u Sjedinjenim Američkim Državama za penzionere i osobe starije od 50 godina. Oracle je Leetaru dostavio dokumente i podatke o njemu, uključujući spisak kategorija koje su mu pripisali. Leetaru je utvrdio da je od ukupno 108 kategorija sa kojima ga je Oracle povezo, njih 85 (oko 78%) bilo potpuno netačno. Ova vrsta podataka koje Oracle prikuplja prodaje se mnogim drugim preduzećima i ako je ova velika količina podataka koji se prikupljaju netačna, prenamijenjeni podaci mogu naknadno izazvati brojne štetne posljedice po pojedince (Leetaru 2018).

3.3. Prenamjena podataka

Velike tehnološke kompanije predstavljaju koncentraciju digitalne moći jer posluju prema poslovnom modelu motivisanom „prikupljanjem, analizom i praćenjem neselektivnih i masovnih podataka“ (Ranking Digital Rights 2020). O efikasnosti ovog modela svjedoče i

prihodi koje ove kompanije ostvaruju od oglašavanja, a koji su najveći na svijetu. Kroz ovaj model i odgovarajuće poslovne prakse velike tehnološke kompanije su „zaključale“ logiku ovog poslovnog modela i pozicionirale se na centralno mjesto u životu ljudi, organizacija, medija, drugih kompanija i manjih preduzeća. Kako ljudi postaju sve više zavisni od usluga velikih tehnoloških kompanija proizvodnja podataka postaje decentralizovana, a njihovo prikupljanje ponovo centralizovano (Helmond 2015, str. 12).

U praksi to znači da ove velike tehnološke kompanije prikupljaju, pohranjuju i analiziraju digitalne tragove koje ljudi ostavljaju na internetu (npr. dijeljenja, lajkovi, posjećivanje i vrijeme provedeno na web stranicama i platformama društvenih medija), koji naizgled postoje samo u digitalnim prostorima. Na primjer, kada neko posjeti neku web stranicu prvi put u direktoriju pretraživača te osobe kreira se datoteka koja se zove kolačić na osnovu koje se stvara jedinstvena poveznica između te osobe i stranice koju je posjetila. Sve aktivnosti na ovoj web stranici, kao što je dodavanje proizvoda u korpu, pohranjuju se u ovim malim datotekama (kolačićima). Iako je njihova namjena da pomognu u efikasnijem kretanju kroz digitalne prostore – kao što je pamćenje korpe od prethodne posjete stranici – zahvaljujući pohranjivanju ovih aktivnosti ti kolačići postaju šema naših online tragova (Pojmovnik UI VE). Ove digitalne šeme omogućavaju kompanijama da prate i mapiraju naše ponašanje na internetu. Da bi se „kombinovali podaci o različitim jastvima“, ove digitalne šeme se često uparuju sa *offline* podacima (Couldry i Mejias 2019, str. 21).

S obzirom na nepostojanje transparentnosti, velike tehnološke kompanije su u poziciji da mogu vršiti prenamjenu, ponovo koristiti i preprodavati naše lične podatke, ad nauseam.⁶ Ove kompanije koriste svoju digitalnu moć za mapiranje naših digitalnih tragova kroz prikupljanje pojedinačnih podataka u brojnim različitim digitalnim prostorima, subjektima, platformama i procesima, kao što je praćenje radi oglašavanja (softver za praćenje koji prikuplja lične podatke o ponašanju i preferencijama korisnika radi izrade ciljanih oglasa (Maréchal i Biddle 2020, str. 25-26)) ili kolačići. Iako se prvobitno prikupljaju i obrađuju za određenu svrhu (npr. kolačići koji prikupljaju podatke radi efikasnijeg snalaženja na stranicama), ovi podaci se lako recikliraju i koriste za druge namjene. Na primjer, Cambridge Analytica i njena matična kompanija, SCL Group, prikupile su podatke o korisnicima Facebooka i prenamijenile ih kako bi ciljale pojedinačne grupe i podstakle njihovo glasačko ponašanje čime su štetno uticali na izbore u 30 različitih zemalja (Ghoshal 2018). Mogućnost prenamjene, ponovnog korištenja i preprodaje naših digitalnih tragova dovela je do pojave nekoliko problematičnih procesa kao što su brokeri podacima – kompanije koje prikupljaju, prepakuju i prodaju lične podatke vezane za različite aspekte života (Couldry i Mejias 2019, str. 52). U decembru 2013. godine, tokom saslušanja u Kongresu SAD-a, javnost je imala priliku da zaviri u svijet brokera podacima i upozna se sa rizicima povezanim s tom praksom. U svjedočenju je otkriveno da brokeri podacima prikupljaju i prodaju podatke žrtava silovanja i nasilja u porodici (ibid, pogledati također: Dixon 2013).

Diskriminatorne prakse oglašavanja

Istraživačka kuće MarkUp otkrila je da Facebook dozvoljava kompanijama, kao što je Exxon Mobil, da različitim političkim grupama na Facebooku ciljano šalju različite oglase. U ovom primjeru se javljaju dva specifična problema, jedan je to što Facebook koristi profiliranje

⁶Preporuka Odbora ministara Vijeća Evrope državam članicama o velikim podacima za kulturu, pismenost i demokratiju (CM/Rec(2017)8) prenamijenjivanje definiše kao „pronalaženje nove namjene određenom predmetu i njegovo ponovno stavljanje u upotrebu dodjeljivanjem alternativne namjene i vrijednosti, odnosno različitog formata i konteksta, što u digitalnom svijetu podrazumijeva stvaranje metapodataka i podataka“. (str. 7).

podataka za utvrđivanje političkih afiniteta korisnika na osnovu sadržaja sa kojim oni stupaju u interakciju i dijele ga, a drugi to što se kompanijama dozvoljava da ciljaju određene grupe sa određenih interfejsa koje je Facebook identifikovao (Merrill 2021). Ovaj primjer ukazuje na mogućnosti platformi da ciljaju korisnike, kao i da te ciljeve stvaraju same platforme pri čemu korisnici i ne znaju kako ih te platforme označavaju ili kategorišu.

3.4. Moć kontrole i djelovanja korisnika

Ekstrakcija podataka i nemogućnost korisnika da taj proces ekstrakcije kontrolišu preovladava u gotovo svakoj grani djelatnosti i zemlji. U industriji dizajna i tehnologije ovaj vid moći djelovanja ili „informisani pristanak“ – izjava ili jasne potvrđne radnje kojima korisnik daje dozvolu za obradu njegovih ličnih podataka (član 4(11) GDPR) – odnosi se na vrste odluka koje pojedinci mogu ili ne mogu donijeti u vezi sa svojim ličnim podacima i istorijatom digitalnog ponašanja. Često tehnologija i proizvodi daju privid kontrole ili moći djelovanja kroz male odluke koje osoba može donijeti, kao što su izgled profila na Facebooku ili sadržaj koji gledaju na Netflixu. Međutim, ovi proizvodi u stvarnosti ostavljaju malo prostora za svrsishodno djelovanje. Sadržaj koji osoba gleda na Netflixu odabire se na osnovu algoritamski priređenog spiska preferencija u gledanju. Slično tome, na Facebooku korisnici ne mogu kontrolisati koje kompanije pristupaju njihovim podacima ili koja vrsta sadržaja je vidljiva na njihovoj vremenskoj traci niti mogu uticati na to koja vrsta proizvoda im se reklamira na Amazonu ili drugim platformama za kupovinu.

Targetov ciljani nadzor

Algoritamsko donošenje odluka o sadržaju prisutno je na svim platformama e-trgovine: počevši od povećanja cijena neophodnih artikala kao što su sredstva za dezinfekciju ruku na Amazonu (Harrison 2020) do toga kako ili zašto se sadržaj prikazuje na algoritamskoj vremenskoj traci korisnika na Facebooku (pri čemu Facebookov algoritam prednost daje dijeljenju pogrešnih informacija i dezinformacija, Merrill i Oremus 2021). Još jedan zanimljiv primjer je iz 2012. godine kada je otkriveno da Target stvara ogromne baze podataka i prati podatke o svojim kupcima kako bi im slao kupone i informacije bez njihovog pristanka. Ovo na prvi pogled ne djeluje kao nešto zlonamjerno ili štetno, međutim Target je prikupljao izrazito specijalizovane informacije bez znanja i pristanka kupaca na osnovu kojih je mogao utvrditi da li su kupci trudni. New York Times je otkrio da je Target na osnovu kupljenih proizvoda kupcima mogao određivati vjerovatnoću trudnoće. Na osnovu ovih podataka kompanija je mogla predvidjeti različite faze trudnoće pa je kupcima slala personalizovane kupone za svaku od tih faza. Zahvaljujući ovom dodatnom kontekstu lakše se može razumjeti zašto su kupci smatrali da su njihovi lični podaci i informacije zloupotrijebljeni (Duhigg 2012).

3.5. Štetni uticaji algoritamskih sistema

Digitalna moć koju tehnološke kompanije imaju i nepostojanje lične autonomije ne utiču na svakoga isto. „Tehnološki zasnovano povećavanje subjektivnosti i predrasuda, kao i statistički nedostaci i greške“ (PACE 2020, paragraf 25) mogu podstaći i povećati društvenu nejednakost i diskriminatorne prakse i tako dodatno marginalizovati obespravljene zajednice i društva koja prolaze kroz poslijeratni period (Keller 2021). Kada je riječ o slobodi izražavanja na internetu, postoji bogata naučna literatura u kojoj se istražuju i razotkrivaju različiti vidovi štetnih uticaja algoritamskih sistema, kao što su intersekcionalna diskriminacija (Noble 2018), pojačana

cenzura (Cobbe 2020, str. 9), praznine u informisanju (Čaušević i Sengupta 2020), manipulacija informacijama velikih razmjera (Nikolić i Jeremić 2020) te prikrivena privatno-javna partnerstva (Feldstein 2020).

Međutim, izvan područja mrežne komunikacije, negativni uticaji algoritamskih sistema također podstiču i održavaju diskriminatorne prakse. Na primjer, softver za prepoznavanje lica koji ne prepoznaje različite boje kože (pogledati: Dio II. Osnovni rječnik UI, također pogledati projekat: Our data bodies). U nekim se slučajevima nemogućnost prepoznavanja etničke pripadnosti i/ili rase osobe može pokazati nezgodnom, na primjer u slučajevima kada se ova tehnologija koristi u zgradama za omogućavanje ulaska stanarima (The Guardian 2019). U drugim slučajevima negativan uticaj oslanjanja na sisteme koji ne mogu adekvatno identifikovati osobine osobe je očigledniji, a tu spadaju i sistemi za provođenje zakona koji ovu tehnologiju koriste u krivičnim istragama (Al-Kawaz et al. 2018). U drugim slučajevima opet – društvene nejednakosti i diskriminatorne prakse – ovi sistemi su ontološki poput socio-tehničkog dizajna zasnovanog na uvjerenju da mnogi aspekti našeg identiteta (rasa, etnička pripadnost, pol, seksualnost itd.) predstavljaju statične i mjerljive varijable (pogledati primjer u nastavku).

Automatizovanje pola na osnovu šarenice oka

Institut inženjera elektrotehnike i elektronike (IEEE) je 2007. godine objavio studiju koju su finansirali UNISYS Corp i nekoliko američkih državnih agencija, uključujući Centralnu obavještajnu agenciju (CIA), u kojoj se tvrdi da modeli mašinskog učenja mogu odrediti pol osobe na osnovu teksture njihove šarenice (Thomas et al. 2007). Ovi modeli su trenirani pomoću slika očiju učesnika u studiji. Na primjeru jedne takve slike date u članku vidi se da na slikama nije prikazana samo šarenica učesnika nego cijelo oko (uključujući gornje i donje trepavice) (ibid, str. 3). U ovoj studiji, koju je provela grupa računarskih stručnjaka i inženjera muškog pola sa Univerziteta Notre Dame, govori se o „modelima rodne klasifikacije tačnosti od blizu 80%” (ibid, str. 5). U ovoj i još nekoliko drugih studija (Tapia et al. 2016; Fairhurst et al. 2015; Tapia et al. 2014, Bansal et al. 2012; Lagree i Bowyer 2011) tvrđeno je da algoritmi mašinskog učenja mogu odrediti pol osobe. Osim toga što predstavlja korak unazad u shvatanju da se rod kao društveni konstrukt ne može kvantifikovati, ova tvrdnja je zasnovana na poimanju sistema umjetne inteligencije kao neutralnih aktera.

Deset godina kasnije IEEE je objavio drugu studiju koju je (opet) provela grupa naučnika muškog pola, od kojih je jedan autor prethodno citirane studije, a druga dvojica su sa istog odsjeka i iz iste institucije (Kuehlkamp et al. 2017). Ova studija iz 2017. se razlikovala po tome što je u njoj priznato da raniji algoritmi za određivanje pola nisu bili uspješni u prepoznavanju efekata poput šminke. Koristeći isti skup podataka koji je korišten u studiji iz 2007. godine ovi naučnici su istrenirali algoritme da uzmu u obzir efekte šminke i utvrdili da je zahvaljujući slikama na kojima je prikazana šminka postignuta „veća tačnost u određivanju pola na osnovu šarenice“. Zapravo, ova studija je pokazala da ovi algoritmi nisu u stanju da odrede pol nego da prepoznaju prisustvo šminke.

IV. Pregled: UI i sloboda izražavanja

Funkcionalna demokratija podrazumijeva otvorenu, živu i nesmetanu javnu raspravu i slobodan protok informacija i iz tog razloga sloboda izražavanja i sloboda medija predstavljaju žilu kucavicu demokratije. Međutim, sve veća dominacija i uticaj velikih tehnoloških

kompanija, potpomognutih algoritmima i UI, mijenjaju način na koji izražavamo svoje stavove, primamo informacije i učestvujemo u javnoj raspravi. Ova radikalna transformacija „odslikava moć i domet kojim ove privatne platforme raspoložu zahvaljujući svojim sistemima moderacije i daje težinu njihovoj ulozi u demokratskoj kulturi“ (Klonick 2017, str. 1663) pa je zato razumijevanje ovih transformativnih procesa i njihove dinamike od ključnog značaja. U ovom poglavlju analiziramo bitne karakteristike ove dinamike prvo kroz sagledavanje odnosa između UI, algoritama i slobode izražavanja, a potom kroz analizu uticaja ovih sistema na slobodu medija.

4.1. Moderacija sadržaja

Moderacija sadržaja pomoću UI i algoritama – krovni pojam za procese praćenja, procjene, odabira i distribucije sadržaja (Bukovska 2020, str. 32-34) zasnovane na praksama profiliranja i nadzora – predstavlja veliki rizik po sposobnost pojedinca da formira informirane stavove (Leslie et al. 2021, str. 14). Radi razumijevanja rizika, obima i složenosti ovih procesa, u narednim paragrafima prikazan je (u pojednostavljenoj verziji) ciklus moderacije sadržaja.

Svaki dio sadržaja na platformama društvenih medija prolazi kroz tri različita nivoa moderacije, naročito sadržaj koji generišu korisnici, odnosno digitalni mediji koje proizvode i distribuiraju pojedinačni korisnici (Jenkins 2006; OSCE SAIFE 2020, str. 3).

Nivo 1: Postavljanje sadržaja na internet

Prije nego što neki sadržaj bude objavljen na internetu skup automatskih filtera za „postavljanje sadržaja“ analizira dati materijal kako bi se utvrdilo da li je on na spisku sadržaja čije je objavljivanje na internetu zabranjeno u skladu sa uslovima korištenja i internim dokumentima platforme (npr. sprečavanje dječije pornografije ili kršenja autorskih prava).

Nivo 2: Čuvari interneta

Sadržaj koji prođe prvi nivo moderacije zatim procjenjuju UI i algoritmi. Ovi automatizovani procesi osmišljeni su za rangiranje, optimizaciju i preporuku sadržaja na osnovu niza kriterija i podataka (npr. kategorija datog sadržaja, lični profili i preferencije publike). U ovoj fazi automatski procesi donose odluke o tome koliko će taj sadržaj biti vidljiv djelujući tako kao arhitekta „izloženosti“. S obzirom na to da o ovim sistemima rangiranja i preporučivanja zasnovanim na UI nema skoro nikakvih informacija, oni su neprobojni i često se nazivaju crnom kutijom (Pasquale 2015).

Nivo 3: Biti ili ne biti

Nakon što sadržaj bude objavljen na internetu, on živi u skladu sa skupom procesa i praksi kojima „upravljaju“ mehanizmi prijavljivanja. Ovi mehanizmi prijavljivanja su definisani u skladu sa uslovima korištenja platforme i pojedinačnim korisnicima omogućavaju da prijave, označe ili blokiraju kategorije sadržaja koje su unaprijed definisane u internim dokumentima platforme. Ako je sadržaj označen, pokreće se mehanizam prijavljivanja u okviru kojeg se na osnovu kombinacije procjena UI i ljudskih procjena donose odluke o prijavi tako da prijavljeni sadržaj (uključujući lične naloge) može biti blokiran, uklonjen ili na neki drugi način sankcionisan.

(Ne)efikasni mehanizmi prijavljivanja i iskustva Zapadnog Balkana

Bez obzira na postojanje ova tri nivoa sistema moderacije sadržaja, sadržaj koji širi mržnju i štetan sadržaj još uvijek je široko rasprostranjen u digitalnoj sferi. Na primjer, u jednom

istraživanju koje je provela organizacija Balkan Insight Reporting Network utvrđeno je da gotovo polovina prijavljenog sadržaja proizvedenog na Zapadnom Balkanu ostaje na internetu (Jeremić et al. 2021). U ovom istraživanju također je utvrđeno da je manje vjerojatno da će problematičan sadržaj koji su označile žene biti uklonjen (62% u odnosu na 38% za muškarce, ibid). Ovi mehanizmi prijavljivanja, koji se često nazivaju mehanizmima zaštite, moraju biti predmet kritike zbog svoje neefikasnosti, netransparentnosti i proizvoljnosti. Naš odnos zavisnosti sa ovim tehnologijama treba ponovo procijeniti. Ljudska intervencija u ovom procesu moderacije sadržaja nije uvijek moguća pa tako problem koji UI i algoritmi uzrokuju na jednom nivou postaje automatizovan na drugom (Bukovska 2020, str. 3).

4.2. Priredivanje i dijeljenje sadržaja

Na drugom nivou moderacije sadržaja digitalna moć velikih tehnoloških kompanija postaje očigledna. Njihova sposobnost globalnog kontrolisanja cirkulacije digitalnog sadržaja često se naziva priredivanje i dijeljenje sadržaja (Gorwa et al. 2020, str. 3; Bukovska 2020, str. 19). Na primjer, pregled vijesti (eng. newsfeed) na Facebooku proizvodi se na osnovu obrade šablona digitalnog ponašanja osobe (vrijeme potrošeno na sličan komentar, prethodni angažman, lični i grupni profili koji služe kao posrednici) radi optimizacije i personalizacije svih sadržaja (Constine 2016; Bernstein et al. 2020, str. 47) – zbog čega ne postoje dva identična profila ili pregleda vijesti na Facebooku.⁷

UI i algoritamski sistemi preporuka donose automatizovane i personalizovane odluke o tome šta krajnji korisnik vidi na svom profilu. S obzirom na količinu digitalnog sadržaja i brzinu kojom se on postavlja na internet, priredivanje i dijeljenje sadržaja pomoću UI i algoritama čini se nužnim. Iz funkcionalne perspektive ovi sistemi omogućavaju ljudima kretanje kroz šumu informacija. Na primjer, jednostavno pretraživanje interneta bi bilo veoma komplikovano da nije poznat istorijat pretraživanja te osobe (Bodó et al. 2019, str. 207). Dakle, automatizovani sistemi priredivanja i dijeljenja sadržaja nisu samo dio problema nego i neprocjenjivi elementi digitalne pokretne trake (Burri et al. 2020, str. 42).

Procesi priredivanja, dijeljenja i preporučivanja sadržaja nisu zasnovani isključivo na UI i algoritmima nego i na poslovnim modelima i logici profita. Velike tehnološke kompanije koje uvode ove sisteme se finansijski održavaju tako što se trude da ljudi provode više vremena na njihovim platformama kako bi maksimalno povećali njihovu izloženost reklamama. Pri tome postoji inherentni rizik da će sistemi preporučivanja pretežno izlagati ljude onom sadržaju koji će najvjerojatnije privući njihovu pažnju – također poznato kao fenomen filter balona (Bodó et al. 2019, str. 209; Kaluža 2021, str. 5). Ovaj vid optimizacije sadržaja pojavio se još 2009. godine kada je Google omogućio personalizovano pretraživanje na osnovu tačaka podataka kao što su lokacija, istorijat pretraživanja itd. (Kaluža 2021, str. 3). Sistemi preporučivanja od tada privlače pažnju stručnjaka, prvenstveno zbog straha da će ljudi „završiti“ zarobljeni u eho-komorama – komunicirati sa istomišljenicima, što može podstaći nasilje i ekstremizam te tako iscjepkati javnu sferu na homogene i izolovane informacione klastere (Burri et al. 2020, str. 43).

Internetski filter baloni se često doživljavaju kao jedan dio složene dinamike koja oblikuje informativni prostor pojedinca. U jednoj studiji provedenoj u Holandiji zaključeno je da ne postoje ozbiljni razlozi za zabrinutost da će ljudi u ovoj zemlji biti zarobljeni u filter balon

⁷Objašnjenje pogledati ovdje: <https://bit.ly/3OuTZbr>.

(Möller et al. 2019). Međutim, druga studija koja je dovela u pitanje ograničeni potencijal internetskog filter balona kao koncepta bavila se algoritamskom personalizacijom sadržaja i uticajem priređivanja i dijeljenja sadržaja na medijsku raznovrsnost. Ova studija je pokazala da određene kategorije ljudi, kao što su mladi i stariji, jesu u opasnosti da budu izloženi pretežno algoritamski biranom sadržaju (Bodó et al. 2019, str. 219). Ovi nalazi se poklapaju sa prethodno pomenutom studijom u Bosni i Hercegovini (pogledati: Poglavlje I. Uvod) koja je pokazala da kod mladih ljudi u ovoj zemlji također postoji veća vjerovatnoća da će konzumirati sadržaj koji preporučuju algoritmi (Hodžić 2019, str. 32, 34).

Problem vezan za internetski filter balon na koji se pokušava ukazati i koji se nastoji objasniti prevazilazi dualističko shvatanje interakcije između korisnika i algoritama. Proces konzumacije informacija je višedimenzionalan – ljudi koriste različite načine pristupa informacijama i aktivno pretražuju sadržaj koji nije na platformama i digitalnim tehnologijama. Osim algoritama, postoje i drugi vanjski društveno-politički i ekonomski faktori koji ograničavaju informacioni prostor pojedinca (Bodó et al. 2019, str. 218; Mazzoli i Tambini 2020, str. 29). Ovi argumenti ne negiraju moć sistema algoritamske personalizacije i preporučivanja koji upravljaju priređivanjem i dijeljenjem sadržaja na globalnom nivou. Zabrinutost vezana za te procese ostaje legitimna pošto je teško predvidjeti kakve će posljedice oni imati na pojedince i društva te njihovo učešće u javnoj sferi (Kaluža 2021, str. 14).

Kada je riječ o platformama za dijeljenje i streaming audiovizuelnog sadržaja, ovi pružatelji usluga se oslanjaju na sisteme preporuka pretežno radi „ocjenjivanja“ i „predviđanja“ preferencija ljudi prije „isporuke“ optimiziranog audiovizuelnog sadržaja (Burri et al. 2020, str. 23). Za razliku od sistema za preporučivanje vijesti koje karakteriše priliv i količina informacija – što iziskuje drugačiji skup metrike algoritamskog odabira – sistemi preporučivanja koje koriste pružatelji usluga audiovizuelnih medija u suštini korisnicima pomažu da bez prevelikog napora pronađu relevantan sadržaj pa stoga obilje i različitost informacija ima drugačiju ulogu u ova dva domena preporučivanja (Bernstein et al. 2020, str. 49). Međutim, logika funkcionisanja UI na kojoj počivaju oba ova domena je u suštini slična.

4.3. Govor mržnje i UI

Sve veća primjena automatizovanih sistema u kontroli i uklanjanju sadržaja koji širi mržnju i štetnog sadržaja pokazala se delikatnom. S obzirom na to da sistemi UI ne prepoznaju kontekst, oni nisu u stanju da razlikuju sarkazam od govora mržnje. Pored toga, automatizovani sistemi preporučivanja su agilni i stalno se usavršavaju, kao i priroda jezika i online interakcija (Yeung 2019, str. 29). Problem se u suštini svodi na rezultate koje ti automatizovani sistemi daju. Pored toga što u nekim slučajevima uklanjaju legitiman sadržaj (koji se naziva lažno negativnim), ovi sistemi također prave propuste u uklanjanju „stvarnog“ govora mržnje (lažno pozitivan) (Bukovska 2020, str. 56). S obzirom na globalnu primjenu ovih sistema za uklanjanje sadržaja, njihova nepouzdanost je pokazatelj razmjera i potencijala koji oni imaju da ograniče individualnu slobodu izražavanja na internetu (Yeung 2019, str. 30; pogledati također: Wagner et al. 2018, str. 17; OSCE SAIFE 2021, str. 17).

Pored ovih problema tehnološke prirode, još uvijek ne postoji univerzalna definicija govora mržnje, a postojeće informacije se često različito tumače i razumiju u različitim dijelovima svijeta. Pored toga, kada je riječ o problemima koji proizilaze iz korištenja automatizovanih sistema za uklanjanje sadržaja koji širi mržnju (npr. Google Perspective), u postupku razvoja ovih automatizovanih sistema često se previđaju dodatni slojevi složenosti lokalnog konteksta. Konačno, odgovori na široko rasprostranjeni govor mržnje i nasilje u digitalnom prostoru ne

bi trebali biti samo tehnološke prirode nego iziskuju međusektorsku saradnju i strateški odgovor većeg broja aktera, uključujući intervencije u oblasti MIL-a.

Algoritmi i uništavanje reputacije

U društvima koja prolaze kroz poslijeratni period, poput bosanskohercegovačkog društva, govor mržnje i diskriminatorne prakse su značajno prisutni u javnoj debati i narativima o miru. Skorašnji online napad – poznat i kao „uništavanje reputacije“ (Turčilo et al. (u pripremi)) – na dvije ugledne žene i kritičke mislioce u BiH (Safetyofjournalists.net, 2021) oslobodio je negativni potencijal društvenih medija i online prostora i završio se prijetnjama i uznemiravanjem i izvan interneta. Problem nije samo u govoru mržnje i njegovom uticaju na pojedince nego i u infrastrukturi medija i logici UI koji omogućavaju širenje ovakvih vidova govora i online ponašanja. Dakle, digitalna moć velikih tehnoloških kompanija je dovela do pojave novih vidova orkestriranih negativnih uticaja i javnog iskrivljivanja pri čemu te kompanije za sada ne pokazuju preveliko interesovanje da se bore sa široko rasprostranjenim i višestrukim vidovima govora mržnje usmjerenog na pojedince, etničke grupe i kritičke mislioce u Bosni i Hercegovini.

4.4. Uvjerljiv krivotvoreni sadržaj (eng. deepfakes)

Uvjerljiv krivotvoreni sadržaj se najbolje može opisati kao video i audio materijal kojeg generiše UI i kojim se stvara „alternativna realnost“ ili „sintetički medij“ u kojem izgleda da neko govori ili radi nešto što nije rekao ili uradio. U svom najprostijem obliku, uvjerljiv krivotvoreni sadržaj se stvara pomoću računarskog programiranja i velikih skupova podataka koji se „hrane“ slikama i zvukom osobe ili predmeta koji bi ovaj sadržaj trebao da imitira (Burri et al. 2020, str. 147). Zbog mogućnosti njihovog iskorištavanja i zloupotrebe, ovakvi sadržaji imaju brojne negativne uticaje – od stvaranja potencijalne najezde netačnih informacija i dezinformacija povezanih sa medijima do izazivanja rodno zasnovanog uznemiravanja ljudi na internetu.

DeepNude

Otkriveno je da se DeepNude može koristiti za postavljanje lica žena u pornografske materijale. Ovakva vrsta štete koju krivotvoreni sadržaji izazivaju tretira se kao vid seksualnog iznuđivanja (Hao 2021). Krivotvoreni sadržaj se može koristiti za dijeljenje slika stvarnih ljudi bez njihovog pristanka (što se ponekad kolokvijalno naziva pornografija iz osvete) i mogu se koristiti kako bi se ljudi ubijedili da povjeruju da je neki video materijal stvaran (Burri et al. 2020, str. 147).

Imitator Toma Cruisea

U viralnom videu na Twitteru i TikToku osoba koja imitira Tom Cruisea je pomoću tehnologije stvaranja uvjerljivog krivotvorenog sadržaja postigla gotovo savršenu sličnost sa Tomom Cruiseom. Ovaj video je naširoko dijeljen pošto se vjerovalo da se radi o pravom glumcu Tomu Cruiseu, ali je nekoliko sati kasnije (Metz 2021) razotkriveno da je u pitanju krivotvoreni sadržaj. Pored toga, postoje javne potrošačke aplikacije koje omogućavaju jednostavnu „zamjenu lica“ pomoću tehnologije stvaranja krivotvorenog sadržaja (Zucconi 2018). Iz ovih brojnih primjera vidimo negativan uticaj tehnologije kreiranja uvjerljivog krivotvorenog sadržaja koja se koristi za generisanje netačnih informacija i dezinformacija, uključujući uzimanje identiteta stvarnih osoba i kršenje privatnosti i moći djelovanja korisnika.

4.5. Perspektive slobode medija

U narednim paragrafima se, kroz prizmu digitalne moći, analizira složeni odnos između velikih tehnoloških kompanija i slobode medija u kontekstu Bosne i Hercegovine. Medijski prostor u Bosni i Hercegovini, kojeg karakteriše nedostatak medijskog pluralizma, oblikovan je relativno velikim brojem medijskih kuća koje dijele malo tržište oglašavanja i koje sve više zavise od društvenih medija kako bi doprle do svoje publike i ostvarile profit (Madoleva 2021; Kostić (u pripremi)). Zapravo veliki broj medijskih kuća u potpunosti zavisi od društvenih medija po pitanju distribucije svog sadržaja čime se platforme društvenih medija pozicioniraju kao moćan, ali uglavnom „nevidljiv“ akter u medijskom prostoru. Ova „nevidljivost“ u stvarnosti znači da su iz „ekonomskih i ideoloških razloga pretraživači i kompanije društvenih medija odlučili da ostanu neopredijeljeni po pitanju sadržaja u smislu filtriranja poruka koje bi mogle imati negativan uticaj na javnost“ (Donovan i Boyd 2019, str. 6).

4.5.1. Medijski pluralizam. Novinarstvo je praksa koja se stalno mijenja, a medijske kuće se još uvijek bore da očuvaju svoju društvenu relevantnost i povjerenje javnosti te da ojačaju svoju poziciju i pregovaračku moć u odnosu na velike tehnološke kompanije. Iz tog razloga, medijski pluralizam – pluralizam izvora, sadržaja i izloženosti, (Mazzoli i Tambini 2020, str. 40) – je radikalno izmijenjen. Ta transformacija podrazumijeva da ove kompanije kontrolišu procese distribucije i moderacije vijesti što dovodi do pojave takozvanog fenomena „zaključanosti“ (Burri et al. 2020, str. 47) koji medijskim akterima na globalnom nivou oduzima pregovaračku moć.

Pored toga, prakse moderacije sadržaja ne samo da su netransparentne i da isključuju medijske agencije nego su i dinamične i agilne te medijskim kućama neprestano oduzimaju prilike da dopru do publike i samim tim smanjuju njihove prihode. U Bosni i Hercegovini, gdje se medijska industrija suočava sa teškom finansijskom situacijom, a nezavisni mediji su izloženi dodatnom političkom pritisku i kontroli (EK 2021, str. 23), manjak prihoda putem društvenih medija predstavlja ozbiljan „udar“ na njihovu finansijsku stabilnost i medijski pluralizam.

Demokratija, laboratorija društvenih medija

Facebook je 2017. godine restrukturisao algoritme svog *NewsFeeda* kako bi povećao vidljivost sadržaja koji dijele i stvaraju prijatelji i oglasa. Sadržaj koji proizvode mali nezavisni časopisi je, zbog ovog vida rekonfiguracije, postao manje vidljiv korisnicima u Slovačkoj, Šri Lanki i Srbiji (Hern 2017). Glavni urednik KRIK-a, Stevan Dojčinović je, u kolumni pisanoj za New York Times, rekao kako velike tehnološke kompanije tretiraju male zemlje kao laboratorije i istakao da: „birajući male zemlje sa krhkim demokratskim institucijama kao zamorčice u eksperimentu, pokazuju ciničnu nebrigu o tome kako njihove odluke utiču na najugroženije“ (Dojčinović 2017). Konačno, naše demokratije⁸ su postale igrališta i probni poligoni kompanija društvenih medija.

4.5.2. Kvalitetno novinarstvo. Primjenom alata kao što su dugmad za dijeljenje i lajkovanje, medijske organizacije – bez obzira na njihovu veličinu, uticaj, geografsko porijeklo i položaj u društvu – sve više usvajaju logiku kompanija društvenih medija pa smo tako svjedoci sve veće platformizacije interneta: „uspon platforme kao dominantnog infrastrukturnog i ekonomskog modela društvene mreže i posljedice širenja platformi društvenih medija na druge online prostore“ (Helmond 2015, str. 12). U ovakvom ambijentu, medijske organizacije iskorištavaju

⁸Pogledati vizuelni prikaz pozadinskih procesa: Share Lab. (2016). Immaterial Labour and Data harvesting - Facebook Algorithmic Factory.

ove intruzivne sisteme na gotovo isti način kao i velike tehnološke kompanije (npr. korištenjem ciljanog oglašavanja, sponzorisanog sadržaja ili prikupljanjem i prodajom korisničkih podataka). Vođene komercijalnim interesima, ove medijske kuće proizvode sadržaj koji komodifikuje i nastoji da zaokupi pažnju pojedinca (npr. sadržaj koji služi kao mamac za klikove (eng. clickbait) ili huškački sadržaj: pogledati: Mattu et al. 2021; Donovan i Boyd 2019, str. 3, 7). Medijske organizacije su implicitno podstaknute da proizvode takav sadržaj i žrtvuju prakse profesionalnog novinarstva kako bi zaštitile svoju finansijsku održivost (Mazzoli i Tambini 2020, str. 41). Kako je to novinar i urednik Janus Rose istakao u jednom intervjuu koji je dao za projekat „Explainable AI“ koji je finansirao OSCE: „logika platformi je suštinski suprotna onome što želimo da proizvedemo kao kvalitetan novinarski sadržaj koji se ponekad ne zasniva u potpunosti na tome šta ljudi žele da čitaju“ (Explainable AI 2020).

Izbori i farme trolova

Kao što je prethodno rečeno, funkcionalna demokratija podrazumijeva otvorenu, živu i nesmetanu javnu raspravu i slobodan protok informacija. Sposobnost zemalja u kojima se građanski i medijski prostor sužava da očuvaju demokratiju je ugrožena, kao što se to vidi iz narednog primjera.

Mjesec dana prije nego što se predsjednik Daniel Ortega kandidovao za reizbor 2021. godine Facebook je uklonio više od 1.000 naloga na Facebooku i Instagramu u pokušaju da uništi farmu trolova – grupisanu mrežu lažnih naloga korištenih za manipulaciju javnim diskursom. Prema Facebooku, ovom farmom trolova prvenstveno je upravljala nikaragvanska vlada i vladajuća stranka u zemlji, Sandinistički front nacionalnog oslobođenja (FSLN). Farma je postala aktivna u aprilu 2018. godine, nakon izbivanja studentskih protesta protiv vlade, sa ciljem da diskredituje učesnike protesta i poveća prisustvo provladinog sadržaja na više platformi društvenih medija: Facebook, TikTok, Twitter, YouTube i Telegram. Treba napomenuti da farma trolova povezana sa vladom Nikaragve nije izolovana pojava nego rastući trend. Tokom 2021. godine Facebook je uklonio nekoliko drugih naloga na društvenim medijima povezanih sa vladom i to u Etiopiji, Ugandi, Sudanu, Tajlandu i Azerbejdžanu (Culliford 2021).

Twitter je, 2020. godine, uklonio oko dvije hiljade bot (lažnih) naloga u Srbiji (The Guardian 2020). Slične manipulativne aktivnosti zabilježene su u izbornom periodu u Bosni i Hercegovini 2019. godine (Cvjetićanin 2019, str. 7, 22). Pored toga, Wall Street Journal je otkrio kako Facebook ima bijele i crne liste sadržaja, profila i stranica (O’Neil 2021), što također ukazuje na nadzorničke prakse kompanija i njihov uticaj na ono što vidimo na internetu.

U ovoj igri moći, konglomerati društvenih medija su osvajači zlatne medalje. Njihova zarada nastavlja da raste kako finansijska održivost novinskih organizacija postaje sve zavisnija od usluga koje pružaju, a oni pri tome ostaju ravnodušni prema potrebama građana i neopredijeljeni u pogledu sadržaja koji priređuju i dijele. Zemlje u kojima je situacija u pogledu slobode medija složena, poput Bosne i Hercegovine, direktno osjećaju brutalnu moć ove dinamike. Medijske organizacije u ovoj zemlji funkcionišu po poslovnom modelu dvostruke neprozirnosti. Drugim riječima, medijske kuće plaćaju velikim tehnološkim kompanijama da promovišu njihov, često nekvalitetan, sadržaj kako bi bili vidljivi i interesantni oglašivačima. Posljedično tome taj sadržaj je često pun dezinformacija i govora mržnje, što utiče na povjerenje javnosti u njih i kvalitet javne debate u zemlji (Sokol 2021, str. 20; Sokol 2020, str. 7; Cvjetićanin 2019, str. 22).

4.5.3. Medijska raznolikost. Poput algoritama Newsfeeda, sistemi za preporučivanje vijesti koriste podatke i obrasce ponašanja korisnika za optimizaciju i isporuku personalizovanog sadržaja. Ako je neko zainteresovan za čitanje sadržaja o sportu, ta osoba će vjerovatno biti izložena sličnom sportskom sadržaju što dovodi do „de facto smanjenja raznovrsnosti vijesti koje se tom korisniku nude“ (Bernstein et al. 2020, str. 52). Smanjivanjem izloženosti i mogućnosti pristupa pojedinca različitim mišljenjima, vrijednostima i narativima „prakse moderacije sadržaja utiču na javni interes, miješaju se u demokratske procese i ugrožavaju informacioni pluralizam i raznolikost“ (Helberger 2019, str. 993). Prema tome, velike tehnološke kompanije diktiraju strukturu slobode medija i često potiskuju medijsku raznolikost.

Agregatori vijesti u sukobu sa medijskom raznolikošću

U postojećem medijskom ambijentu pojedinci postaju sve zavisniji od agregatora vijesti koji prikupljaju i sastavljaju vijesti iz različitih izvora na jednom mjestu (OSCE RFoM 2021, str. 66; Newman et al. 2020, str. 11; Bernstein et al. 2020, str. 47). Iako se u sistemima agregatora vijesti koriste i ljudski i algoritamski modeli odabira, skorašnja studija je pokazala da Appleov agregator u „News Top Stories“, gdje sadržaj priređuju i dijele ljudi, nudi raznovrsniji sadržaj od sadržaja na „Trending Stories“ na istoj platformi koji se bira algoritamski (Bandy i Diakopoulos 2020, str. 43). Pored medijske raznolikosti, agregatori vijesti utiču na medijski ambijent na još dva načina: agregatori imaju tendenciju da ljude zadržavaju u okviru svog algoritamskog odabira vijesti čime im ograničavaju korištenje drugog sadržaja vijesti (također poznato kao „efekat supstitucije“) na štetu finansijske održivosti medija i protoka prihoda. Međutim, ovi agregatori vijesti su također omogućili veću vidljivost manje „popularnim medijskim kućama zahvaljujući takozvanom „efektu širenja tržišta“ (OSCE SAIFE 2021, str. 67).

4.6. Preliminarni zaključak

Prethodno analizirana pitanja generalno ilustruju mjeru do koje su UI i algoritamski sistemi vođeni profitom ukorijenjeni u naše živote. Iako zasnovani na automatizaciji, prethodno istaknuti problemi nadilaze samu automatizaciju. Couldry i Mejias u svojoj knjizi „The costs of connection. How Data Is Colonizing Human Life and Appropriating It for Capitalism“ [Cijena konekcije. Kako podaci kolonizuju život ljudi i prisvajaju ga u korist kapitalizma] otkrivaju da problem nije u jednoj platformi ili određenoj tehnologiji nego u isprepletenoj kombinaciji šest ključnih faktora: postojeća (digitalna) infrastruktura za ekstrakciju, poredak u nastajanju (što podrazumijeva podatkovne prakse koje uvlače ljude u ovu infrastrukturu), ekonomski sistem izgrađen na toj infrastrukturi i poretku, racionalnost koja daje smisao tim praksama, novi model znanja koji preobličava svijet i „sve što se može znati o ljudskom životu“ (Couldry i Mejias 2019, str. 192). Iz makro perspektive, ovih šest faktora ilustruju da su velike tehnološke kompanije u stanju da iskoriste „moć mišljenja“ ili moć uticaja na „proces formiranja individualnog i javnog mišljenja“ (Helberger 2020, str. 846). Ukratko, „ove platforme mijenjaju samu strukturu i ravnotežu medijskog tržišta i time direktno i trajno utiču na pluralističku javnu sferu“ (ibid.). Moć mišljenja neizbježno oblikuje našu slobodu izražavanja (Burri et al. 2020, str. 62) pošto mjera u kojoj smo u mogućnosti da uživamo slobodu izražavanja zavisi od vrste informacija kojima smo izloženi i mjeri u kojoj smo im izloženi.

V. Regulatoriva i politike: kratki pregled

U prethodnim dijelovima ove studije opisano je kako algoritamski sistemi zasnovani na podacima oblikuju mogućnost ljudi da primaju i prenose informacije u današnjem konvergentnom i „platformizovanom“ svijetu te se postavlja pitanje: *ko je odgovoran u situaciji kada UI utiče na slobodu izražavanja i slobodu medija?* Da bi smo odgovorili na ovo pitanje u ovom dijelu je dat kratak prikaz okvira politika vezanih za UI kroz prizmu Vijeća Evrope i Evropske unije. Treba napomenuti da ne postoji obavezujući regulatorni okvir koji reguliše zavrzlane automatizovane moderacije sadržaja. Dokumenti navedeni u nastavku teksta izvedeni su iz člana 10 Evropske konvencije o ljudskim pravima koji garantuje zaštitu slobode izražavanja i pokriva sve vidove izražavanja, uključujući izražavanje kroz algoritamske sisteme i druge komunikacione tehnologije.

Prije nego što je UI ušla u široku upotrebu, Vijeće Evrope je bilo uključeno u brojna pitanja koja su se ticala digitalnog okruženja u nastajanju, kao što je uloga i odgovornost internetskih posrednika i značaj interneta i novih medija u eri tehnološke konvergencije.⁹ Kako su sistemi UI postajali sve rašireniji Konvencija o zaštiti lica u pogledu automatske obrade ličnih podataka (u daljem tekstu Konvencija 108) je ažurirana kako bi se fokusirala na zaštitu pojedinaca i njihovih informacija u svijetu velikih podataka (T-PD(2017)01). U ovim smjernicama je istaknut značaj ljudske intervencije u automatizovanim procesima i informisanog pristanka korisnika. U istom smislu, u Smjernicama uz Konvenciju 108 o umjetnoj inteligenciji i zaštiti podataka (T-PD(2019)01) naglašeno je da se prakse obrade podatka pomoću UI moraju uskladiti sa opštim principima zaštite podataka i ponovo je istaknut značaj informisanog pristanka i moći djelovanja korisnika (pogledati: T-PD(2019)01, Dio I). U ovim smjernicama se također predlaže procjena uticaja rizika radi procjene „eventualnih negativnih posljedica primjene UI po ljudska prava i osnovne slobode“ (pogledati: T-PD(2019)01, Dio II) kako bi se garantovala ljudska intervencija u slučajevima kršenja ljudskih prava povezanih sa UI (pogledati: T-PD(2019)01, Dio III). Pored toga, u Preporuci Odbora ministara državama članicama o velikim podacima za kulturu, pismenost i demokratiju (CM/Rec(2017)8) ponovo je potvrđen značaj pristupa ljudskih prava (pogledati: CM/Rec(2017)8, Preambula) i dijaloga različitih aktera (paragraf 3) i istovremeno istaknuta važnost moći djelovanja korisnika i njihove kontrole na procesima algoritamskog donošenja odluka koji „određuju kulturna obilježja, preferencije i ponašanje“ (pogledati: CM/Rec(2017)8, Preambula).

Dok prethodno pomenute politike naglašavaju potrebu za intervencijama koje se tiču ljudskih prava i ističu izazove koje UI i algoritamski sistemi predstavljaju, Deklaracija Odbora ministara o manipulativnim sposobnostima algoritamskih procesa bavi se zloupotrebom zasnovanom na UI i njenim uticajem na društvom. U ovom dokumentu se predlaže da mehanizmi zaštite od algoritamske manipulacije i ubjeđivanja trebaju ići izvan okvira zaštite podataka i „pozabaviti se značajnim uticajima ciljane upotrebe podataka u odnosu na društvo i zaštitom ljudskih prava u širem smislu“ (CM Decl (13/02/2019)1, 9.b), kao i da se napravi jasna razlika „između dopuštenih vidova ubjeđivanja i neprihvatljive manipulacije“ (ibid., 9.c). Parlamentarna skupština Vijeća Evrope (PACE/2341) je, 2020. godine, usvojila rezoluciju u kojoj se ističu manipulativne prakse koje mogu oslabiti i ugroziti demokratiju (uplitanje u izborni proces, političko mikrotargetiranje, pojačavanje propagande i polarizacija i erozija kritičkog razmišljanja (paragraf 14.4) i napominje značaj saradnje između različitih aktera radi

⁹Pogledati, na primjer: Preporuku (CM/Rec(2018)7) Odbora ministara državama članicama o smjernicama o poštovanju, zaštiti i ostvarivanju prava djeteta u digitalnom okruženju.; Preporuku CM/Rec(2018)2 Odbora ministara državama članicama o ulogama i odgovornostima internetskih posrednika.; Preporuku 2102(2017) Parlamentarne skupštine Vijeća Evrope (PACE) o tehnološkoj konvergenciji, umjetnoj inteligenciji i ljudskim pravima.

identifikacije „skupa opšteprihvaćenih principa o načinu odgovora na probleme vezane za primjenu UI“ (paragraf 9).

Među ovim postojećim dokumentima, najdetaljnije smjernice su date u Preporuci CM/Rec(2020)1 Odbora ministara o uticaju algoritamskih sistema na ljudska prava. U ovoj politici se predlaže pristup zasnovan na ljudskim pravima kojim će se državnim i privatnim akterima onemogućiti korištenje praksi (uključujući razvoj i korištenje algoritamskih sistema) koje ograničavaju ili utiču na ljudska prava. Kao garancija postupanja ovih aktera u skladu sa datim pristupom predlaže se osnivanje nadzornog i nezavisnog tijela zaduženog za „provjeru, nadzor i koordinaciju poštovanja relevantnog zakonodavnog i regulatornog okvira u skladu sa ovom preporukom“ (paragraf 4) koje će učestvovati u redovnim konsultacijama sa različitim akterima, uključujući javne institucije iz oblasti obrazovanja i kulture, kako bi se garantovalo „detaljno praćenje projektovanja, razvoja i primjene algoritamskih sistema te rasprava i postupanje po tim pitanjima“ (paragraf 5).

Odbor ministara VE je osnovao Ad-hoc odbor za umjetnu inteligenciju (CAHAI) prvenstveno sa ciljem ispitivanja postojanja potrebe za posebnom regulacijom sistema UI sa aspekta ljudskih prava, demokratije i vladavine prava i načina na koje bi se to moglo uraditi. Studija izvodljivosti CAHAI-a, usvojena u decembru 2020, uvodi ideju da bi se ljudska prava trebala posmatrati kroz „socio-tehničku“ prizmu UI kako bi se razumjele vrijednosti ljudi koji je koriste i stvaraju (Leslie et al. 2021, str. 13). CAHAI je predložio sljedećih devet ključnih principa na kojima počiva njihov regulatorni okvir: ljudsko dostojanstvo, ljudska sloboda i autonomija, sprečavanje negativnih uticaja, nediskriminacija, rodna jednakost, pravičnost i raznolikost, transparentnost i objašnjivost sistema UI, zaštita podataka i pravo na privatnost, odgovornost i obaveze, demokratija i vladavina prava (ibid, str. 19). CAHAI je u decembru 2021. usvojio i podnio Odboru ministara dokument pod nazivom „Potencijalni elementi pravnog okvira za umjetnu inteligenciju zasnovanog na standardima Vijeća Evrope o ljudskim pravima, demokratiji i vladavini prava“ u kojem se ponavljaju neki od regulatornih elemenata predloženi u Studiji izvodljivosti. Slično elementima koji se pominju u Rezoluciji 2341 PACE-a – poput manipulacije izborima i profiliranja (Team AI Regulation 2021), – ovi predloženi regulatorni elementi, između ostalog, ističu značaj bavljenja uticajem UI na javno mnijenje i njegovim efektom odvratanja ljudi od javnog učešća.

Pored ovih regulatornih inicijativa VE, EU je također pokrenula niz regulatornih intervencija. Realizacijom ovih intervencija EU ima za cilj da podrži tehnološke inovacije među svim tržišnim akterima (uključujući velike tehnološke kompanije) uz istovremeno rješavanje problema njihovih potencijalnih negativnih uticaja (Daly et al. 2021). To se vidi u Bijeloj knjizi o umjetnoj inteligenciji iz 2020: Evropski pristup izvrsnosti i izgradnji povjerenja, koju je objavila Evropska komisija. Okviri koje oni u ovom dokumentu predlažu za rješavanje rizika koji se tiču ljudskih prava praćeni su nizom sektorskih regulatornih instrumenata – a jedan od njih je i nacrt Akta o digitalnim uslugama (DSA).¹⁰

Sa aspekta slobode izražavanja i slobode medija, DSA je najsveobuhvatniji i prvi regulatorni instrument na nivou EU koji se bavi digitalnom moći velikih tehnoloških kompanija. U ovom aktu se predlaže skup različitih obaveza (npr. obavezna revizija, pravila o transparentnosti, pristup istraživačima itd.) internetskih posrednika koji su podijeljeni u nekoliko kategorija, uključujući „vrlo velike online platforme“ odnosno velike tehnološke kompanije, kako smo ih prethodno nazivali (OSCE SAIFE 2021). Radi garantovanja ravnopravnijih uslova za aktere

¹⁰Evropska komisija. Prijedlog uredbe Evropskog parlamenta i Vijeća o jedinstvenom tržištu digitalnih usluga (Akt o digitalnim uslugama) i izmjeni Direktive 2000/31/EZ.COM/2020/825 final.

na ovom digitalnom tržištu, EU je predložila nacrt Akta o digitalnim tržištima (DMA).¹¹ Pored ovih regulatornih mehanizama na nivou politika usvojeni su i tematski instrumenti koji se bave problematikom govora mržnje¹² i dezinformacija.¹³ U ovim tematskim instrumentima primijenjeni su modeli učešća različitih aktera i multidisciplinarni modeli uz posebno učešće velikih tehnoloških kompanija, poput Facebooka, Googlea i TikToka, u rješavanju problema vezanih za dezinformacije (Burri et al. 2020, str. 40).

U mapiranju okvira politika vezanih za UI kroz prizmu VE i EU prethodno pomenuti dokumenti ilustruju dva glavna stava: prvo, sve veći konsenzus da UI i algoritmi predstavljaju izazov za niz ljudskih prava. Kako bi se odgovorilo na te izazove, pristup zasnovan na ljudskim pravima – skup normi i vrijednosti kojim se štite ljudska prava, demokratski principi i vladavina prava – javlja se kao istaknuti aspekt regulatornog projektovanja, razvoja, uvođenja i primjene UI i algoritamskih sistema (PACE 2020, paragraf 11; CM/Rec(2020)1). Drugi stav naglašava odgovornost, značaj i uključivanje različitih aktera i multidisciplinarnih perspektiva u procese planiranja politika (CoE MSI-AUT(2018)06, paragraf 5; CM/Rec(2020)1, paragraf 5. B.1.3; Leslie et al. 2021, str. 31; Bernstein et al. 2020, str. 48).

Realizacija nastojanja iz ovih propisa dodatno ilustruje rastuću potrebu zauzdavanja digitalne moći velikih tehnoloških kompanija. Kao što je već pokazano u ovoj studiji, UI i algoritamski sistemi – ove kompanije oduzimaju moć djelovanja pojedinca, učutkuju i obeshrabruju manjinske grupe i povećavaju društvene tenzije. U analizi navedenih regulatornih okvira identifikovan je normativni okvir koji treba uvesti radi smanjivanja negativnih uticaja automatizovanih tehnologija. U narednom dijelu je taj okvir detaljnije obrađen u odnosu na intervencije MIL-a.

5.1. Medijska i informacijska pismenost (MIL) i UI: perspektive u zemlji

Kao što to ističe Vijeće Evrope, intervencije medijske i informacijske pismenosti su osnovni instrumenti za primjenu mehanizama zaštite ljudskih prava, demokratije i vladavine prava. MIL se konceptualno stalno mijenja kako bi bio u skladu sa novim socio-tehnološkim izazovima (Chapman & Oermann 2020, str. 10) pa tako kao okvir nudi neograničen potencijal za učešće pojedinaca i društava. Prethodno analizirani dokumenti prepoznaju ovaj potencijal ključne medijske, informacijske i digitalne pismenosti, na primjer, da bi se razumio značaj informisanog pristanka i posljedice algoritamskih sistema, intervencije MIL-a bi trebale biti osmišljene kao obrazovne kompetencije (CM/Rec(2020)1, paragraf 7; CM/Rec(2017)8; CoE MSI-AUT(2018)06, paragraf 7; T-PD(2017)01, paragraf 9). Pored toga, kako bismo smanjili svoju izloženost prijetnjama povezanim sa UI „[...]javnost treba bolje upoznati sa činjenicom da se algoritamski alati uveliko koriste u komercijalne, a sve više i u političke svrhe, kao i za ostvarenje ambicija antidemokratskih i nedemokratskih snaga da dođu na vlast, u svrhe ratovanja i nanošenje direktne štete“ (CM Decl(13/02/2019)1, paragraf 9.e).

U tom kontekstu, buduće intervencije MIL-a ne bi trebale biti usmjerene samo na tehnološke aspekte UI i algoritamskih sistema nego i na:

¹¹ Evropska komisija. Prijedlog uredbe Evropskog parlamenta i Vijeća o konkurentnim i fer tržištima u digitalnom sektoru (Akt o digitalnim tržištima). COM/2020/842 final.

¹² Evropska komisija.(2016). Kodeks postupanja EU radi suzbijanja nezakonitog govora mržnje na internetu. (Usvojen 16. maja 2016).

¹³ Evropska komisija.(2018). Kodeks dobre prakse u suzbijanju dezinformacija.

- njihove vrijednosti, ciljeve i negativne uticaje na pojedince i grupe sa aspekta društvene kohezije i mirovnih procesa;
- prakse iskorištavanja podataka i poslovne modele velikih tehnoloških kompanija;
- lokalne kontekste, lokalne negativne uticaje i prilike koje se na ovaj način oduzimaju ljudima.

Na ovaj način kroz intervencije MIL-a će se predstaviti složen i slojevit svijet digitalne moći i uticaja UI na naš svakodnevni život. Pri realizaciji ovih intervencija posebna pažnja se mora posvetiti starijoj populaciji (preko 65 godina), odraslima bez formalnog obrazovanja ili sa vrlo niskim stepenom obrazovanja, njegovateljima i svima koji aktivno ne koriste ili nisu u interakciji sa ovim digitalnim tehnologijama (devet od deset ljudi u Bosni i Hercegovini, pogledati: VE 2021, str. 5, 55, 65). U tom smislu, buduće namjenski osmišljene i ciljane intervencije MIL-a, koje obuhvataju mnoštvo povezanih pitanja (verifikacija resursa, svijest o rasprostranjenosti primjene UI i algoritama, moderacija sadržaja i upravljanje ličnim informacionim prostorima, ciljano oglašavanje i profiliranje (pogledati: Prilog A)), iziskuju istinski angažman i saradnju različitih državnih i nedržavnih aktera.

VI. Zaključak

Kao što je pokazano u ovoj studiji, rizici koje UI i algoritmi predstavljaju su nestalni, nepostojani i promjenjivi. Svjedoci smo radikalne promjene koja velikim tehnološkim kompanijama omogućava da steknu digitalnu moć u cijelom digitalnom svijetu. Zahvaljujući toj moći, sistemi i tehnologije koje te kompanije uvode prisvajaju i iskorištavaju sve aspekte ljudskog života. Radi zaštite naših ljudskih prava te kompanije i njihovi socio-tehnološki sistemi moraju biti smješteni u širi okvir medijske i informacijske pismenosti kako bi se razotkrile njihove vrijednosti, ciljevi i moć djelovanja. Ova studija, kao jedan od koraka u uvođenju ovih mehanizama zaštite, analizira intervencije MIL-a kao okvir za razumijevanje negativnih uticaja ove višedimenzionalne mreže na našu autonomiju i ljudska prava istovremeno ističući značaj pristupa zasnovanog na ljudskim pravima i učešća različitih aktera ne samo kao obaveze nego i kao sastavnog dijela okvira MIL-a.

U tom kontekstu, holističke i kritičke intervencije MIL-a pojavljuju se kao društveni pokretači koji pojedincima omogućavaju da shvate razmjere i složenost negativnih uticaja ovih tehnologija i tako promovišu sposobnost pametnog korištenja usluga velikih tehnoloških kompanija. Mi trenutno prolazimo kroz prelazni paradigmatički period. Sa našom sve većom zavisnošću od UI i algoritamskih tehnologija rađa se novi komunikacioni i digitalni poredak. Trenutno je teško predvidjeti dalji uticaj UI i algoritama na naš život i slobodu izražavanja pa studije ove vrste treba posmatrati kao tekući i iterativni proces koji se stalno mora ažurirati.

Prilog A

Resursi MIL-a

U tabelama su prikazani korisni resursi koji su organizovani na sljedeći način:

1. Tabela 1: Pismenost u oblasti UI
2. Tabela 2: komplet materijala o digitalnoj sigurnosti i načini jačanja moći djelovanja i kontrole korisnika nad podacima
3. Tabela 3: resursi za provjeru vjerodostojnosti digitalnog sadržaja
4. Tabela 4: studije slučaja i dodatne ilustracije

Tabela 1: Pismenost u oblasti UI

Algorithmic literacy , Kids Code Jeuness, CCUNESCO and UNESCO
A collection of critical essay and videos , C.Sinders and B.Kostic, OSCE
Trips and tricks , learn more about AI, C.Sinders and B.Kostic, OSCE
Digital resilience , handbook for teachers, CoE
AI glossary , CoE
Digital school for children , Digital School Slovenia
AI literacy resources , CoE data visualisation of AI initiatives
AI Literacy , Net Literacy
AI Literacy 101 , Schouten, Towards Data Science
What to Read: A Biassed Guide to AI Literacy for the Beginner , P. Agre, MIT Libraries

Tabela 2: Komplet materijala o digitalnoj sigurnosti

Safety tool-boxes , Front Line Defenders and Tactical Tech
Toolkit , Share Foundation (available in BHS)
Privacy protection tools and a video explainer , Share Foundation (available in BHS)
EFF Tools to protect privacy, Electronic Frontier Foundation
XYZ , gender and digital safety, Tactical Tech
Data resilience , Our Data Bodies
Totem project, courses , Free Press Unlimited

Tabela 3: Resursi za provjeru vjerodostojnosti

A Beginner’s Guide to Social Media Verification , Bellingcat
Insights and Recommendations for AI and Media Integrity , Partnership on AI.
Understanding and countering deepfakes , Witness
Analysis of browser extensions to flag suspicious content , Reuters Institute
Facebook split screen , Markup

Tabela 4: Studije slučaja i ilustracije

A certain kind of doom scrolling , Caroline Sindors
Can data die? , The Pudding
Scary side of reality , Deepfakes, MIT Technology Review
Human rights violations , case studies and resources, Ranking Digital Rights
AI human factory and Facebook pyramid , Share Foundation
Meta-verse , Basic information, Washington post
Data Is power , Privacy International
How to make sure you don’t take personalization too far , Harvard Business Review
Understanding difference between AI and Machine Learning , Microsoft

Prilog B:

Rječnik pojmova u oblasti UI

Praćenje radi oglašavanja (ad-tracking): praksa prikupljanja ličnih uvida o ponašanju i preferencijama pojedinaca radi izrade ciljanih i personalizovanih oglasa.

Algoritmi: skup instrukcija koje su pripremili ljudi sa programiranim procedurama za pretvaranje ulaznih podataka u željene izlazne rezultate na osnovu posebnih izračuna.

Negativni uticaji algoritamskih sistema: podsticanje i povećavanje društvene nejednakosti i diskriminatornih praksi u digitalnom (hibridnom) prostoru čime se dodatno marginalizuju obespravljene zajednice i društva koja prolaze kroz poslijeratni period (npr. intersekcionalna diskriminacija, pojačana cenzura, praznine u informisanju, manipulacija informacijama velikih razmjera).

Umjetna inteligencija (UI): skup nauka, teorija i tehnika koje imaju za cilj da reprodukuju, pomoću mašine, kognitivne sposobnosti ljudskog bića.

Subjektivnost UI: UI su napisali, stvorili i programirali ljudi. Računski kod i baze podataka koje pokreću algoritme, a koji su gradivni elementi UI, odraz su ljudskih vrijednosti, subjektivnosti, potreba i želja.

Automatizovani sistem odlučivanja: tehnološki procesi u kojima se ljudski procesi odlučivanja unapređuju i zamjenjuju računarskim procesorima koji odgovaraju na različita pitanja poput klasifikacije (npr. ženski-muški-nebinarni pol) ili daju procjene kao u slučaju kreditne sposobnosti, rizika od kriminala.

Moderacija sadržaja: procesi praćenja, procjene, odabira i distribucije sadržaja zasnovani na praksama profiliranja i nadzora.

Kolačići: datoteka koja se stvara u direktoriju pretraživača kako bi se stvorila poveznica između pojedinca i stranice koju je posjetio. Sve aktivnosti na toj web stranici pohranjuju se u ovim malim datotekama.

Prenamjena podataka: praćenje i prikupljanje ličnih i ostalih podataka pojedinca u više različitih digitalnih prostora i platformi za određenu svrhu koji se u kasnijoj fazi recikliraju i koriste za druge namjene.

Uvjerljivi krivotvoreni sadržaj: video i audio materijal kojeg generiše UI i kojim se stvara „alternativna realnost“ ili „sintetički medij“ u kojem izgleda da neko govori ili radi nešto što nije rekao ili uradio.

Digitalna moć UI: koncentracija podataka informacija i uticaja kojim raspolaže mali broj tehnoloških kompanija u kombinaciji sa rastućom zavisnošću pojedinaca, društava i institucija o njihovim uslugama i nedostatkom djelotvorne demokratske kontrole.

Eho komora: Okruženje u kojem je osoba izložena samo sličnim uvjerenjima i mišljenjima.

Uticaj UI na donošenje odluka ljudi: UI i procesi zasnovani na podacima su duboko ukorijenjeni i oblikuju kontekst u kojem pojedinci donose odluke, a samim tim utiču na ličnu autonomiju ljudi.

Internetski balon: Sistem preporučivanja koji izlaže ljude sličnom homogenom sadržaju koji odgovara njihovim interesovanjima i preferencijama.

Mašinsko učenje: vid (polu) autonomnog učenja i razvojna tehnika UI osmišljena radi unapređenja kvaliteta automatizovanog odlučivanja na osnovu prepoznavanja obrazaca i „pravilnosti“ kako bi se određeni zadaci obavljali bez intervencije čovjeka.

Čovjek u petlji: potreba za ljudskom intervencijom u ovim automatizovanim procesima koja će imati nadzornu i korektivnu funkciju kako bi se garantovalo poštovanje i zaštita ljudskih prava pojedinca.

Platformizacija interneta: uspon platforme kao dominantnog infrastrukturnog i ekonomskog modela društvene mreže i posljedice širenja platformi društvenih medija na druge online prostore.

Profiliranje: izvođenje ličnih podataka i obrazaca ponašanja iz naizgled nezanimljivih podataka pomoću tehnologija UI kako bi se proizveli digitalni profili koji se koriste za informisanje budućih odluka pojedinaca (npr. za potrebe praćenja radi oglašavanja).

Odgovorna UI: oblast koja nudi smjernice za projektovanje, razvoj i korištenje UI u skladu sa etičkim standardima i standardima ljudskih prava.

Društvo u petlji: sistem koji uključuje vrijednosti i interese različitih društvenih grupa, etičkih principa i principa ljudskih prava, kao i mogućnost učešća u projektovanju, razvoju, uvođenju i dostupnosti mehanizama pravne zaštite kod automatizovanih sistema donošenja odluka.

Korisnička kontrola: mjera u kojoj pojedinac može ili ne može donositi odluke u vezi sa korištenjem njegovih ličnih podataka i digitalnog ponašanja.

Reference

Pravni tekstovi i politike

Vijeće Evrope. (1950). Evropska konvencija o zaštiti ljudskih prava i osnovnih sloboda, kako je izmijenjena protokolima 11 i 14 (ETS br. 5, 04/11/1950).

Vijeće Evrope. (1997). Preporuka (R(97)20) Odbora ministara državama članicama o govoru mržnje

Vijeće Evrope. (1981). Konvencija o zaštiti lica u pogledu automatske obrade ličnih podataka (ETS br. 108, 28.1.1981).

Vijeće Evrope. (2017). Smjernice uz Konvenciju 108 o zaštiti lica u pogledu obrade ličnih podataka u svijetu velikih podataka. (T-PD(2017)01).

Vijeće Evrope. (2017). Preporuka Odbora ministara državama članicama o velikim podacima za kulturu, pismenost i demokratiju. (CM/Rec(2017)8).

Vijeće Evrope. (2018). Nacrt preporuke Odbora ministara državama članicama o uticaju algoritamskih sistema na ljudska prava. (MSI-AUT(2018)06rev3).

Vijeće Evrope. (2018). Preporuka (CM/Rec(2018)7) Odbora ministara državama članicama o smjernicama za poštovanje, zaštitu i ostvarivanje prava djeteta u digitalnom okruženju.

Vijeće Evrope. (2018). Preporuka CM/Rec(2018)2 Odbora ministara državama članicama o ulogama i odgovornostima internetskih posrednika.

Vijeće Evrope. (2019). Deklaracija (CM Decl(13/02/2019)1) Odbora ministara o manipulativnim mogućnostima algoritamskih procesa.

Vijeće Evrope. (2019). Smjernice uz Konvenciju 108 o umjetnoj inteligenciji i zaštiti podataka. (T-PD(2019)01).

Vijeće Evrope. (2020). Preporuka (CM/Rec(2020)1) Odbora ministara državama članicama o uticaju algoritamskih sistema na ljudska prava.

Vijeće Evrope. (2020). PACE. Need for democratic governance of artificial intelligence. Dok. 15150. 24. septembar 2020.

Evropska komisija. (2016). Uredba (EU) 2016/679 Evropskog parlamenta i Vijeća o zaštiti fizičkih lica u vezi sa obradom ličnih podataka i o slobodnom kretanju tih podataka te o stavljanju van snage Direktive 95/46/EZ (Opšta uredba o zaštiti podataka). (Usvojena 27. aprila 2016).

Evropska komisija. (2016). Kodeks postupanja EU radi suzbijanja nezakonitog govora mržnje na internetu. (Usvojen 16. maja 2016).

Evropska komisija. (2018). Kodeks dobre prakse u suzbijanju dezinformacija.

Evropska komisija. (2020) Prijedlog uredbe Evropskog parlamenta i Vijeća o konkurentnim i fer tržištima u digitalnom sektoru (Akt o digitalnim tržištima). COM/2020/842 final.

Evropska komisija. (2020). Prijedlog uredbe Evropskog parlamenta i Vijeća o jedinstvenom tržištu digitalnih usluga (Akt o digitalnim uslugama) i izmjeni Direktive 2000/31/EZ. COM/2020/825 final.

Evropska komisija. (2020). Bijela knjiga o umjetnoj inteligenciji: Evropski pristup izvrsnosti i izgradnji povjerenja.

Parlamentarna skupština Vijeća Evrope (PACE). Preporuka 2102(2017) o tehnološkoj konvergenciji, umjetnoj inteligenciji i ljudskim pravima.

Parlamentarna skupština Vijeća Evrope (PACE). (2020) Need for democratic governance of artificial intelligence. Rezolucija PACE/2341.

Parlamentarna skupština Vijeća Evrope (PACE). (2020). Need for democratic governance of artificial intelligence. Izvještaj PACE/15150.

Bibliografija

Al-Kawaz, H., N, Clarke., Furnell, S., Li, F., Abdulrahman, A. (2018). Advanced facial recognition for digital forensics. U A. Jøsang (ur.), *Proceedings of the 17th European Conference on Information Warfare and Security: ECCWS 2018*. Academic Conferences and Publishing International Limited.

Bandy, J., i Diakopoulos, N. (2020). Auditing news curation systems: A case study examining algorithmic and editorial logic in Apple News. U *Proceedings of the International AAAI Conference on Web and Social Media* (tom 14, str. 36-47).

Bernstein, A., De Vreese Claes., Helberger, N., S, Wolfgang., Z, Katharina. (2020). Diversity, Fairness, and Data-Driven Personalization in (News) Recommender System, *Dagstuhl Manifestos*, Dagstuhl Perspectives Workshop tom 9, izdanje 1.

Bansal, A., Agarwal, R., i Sharma, R. K. (2012). SVM based gender classification using iris images. U *2012 Fourth International Conference on Computational Intelligence and Communication Networks* (str. 425-429).

Baujard, T., Tereszkiwicz, R., de Swarte, A., Tuovinen, T. (2019). Entering the new paradigm of artificial intelligence and series. Council of Europe and Eurimages. Bloch-Wehba, H. (2020). Automation in moderation. *Cornell International Law Journal*. 53, 41.

Bodó, B., Helberger, N., Eskens, S., i Möller, J. (2019). Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. *Digital journalism*, 7(2).

Bukovska, B. (2020). Spotlight on Artificial Intelligence and Freedom of Expression. Predstavnik OSCE-a za slobodu medija (RFoM).

Burri, M., Eskens, S., Farish, K., Frosio, G., Guidotti, R., Jääskeläinen, A., Pin, A., Raižytė, J. (2020). Artificial intelligence in the audiovisual sector. Evropska audiovizuelna opservatorija.

Causevic, A. i Sengupta, A. (2020). Whose Knowledge Is Online? Practices of Epistemic Justice for a Digital New Deal. IT For change. Čaušević et al. 2020

Chapman, M., i Oermann, M. (2020). Supporting Quality Journalism through Media and Information Literacy. Vijeće Evrope. Strasbourg.

Cobbe, J. (2020). Algorithmic Censorship by Social Platforms: Power and Resistance. Philosophy & Technology.

Couldry, N., i Mejias, U. A. (2019). The costs of connection. How Data Is Colonizing Human Life and Appropriating It for Capitalism. Stanford University Press. Redwood City.

Ad-hoc odbor Vijeća Evrope za umjetnu inteligenciju (CAHAI). (2020). Toward Regulation of AI systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law.

Ad-hoc odbor Vijeća Evrope za umjetnu inteligenciju (CAHAI). (2020). Feasibility Study. CAHAI(2020)23.

Komesar Vijeća Evrope za ljudska prava. (2019). Umjetna inteligencija: 10 koraka za zaštitu ljudskih prava.

Vijeće Evrope (2020). Digital Resistance: An empowering handbook for teachers on how to support their students to recognise fake news and false information found in the online environment.

Vijeće Evrope. (2021). Zaključci sa konferencije: Human Rights in the Era of AI Europe as international Standard Setter for Artificial Intelligence.

Vijeće Evrope. (2021). Medijske navike i stavovi: Studija o medijskim navikama odraslih u Bosni i Hercegovini.

Gillespie, T. (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. London: Yale University Press.

Cvjetićanin, T. (2019). Dezinformacije u online sferi: slučaj BiH. OCD Zašto ne? Sarajevo.

Gerlitz, C. i Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society* 15(8).

Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B., & Wang, W. W. (2021). AI, Governance and Ethics: Global Perspectives. In H-W. Micklitz, O. Pollicino, A. Amnon, A. Simoncini, G. Sartor, & G. De Gregorio (ur.), *Constitutional Challenges in the Algorithmic Society* Cambridge University Press.

Dixon, P. (2013). Svjedočenje u Kongresu: What Information Do Data Brokers Have on Consumers?. World Privacy Forum.

Donovan, J. i Boyd, D. (2019). Stop the presses? Moving from strategic silence to strategic amplification in a networked media ecosystem. *American Behavioral Scientist*, 1–18. SAGE Publication.

Evropska komisija Resorni odjel za vanjsku politiku. (2021). Mapiranje lažnih vijesti i dezinformacija na Zapadnom Balkanu i identificiranje načina za efikasnu borbu protiv njih. Brussels.

Evropska komisija. (2020). Radni dokument Komisije. Izvještaj o Bosni i Hercegovini za 2020. godinu. Uz Saopštenje Komisije Evropskom parlamentu, Vijeću, Evropskom ekonomskom i socijalnom odboru i Odboru regija (SWD(2020) 350 final). Brussels.

European Digital Rights Initiative. (2020). DSA: Platform Regulation Done Right.

Fairhurst, M., Da Costa-Abreu, M. i Erbilek, M. (2015). Exploring gender prediction from iris biometrics. U *2015 International Conference of the Biometrics Special Interest Group (BIOSIG)* (str. 1-11). IEEE.

Gillespie, T. (2014). The relevance of algorithms. In *Media technologies: Essays on communication, materiality, and society*. MIT Press.

Gorwa, R., Binns, R. i Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*. SAGE.

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, 7(8), 993-1012.

Helberger, N. (2020). The political power of platforms: How current attempts to regulate misinformation amplify opinion power. *Digital Journalism*, 8(6), 842-854.

Helberger, N., Eskens, S. J., van Drunen, M. Z., Bastian, M. B. i Möller, J. E. (2019). Implications of AI-driven tools in the media for freedom of expression. U *Artificial intelligence–Intelligent politics: Challenges and opportunities for media and democracy*.

Helmond, A. (2015). The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society*.

Hodžić, S. (2019). Surfanje po tankom ledu: mladi, mediji i problematičan sadržaj. Media Centar Sarajevo. Sarajevo.

Jenkins, H. (2006). *Convergence Culture: Where Old and New Media Collide*. New York. New York University Press.

Jørgensen, R. F. (2019). *Human rights in the age of platforms*. The MIT Press.

Kalluri, P. (2020). Don't ask if AI is good or fair, ask how it shifts power. World View. *Nature* Vol. 583. Springer.

Keller, D. (2021). Amplification and Its Discontents. Occasional Papers. Knight First Amendment Institute. Columbia University.

Kaluža, J. (2021). Habitual Generation of Filter Bubbles: Why is Algorithmic Personalisation Problematic for the Democratic Public Sphere?, *Javnost*.

Klonick, K. (2017). The New Governors: The People, Rules, and Processes Governing Online Speech. 131 *Harv. L. Rev.* 1598.

Kostić, B. (u pripremi). Report: Content moderation on Social Media and Map of Stakeholders in Bosnia and Herzegovina. ARTICLE 19.

Kostić, B. (2021). Veštačka inteligencija - Uticaj na slobodu izražavanja, medijske perspektive i regulatorni trendovi (2021) OSCE Srbija. dostupno na: <https://www.osce.org/sr/mission-to-serbia/479672>

Kuehlkamp, A., Becker, B. i Bowyer, K. (2017). Gender-from-iris or gender-from-mascara?. In *2017 IEEE Winter conference on applications of computer vision (WACV)* (str. 1151-1159). IEEE.

Lagree, S. i Bowyer, K. W. (2011). Predicting ethnicity and gender from iris texture. U *2011 IEEE international conference on technologies for homeland security (Hst)* (str. 440-445). IEEE.

Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., Briggs, M. (2021). Artificial intelligence, human rights, democracy and the rule of law. *Vijeće Evrope*.

Leerssen, P. (2020). The Soap Box as a Black Box: Regulating transparency in social media recommender systems. *European Journal of Law and Technology*, 11(2).

Maréchal, N. i Biddle, E. R. (2020). It's not just the content, it's the business model: democracy's online speech challenge. *New America — Ranking Digital Rights*, 17. mart.

Mazzoli, M. E. i Tambini, D. (2020). Prioritisation uncovered: The Discoverability of Public Interest Content Online. *Vijeće Evrope*.

Möller, J., Helberger, N. i Makhortykh, M. (2019). Filter Bubbles in The Netherland. Hilversum: Commissariaat voor de Media.

Molnar, P. (2020). Technological Testing Grounds: Border tech is experimenting with people's lives. *Evropska digitalna prava (EDRi)*. Brussels.

Mozilla. (2021). YouTube Regrets. A crowdsourced investigation into YouTube's recommendation algorithm.

Newman, N., Fletcher, R., Schulz, A., Andl, S. i Nielsen, R. K. (2020). Reuters Institute Digital News Report 2020. Reuters Institute for the Study of Journalism. Oxford University.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Onuoha, M. i D, Nucera. (2018). *A People's Guide to AI : Artificial Intelligence*. Open Society Foundations.

Predstavnik za slobodu medija Organizacije za sigurnost i saradnju u Evropi (OSCE RFoM). (2021). *A Policy Manual. Spotlight on Artificial Intelligence and Freedom of Expression # SAIFE*.

Predstavnik za slobodu medija Organizacije za sigurnost i saradnju u Evropi (OSCE RFoM). (2020). *Non-paper on the impact of artificial intelligence on freedom of expression. #SAIFE*.

Partnership on AI. (2020). *The Deepfake Detection Challenge: Insights and Recommendations for AI and Media Integrity*.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press. Cambridge Massachusetts.

Privacy International i Article 19. (2018). *Privacy and Freedom of expression in the age of artificial intelligence*. London.

Privacy International. (2017). *Data is power: Profiling and Automated Decision-Making in GDPR*. Privacy International. London.

Rahwan, I. (2017). *Society in the Loop: Programming the Algorithmic Social Contract, Ethics of Information Technology*.

Rakova, B., Yang, J., Cramer, H. i Chowdhury, R. (2021). *Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1-23*.

Ranking Digital Rights. (2020). *Human rights risk scenarios: Algorithms, machine learning and automated decision-making (Consultation Draft)*.

Sinders, C. (2021). *Designing Against Dark Patterns*. German Marchall Fund of the United States.

Sokol, A. (2020). *Modeli propagande, dezinformacija i mržnje u medijima i komunikacijama u Bosni i Hercegovini*. SEENPM, Tirana, Mirovni institut, Ljubljana i Fondacija Mediacentar Sarajevo. Sarajevo.

Sokol, A. (2021). *Mrzilački narativi u medijima i korisnički generiranom sadržaju*. Mediacentar Sarajevo. Sarajevo.

Tapia, J. E., Perez, C. A. i Bowyer, K. W. (2014). *Gender classification from iris images using fusion of uniform local binary patterns*. U *European Conference on Computer Vision* (str. 751-763). Springer International Publishing.

Tapia, J. E., Perez, C. A. i Bowyer, K. W. (2016). *Gender classification from the same iris code used for recognition*. *IEEE Transactions on Information Forensics and Security, 11(8), 1760-1770*.

Thomas, V., Chawla, N. V., Bowyer, K. W. i Flynn, P. J. (2007). Learning to predict gender from iris images. *UFirst IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*.

Turčilo, L. i Buljubašić, B. (u pripremi). Uništavanje reputacije na bh način: govor mržnje botova u online prostoru kao sredstvo sužavanja javnog prostora za alternativna mišljenja u Bosni i Hercegovini. U *Konferencija Komentari, govor mržnje, dezinformacija i regulacija javne komunikacije. Agencija za elektroničke medije i Medijska istraživanja. Zagreb.*

UC Berkeley School of Law Human Rights Center Research Team.(2019). Memorandum on Artificial Intelligence and Child Rights.

UNESCO i EQUALS Skills Coalition. (2019). I'd blush if I could: closing gender divides in digital skills through education.

UNESCO. (2020). Artificial Intelligence and Inclusion: Compendium of Promising Initiatives. Mobile Learning Week 2020.

Vijeće Ujedinjenih nacija za ljudska prava (UNHRC). (2018) Izvještaj nezavisne međunarodne misije za utvrđivanje činjenica o Mijanmaru. (A/HRC/39/64).

Wagner, Ben, Wolfgang Schulz, Karmen Turk, Bertrand de la Chapelle, Julia Hörnle, Tanja Kersevan-Smokvina, Mathias C. Kettemann, Dörte Nieland, Arseny Nedyak, Pēteris Podvinskis, Thomas Schneider, Sophie Stalla-Bourdillon i Dirk Voorhoof. (2018) Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications. DGI(2017)12. Vijeće Evrope.

Wardle, C. i Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Vijeće Evrope.

Yeung, K. (2019). A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Vijeće Evrope.

Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. Vijeće Evrope.

Novinski članci, blogovi i kolumne

Ackerman, E. (2021). This Year Autonomous Truck will take on the road with no one on board. IEEE Spectrum.

Appen. (2019). Human in the loop Machine Learning.

Udruženje za unapređenje umjetne inteligencije (AAAI).

BBC. (2019). Apple's 'sexist' credit card investigated by US regulator.

Bell, G. (2020). Touching the future. The Griffith Review.

Brandom, R. (2021). Social media vs. the world: What comes after deplatforming. The Verge.

Glosar pojmova UI Vijeća Evrope

Constine, J. (2016). How Facebook News Feed Works. Techcrunch.

Culliford, E. (2021). Facebook says it removed troll farm run by Nicaraguan government. Swissinfo.ch.

Dojčinović, S. (2017). Hey, Mark Zuckerberg: My Democracy Isn't Your Laboratory. The New York Times.

Duhigg, C. (2012). How Companies Learn Your Secrets. The New York Times.

Evropska komisija. (2019). The Digital Markets Act: ensuring fair and open digital markets.

Evropska građanska inicijativa (ECI). (2021). Reclaim your face.

Explainable AI. (2020). Speech and AI.

Feldstein, S. (2020). When It Comes to Digital Authoritarianism, China is a Challenge — But Not the Only Challenge. Carnegie Endowment for International Peace.

Gilman, M. (2020). AI algorithms intended to root out welfare fraud often end up punishing the poor instead. The Conversation.

Globalni internetski forum za suzbijanje terorizma (GIFCT). Sharing Consortium Hashtag.

Google. Google Perspective.

Ghoshal, D. (2018). Mapped: The breathtaking global reach of Cambridge Analytica's parent company. Quartz.

Hao, K. (2021). A horrifying new AI app swaps women into porn videos with a click. MIT Review.

Harrison, S. (2020). Why Am I Paying \$60 for That Bag of Rice on Amazon.com?. The Markup. (2020).

Hern, A. (2017). Facebook moving non-promoted posts out of news feed in trial. The Guardian.

Horwitz, J. i Seetharama, D. (2020). Facebook executives shut down efforts to make the site less divisive. The Wall Street Journal.

IBM Cloud Education. (2020). Artificial Intelligence (AI).

Jeremić, I. i Stojanovic, M. (2021). Facebook, Twitter Struggling in Fight against Balkan Content Violations, Balkan Insight, BIRN.

Kastrenakes, J. (2016). Microsoft made a chatbot that tweets like a teen. The Verge.

Leetaru, K. (2018). The Data Brokers So Powerful Even Facebook Bought Their Data - But They Got Me Wildly Wrong. The Forbs.

Leufer, D. (2021). Computers are binary, people are not: how AI systems undermine LGBTQ identity. Access now.

Lewis, P. i McCormick, E. (2018). How an ex-YouTube insider investigated its secret algorithm. The Guardian.

Madoleva, S. (2021). Media Freedom in Bosnia and Herzegovina. Konrad Adenauer Stiftung.

Mattu, S., Yin, L., Waller, A. i Keegan, J. (2021). How We Built a Facebook Inspector. The Markup.

Merrill, J. (2021). How Facebook's Ad System Lets Companies Talk Out of Both Sides of Their Mouths. The Markup.

Merrill, J. i Oremus, W. (2021) Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation. The Washington Post.

Metz, R. (2021). How a deepfake Tom Cruise on TikTok turned into a very real AI company. CNN.

Newton, C. (2021). The tier list: how Facebook decides which countries need protection. The Verge.

Nikolic, I. i Jeremic, I. (2020). 'Vox Populi': How Serbian Tabloids and Twitter Bots Joined Forces. Balkan Insight. BIRN.

O'Neil, C. (2021). Facebook's VIP 'Whitelist' Reveals Two Big Problems. Bloomberg.

Our Data Bodies.

Ohlheiser, A. (2019) The one word that lets politicians get away with breaking the rules on social media. Washington Post.

Pirkova, E. (2020). How the Digital Services Act could hack Big Tech's human rights problem. Access now.

Safetyofjournalists.net. (2021). The brutal campaign against Lejla Turcilo and Borka Rudic from Sarajevo must stop.

Statista.com. (2021). Facebook users in Bosnia and Herzegovina from September 2018 to July 2021.

Team AI Regulation. (2021). The Council of Europe's recommendation for a legal framework on AI. AI Regulation.

The Guardian. (2019). New York tenants fight as landlords embrace facial recognition cameras.

The Guardian. (2020). Twitter deletes 20,000 fake accounts linked to Saudi, Serbian and Egyptian governments.

The Markup. (2021). How Facebook's Ad System Lets Companies Talk Out of Both Sides of Their Mouths.

Zucconi, A. (2018). An Introduction to Deepfakes. AlanZucconi.com.

Ova studija daje kritički uvid u postojeće stanje u oblasti UI i izazove koje ona postavlja za ljudska prava. Treba napomenuti da ovom studijom nisu obuhvaćeni svi relevantni aspekti UI i ljudskih prava budući da brojne teme vezane za UI ne spadaju u predmet ove studije – kao što su primjena UI u određenim sektorima, problematika zaštite podataka nevezano za UI i slobodu izražavanja, kao i standardi praćenja i ocjene poštovanja ljudskih prava. Neki od ovih aspekata spominju se kao ilustrativni primjeri ili prateći argumenti, ali je fokus na artikulaciji i mapiranju postojećih stručnih i regulatornih diskusija kroz prizmu UI i slobode izražavanja.

www.coe.int/freedomofexpression

www.coe.int

Vijeće Evrope je vodeća organizacija za ljudska prava na kontinentu. Čini je 46 država članica, uključujući sve članice Evropske unije. Sve države članice Vijeća Evrope su potpisnice Evropske konvencije o ljudskim pravima, instrumenta za zaštitu ljudskih prava, demokratije i vladavine prava. Evropski sud za ljudska prava nadzire implementaciju Konvencije u državama članicama.