U FOKUSU

VEŠTAČKA INTELIGENCIJA I SLOBODA IZRAŽAVANJA

Priručnik za razvoj politika

OSCE
The Representative on Freedom of the Media

Ova publikacija je deo projekta „Veštačka inteligencija i sloboda izražavanja u fokusu“ (#SAIFE). Stavovi, nalazi, tumačenja, preporuke i zaključci ovde izrečeni su stavovi autora i ne predstavljaju nužno zvaničan stav OEBS-a i/ili njegovih država članica.

Stavovi izrečeni u publikaciji pripadaju isključivo autoru i njegovim saradnicima i ne predstavljaju nužno zvaničan stav Misije OEBS-a u Srbiji.

6a Wallnerstrasse
1010 Beč, Austrija
Telefon +43-1-514-36-68-00
e-mail: pm-fom@osce.org
https://www.osce.org/fom/ai-free-speech

ISBN: 978-92-9234-744-4

Veštačka inteligencija i sloboda izražavanja u fokusu
Priručnik za razvoj politika

Autori
Eliska Pirkova, Matthias Kettemann, Marlena Wisniak, Martin Scheinin, Emmi Bevensee, Katie Pentney, Lorna Woods, Lucien Heitz, Bojana Kostic, Krisztina Rozgonyi, Holli Sargeant, Julia Haas, i Vladan Joler

Stručni saradnici
Deniz Wagner i Julia Haas

Eksperti
Jennifer Adams, Susie Alegre, Asha Allen, Andreas Marckmann Andreassen, Nikolett Aszodi, Jef Ausloos, Josephine Ballon, Joan Barata, Nadia Bellardi, Susan Benesch, Guy Berger, Frederik Zuiderveen Borghesius, Irina Borogan, Jonathan Bright, Elda Brogi, Amy Brouillette, Joanna J. Bryson, Pete Burnap, Camilla Bustani, Ignacio Talegon Campoamor, Maja Cappello, Marcelo Daher, Anita Danka, Nicholas Diakopoulos, Aijamal Djanybaeva, Leyla Dogruel, Maria Donde, Sead Dzigal, Franccesca Fanucci, Marc Fumagalli, Maximilian Gahntz, Jana Gajdošová, Maya Indira Ganesh, Lalya Gaye, Brandi Geurkink, Arzu Geybulla, Michele Gilman, Nadine Gogu, Gabrielle Guillemin, Rustam Gulov, Ben Hayes, Natali Helberger, Georgia Holmer, Andrea Huber, Karolina Iwańska, Sam Jeffers, Elliot Jones, Pascal Jürgens, Agnes Kaarlep, Frederike Kaltheuner, Kari Karppinen, Susan Kerr, Benjamin Kille, Yoojin Kim, Wolfgang Kleinwächter, Beata Klimkiewicz, Djordje Krivokapić, Ľuboš Kukliš, Andrey Kuleshov, Joanna Kulesza, Collin Kurre, Susan Landau, Paddy Leerssen, Emma Llansó, James MacLaren, João Carlos Magalhães, Samvel Martirosyan, Estelle Massé, Kyle Matthew, Eleonora Maria Mazzoli, Tarlach McGonagle, Marko Milosavljevič, Mira Milosevic, Iva Nenadić, Marielza Oliveira, Rebekah Overdorf, Roya Pakzad, Sejal Parmar, Patrick Penninckx, Jon Penny, Carlos Perez-Maestro, Emilija Petreska-Kamenjarova, Andrej Petrovski, Courtney Radsch, Otabek Rashidov, Judith Rauhofer, David Reichel, Moritz Riesewieck, Katitza Rodriguez, Asja Rokša-Zubčević, Bianca Schönberger, Christopher Schwartz, Lisa Seidl, Murtaza Shaikh, Jat Singh, Vanja Škorić, Andrei Soldatov, Maria Luisa Stasi, Nikolas Suzor, Damian Tambini, Dhanaraj Thakur, Gulnura Toralieva, Max van Drunen, Vitaly Vasilchenko, Francisco Vera, Kristina Voko, Diana Vlad Calcic, Ben Wagner, Douglas Wake, Hilary Watson, Agnieszka Wawrzyk, Veszna Wessenauer, i Andrej Zwitter

Pregled vršili
OHCHR Ujedinjenih nacija, UNESCO, Savet Evrope, Evropska audiovizuelna opservatorija, Evropska komisija, Agencija za osnovna prava Evropske unije, Evropska radiodifuzna unija, OEBS (Sekretarijat, ODIHR, HCNM)

Urednik primerka
Tom Popper

Dizajn i prelom
Peno Mishoyan

Sadržaj

Predgovor

Poštovani čitaoci,

Sa zadovoljstvom predstavljam publikaciju moje Kancelarije koja stavlja u središte pažnje veštačku inteligenciju i slobodu izražavanja (SAIFE). Ova publikacija je kulminacija dvogodišnjeg istraživanja i nekoliko stručnih radionica, i sažima znanje više od stotinu najpoznatijih naučnika i stručnjaka koji rade u oblasti slobode medija, ljudskih prava, tehnologije i bezbednosti.

U 2022. godini navršava se 25 godina od početka mandata Predstavnika OEBS-a za slobodu medija. 1997. godine, kada je ova institucija osnovana, samo 1,7 posto globalne populacije je bilo onlajn, a digitalne tehnologije koje podržavaju onlajn komunikaciju bile su nove i iskreno optimistične.

Dvadeset pet godina kasnije, broj ljudi koji pristupaju internetu porastao je na više od 80 posto širom regiona OEBS-a. Ova monumentalna ekspanzija je bila izuzetno korisna za slobodu izražavanja, slobodan protok informacija i mogućnost traženja, primanja i prenošenja informacija i ideja svih vrsta preko granica i širom sveta.

To je ključno za ekonomsko, javno i političko učešće, za demokratizaciju, za obrazovanje i zdravstvo, za pozivanje vlasti na odgovornost i za rasvetljavanje ratnih zločina i drugih kršenja ljudskih prava. Istovremeno, to je takođe dovelo do masovnog praćenja, kao i visoko-tehnološkog kriminala i širenja nelegalnog i štetnog sadržaja onlajn.

Upravljanje neizmernim mnoštvom informacija onlajn postalo je nemoguće bez tehnologija mašinskog učenja i drugih oblika veštačke inteligencije (VI). VI tehnologije postaju glavni alati za oblikovanje i arbitriranje sadržaja onlajn; VI se koristi za odlučivanje o tome koji sadržaj se uklanja, koji sadržaj ima prioritet ili kome se distribuira. Te odluke se sprovode pomoću tehnologije koju razvija i primenjuje nekolicina onlajn platformi – čuvara ulaza u digitalni svet.

To su moćne kompanije sa sposobnošću da oblikuju i arbitriraju politički i javni diskurs. Nema sumnje da način na koji se onlajn informacije priređuju i dele (kuriraju) i moderiraju ima direktan i značajan uticaj na globalni mir, stabilnost i sveobuhvatnu bezbednost. Takvu moć mora da prati odgovornost. I pored toga, ovi novi čuvari – i njihova poslovna praksa – razvijaju se brzinom koja nadmašuje bilo koji pravni ili regulatorni okvir za korišćenje veštačke inteligencije za oblikovanje našeg informacionog prostora onlajn.

Nalazimo se na prekretnici.

Države članice OEBS-a moraju se ujediniti kako bi pronašle multilateralna rešenja za izazove sa kojima se suočava njihov zajednički informacioni prostor. One to moraju da obave tako što će ljudska prava staviti u centar razvoja i primene VI za priređivanje onlajn sadržaja i moderaciju sadržaja.

Navedeni izazovi su dalekosežni, a rešenja se mogu pronaći samo kroz delovanje većeg broja različitih zainteresovanih strana. Što se tiče izazova koji se odnose na slobodu medija i slobodu izražavanja, nadam se da će ova publikacija pomoći državama članicama OEBS-a, kreatorima politike, akademskoj zajednici i medijskim profesionalcima, u regionu i šire, da spoznaju kako da zajednički razviju mere zaštite ljudskih prava u okviru svojih nacionalnih, regionalnih i međunarodnih kapaciteta.

decembar, 2021. godine

Teresa Ribeiro
Predstavnica OEBS-a za slobodu medija

Ključne preporuke za države članice OEBS-a

1. Štititi i promovisati slobodu izražavanja i druga **ljudska prava kao okosnicu** strategija i politika u vezi sa veštačkom inteligencijom.

2. Očuvati i negovati **internet** kao prostor za **demokratsko učešće** i predstavljanje, i za **medijski pluralizam**.

3. Razviti politike **zasnovane na dokazima**, izgrađene na **inkluzivnim procesima**, kako bi se odgovorilo na izazove usmerene na slobodu mišljenja, slobodu informisanja i slobodu izražavanja.

4. Promovisati usklađenost sa **Vodećim principima UN-a o poslovanju i ljudskim pravima**, kako bi se sprečilo davanje prioriteta ostvarivanju najvećeg mogućeg profita na štetu ljudskih prava i demokratskih vrednosti.

5. Obavezati onlajn platforme da sprovode **dubinsku analizu ljudskih prava**, uključujući procene uticaja na ljudska prava **(HRIA)** za svoje politike upravljanja sadržajem i automatsko donošenje odluka, kao i za poslovne prakse, kao što su prikupljanje podataka (data harvesting), targetirano oglašavanje i dizajn interfejsa.

6. Obezbediti **jasnoću, objašnjivost i pristupačnost** u korišćenju veštačke inteligencije za moderaciju sadržaja, priređivanje i deljenje sadržaja i targetirano oglašavanje.

7. Osigurati da zaštita ljudskih prava nije u potpunosti eksternalizovana ili automatizovana, i obezbediti transparentnost u vezi sa bilo kojim **javno-privatnim partnerstvom.**

8. Uvesti jake okvire **transparentnosti**, uključujući obavezujuće sveobuhvatne izveštaje o transparentnosti koji sadrže detaljne informacije o korišćenju veštačke inteligencije.

9. Omogućiti uspostavljanje **snažnih mehanizama pravnih lekova** protiv cenzure i praćenja, uključujući i mehanizme provere od strane pojedinaca i nezavisne žalbene mehanizme.

10. Garantovati snažnu **odgovornost**, uključujući **nezavisni nadzor i nezavisnu reviziju**, posebno u pogledu poštovanja ljudskih prava i nediskriminacije.

11. Poštovati pravo na **privatnost** i zaštitu podataka, uključujući utvrđivanje ograničenja za oglašavanje zasnovano na praćenju i obezbeđivanje jasne transparentnosti i uključivanje korisnika u poslovne prakse zasnovane na praćenju i profilisanju.

12. Promovisati medijsku i **digitalnu pismenost** i pospešiti **osnaživanje korisnika, njihovo posredovanje i kontrolu** nad upravljanjem sadržajem i korišćenjem njihovih podataka, uključujući pružanje mogućnosti da odustanu od automatizovanog donošenja odluka.

13. Rešavati neuravnotežene i **monopolizovane tržišne kapacitete** i promovisati pluralizam i tehnološke i medijske inovacije.

14. Angažovati se na **multilateralnom nivou** kako bi se osigurala zaštita ljudskih prava u razvoju i primeni VI za priređivanje, deljenje i moderaciju onlajn sadržaja.

Uvod: Poštovanje principa Helsinškog završnog akta u digitalnom dobu

Prošle godine se navršilo 45 godina od potpisivanja Helsinškog završnog akta iz 1975. godine. Taj akt, rezultat Prvog samita šefova država i vlada KEBS-a, postao je kamen temeljac evropskog političkog poretka. Države istoka i zapada zajedno su se dogovorile o deset principa kojim će se voditi u svom delovanju, uključujući poštovanje suverene jednakosti i, u principu VII, poštovanje ljudskih prava i osnovnih sloboda. Helsinški završni akt dalje sadrži obaveze o saradnji između država, uključujući naučnu i tehnološku saradnju. Čak i računari imaju ulogu u Helsinškom završnom aktu: saradnja se smatra neophodnom, posebno u pogledu razvoja „telekomunikacionih i informacionih sistema; tehnologija povezanih sa računarima i telekomunikacijama, uključujući njihovu upotrebu za [...] automatizaciju, za proučavanje ekonomskih problema, u naučnom istraživanju i za prikupljanje, obradu i širenje informacija".[1]

Saradnja i multilateralni pristupi su potrebni više nego ikada, a novi akteri koji oblikuju način na koji se informacije obrađuju, pojačavaju i priređuju zahtevaju nove regulatorne pristupe za izazove sa kojima se suočavaju ljudska prava u današnjem informacionom pejzažu. Dok države snose primarnu obavezu da poštuju, štite i ispunjavaju ljudska prava, internet posrednici, a posebno nekoliko dominantnih platformi za društvene medije,[2] sve više utiču na ostvarivanje tih prava. Na internetu se može videti novi kvazi-normativni poredak koji dovodi u pitanje tradicionalna shvatanja normativnosti.[3] U današnjem digitalnom svetu, ostvarivanje slobode izražavanja se sve više reguliše u privatnim, hibridnim i javnim prostorima koje oblikuju privatne kompanije, države i korisnici u različitim, visoko asimetričnim odnosima moći. Štaviše, ovi

1 Konferencija za evropsku bezbednost i saradnju, Helsinški završni akt, https://www.osce.org/helsinki-final-act.

2 Onlajn platforme ispunjavaju širok spektar funkcija, uključujući skladištenje i prenošenje informacija. To uključuje društvene mreže, pretraživače, oglasne mreže i platforme za e-trgovinu. Ova publikacija se fokusira na onlajn platforme koje prvenstveno karakteriše olakšavanje interakcije na internetu između pojedinaca, kroz ponudu komunikativnog prostora. Neke platforme prvenstveno hostuju i priređuju sadržaj, druge dodatno olakšavaju digitalnu trgovinu. Platforme koje se prvenstveno bave olakšavanjem komunikacije, uključujući i u komercijalne svrhe, obično se nazivaju društvene mreže.

3 Matthias C. Kettemann, The Normative Order of the Internet (Oxford: OUP, 2020).

onlajn ekosistemi su utrli put za nove oblike upravljanja izražavanjem, uključujući i one koje izvode algoritmi i veštačka inteligencija (VI). Uglavnom kvazi-normativni standardi internet posrednika univerzalno određuju kako se upravlja slobodom izražavanja, i po obimu i po intenzitetu. Ovo upravljanje sadržajem je obično van svake javne kontrole i često se vrši netransparentnim masovnim automatizovanim donošenjem odluka, bez garancije usklađenosti sa međunarodnim okvirom ljudskih prava.

Mnoge grupe za ljudska prava već godinama dižu uzbunu, ukazujući na stalna kršenja ljudskih prava koja su rezultat netransparentnog automatizovanog donošenja odluka. Internet posrednici, kao što su društvene mreže, postali su neophodni za privatnu komunikaciju i javni diskurs, a njima upravljaju algoritmi koji određuju pristup informacijama, a time i proces kreiranja mišljenja. Preovlađujući poslovni modeli najmoćnijih internet posrednika zasnovani su na praćenju, i često iskorišćavaju psihološku osetljivost i druge slabosti pojedinaca. Izgrađeni na temeljima masovnog prikupljanja i analize korisničkih podataka, ovi poslovni modeli su deo tržišnog ekosistema koji je profesorka sa Harvarda, Šošana Zubof (Shoshana Zuboff), označila kao „nadzorni kapitalizam".[4] Dokazi sugerišu da su poslovni modeli zasnovani na praćenju doveli do izobličenja našeg informacionog okruženja na načine koji su u suprotnosti sa pluralizmom, raznovrsnošću i demokratskim procesima i donošenjem odluka. Nedavna otkrića uzbunjivača Frensisa Haugena (Frances Haugen) samo su potvrdila ove navode, naglašavajući potrebu da države uspostave model upravljanja platformama usredsređen na ljudska prava. Uviđajući potrebu da se osigura zaštita ljudskih prava, mnogi pozivaju na pojačanu državnu regulativu. Međutim, regulisanje internet posrednika, posebno regulisanje njihove upotrebe VI i algoritamskih sistema sa ciljem ublažavanja njihovih društvenih rizika, izazovno je i višestrano.

Generalno, postoji vrlo malo primera dobre prakse upravljanja sadržajem u skladu sa ljudskim pravima, a neke dobrovoljne obaveze za poboljšanje zaštite ljudskih prava koje su obećali internet posrednici su se na kraju pokazale nedovoljnim. Stoga je došlo vreme da se krene ka principima

4 Ranking Digital Rights, It's the Business Model: How Big Tech's Profit Machine is Distorting the Public Sphere and Threatening Democracy (2021).

onlajn ekosistema usredsređenim na ljudska prava i da se, u tom pogledu, doprinese poštovanju principa Helsinškog završnog akta u digitalnom dobu. Takav akt bi ponovo mogao da ujedini one koji vide internet kao produžetak svojih nacionalnih granica i one koji žele da vode politiku koja je više fokusirana na ljudska prava. Time bi se ponovo afirmisao sam temelj OEBS-a: ljudska prava kao sastavni deo njegove sveobuhvatne bezbednosti, onlajn i oflajn.

Cilj projekta Veštačka inteligencija i sloboda izražavanja u fokusu (SAIFE) je da pruži smernice državama članicama OEBS-a o tome kako da ispune svoju pozitivnu obavezu zaštite ljudskih prava pojedinaca prilikom kreiranja regulatornih odgovora na nove izazove sa kojima se suočava pravo na slobodu izražavanja u digitalnom dobu. Organizovane su četiri stručne radionice kako bi se identifikovao stvarni i predvidivi negativan uticaj koji automatizovane metode i metode zasnovane na veštačkoj inteligenciji za otkrivanje, procenu, prireĎivanje i personalizaciju onlajn sadržaja imaju na ljudska prava pojedinaca. Stručne radionice su stavile akcenat na individualno pravo na slobodu izražavanja i mišljenja, kao i prava na društvenom nivou, uključujući slobodu medija. Radionice su rezultirale skupom preporuka usredsređenih na ljudska prava sa ciljem da se identifikuju adekvatne mere za ljudska prava i proceduralne mere zaštite za rešavanje individualnih i društvenih rizika koji proističu iz neopravdane upotrebe VI u upravljanju sadržajem.

Struktura i rezime

U okviru projekta SAIFE, Kancelarija predstavnice OEBS-a za slobodu medija, zajedno sa Access Now, organizovala je četiri stručne radionice u prvoj polovini 2021. godine. Ove stručne radionice su razotkrile i analizirale glavne izazove koje alati VI postavljaju pred ljudska prava, posebno pravo na slobodu izražavanja i mišljenja, kao i slobodu medija i pluralizam. Radionice su se fokusirale na četiri glavna tematska pitanja:

- Moderacija sadržaja – bezbednost
  Alati zasnovani na veštačkoj inteligenciji primenjeni u predstavljanju sadržaja za otkrivanje i procenu nelegalnog sadržaja onlajn,

uključujući bezbednosne pretnje, kao što su ekstremistički i teroristički sadržaji.

- **Moderacija sadržaja – govor mržnje**
  Alati zasnovani na veštačkoj inteligenciji koji se koriste za otkrivanje i procenu potencijalno štetnog, ali legalnog sadržaja, sa posebnim fokusom na govor mržnje onlajn i algoritamsku diskriminatornu pristrasnost.

- **Priređivanje i deljenje sadržaja – medijski pluralizam**
  Alati zasnovani na veštačkoj inteligenciji dizajnirani za priređivanje, deljenje i personalizaciju onlajn sadržaja, sa fokusom na sisteme za preporuke sadržaja i njihov uticaj na medijski pluralizam.

- **Priređivanje i deljenje sadržaja – praćenje**
  Alati zasnovani na veštačkoj inteligenciji koji se koriste u oglašavanju zasnovanom na praćenju i njihova veza sa priređivanjem i deljenjem sadržaja kroz profilisanje pojedinaca i predviđanje budućeg ponašanja.

Ovaj izveštaj sadrži glavne nalaze ovih stručnih radionica, kao i preporuke za politike upućene državama članicama OEBS-a, uz priznavanje da je za delotvorno i održivo rešavanje složenih izazova koje moderacija sadržaja i priređivanje i deljenje sadržaja postavljaju pred slobodu izražavanja neophodno uključivanje brojnih zainteresovanih strana. Preporuke za države članice OEBS-a iznete su tokom radionica i razmotrene od strane renomiranih stručnjaka iz oblasti slobode izražavanja, medijskog pluralizma i veštačke inteligencije. Ova publikacija je zasnovana na izveštajima o rezultatima svake stručne radionice. Izveštaje o rezultatima zajedno su izradili predsedavajući imenovan da vodi rad pojedinačnih ekspertskih grupa, izvestioci odgovarajućih stručnih radionica i Eliška Pírková iz partnera projekta, Access Now, a uključivali su i konsultacije sa svim ekspertima i posmatračima učesnicima dotičnih radionica. Izveštaj prati strukturu tematskih oblasti kojima se bavila svaka ekspertska grupa. On može da se posmatra kao četiri odvojena segmenta, od kojih svaki daje preporuke o politici usredsređenoj na ljudska prava za države članice OEBS-a.

VI u moderaciji sadržaja

Rezultati prve dve stručne radionice, fokusirane na upotrebu veštačke inteligencije u moderaciji sadržaja za targetiranje nezakonitog sadržaja i potencijalno štetnog sadržaja, kao što je govor mržnje, spojeni su u jedan zajednički odeljak. Ovaj odeljak pruža skup preporuka o politikama koje imaju za cilj da pomognu u sprečavanju negativnog uticaja koji alati veštačke inteligencije za moderaciju sadržaja imaju na pravo na traženje, primanje i prenošenje informacija i ideja svih vrsta.

Moderacija sadržaja – bezbednost

Alati zasnovani na veštačkoj inteligenciji primenjeni u moderaciji sadržaja za otkrivanje i procenu nelegalnog sadržaja na mreži, uključujući bezbednosne pretnje, kao što su ekstremistički i teroristički sadržaji

Jedna od dve radne grupe koje su se bavile moderacijom sadržaja fokusirala se na automatizovane sisteme zasnovane na veštačkoj inteligenciji koji se koriste za otkrivanje i reagovanje na nezakonit sadržaj i naloge povezane sa širenjem takvog sadržaja. Ova praksa uključuje tehnologije filtriranja i poklapanja hešova koje se primenjuju da blokiraju učitavanje (upload), kao i alate za uklanjanje ili snižavanje ranga (de-rank) sadržaja ex post, često sa prekograničnim efektom. Značajni izazovi se pojavljuju kada se tehnologije veštačke inteligencije koriste za praćenje nacionalnog zakona ili čak da bi se omogućilo praćenje digitalnih komunikacija ljudi od strane organa za sprovođenje zakona pod opravdanjem bezbednosti i javne sigurnosti. Pojedinačna i grupna anonimnost mogu da budu pod posebnim pritiskom, što može da dovede do zastrašujućih efekata na slobodu izražavanja i slobodu medija, kao i na bezbednost novinara. Iako je uticaj moderacije sadržaja zasnovanog na veštačkoj inteligenciji na nezakonito ponašanje i dalje nejasan, VI tehnologije su slepe za kontekst i sklone su prešrokoj primeni pravila koja žele da nametnu. To znači da one redovno generišu takozvane lažne pozitivne i lažne negativne rezultate u identifikaciji verovatno nelegalnog onlajn sadržaja. Rezultat mogu da budu proizvoljna ograničenja legitimnog izražavanja ili neuspeh da se ograniči nelegalno izražavanje.

Radna grupa je istakla potencijalni negativan uticaj koji korišćenje alata zasnovanih na veštačkoj inteligenciji u moderaciji sadržaja ima

na slobodu izražavanja pojedinaca, i šire društvene rizike koje oni predstavljaju za slobodu medija, demokratiju i vladavinu prava. Preporuke o politikama koje je predstavila radna grupa koja istražuje moderaciju sadržaja i nelegalni sadržaj omogućavaju državama članicama OEBS-a da identifikuju, analiziraju i procene značajne sistemske rizike koji proizilaze iz sistema za moderaciju sadržaja, uključujući kada se koriste za sprečavanje brzog širenja nelegalnog onlajn sadržaja. Ove preporuke su kombinovane sa preporukama radne grupe za legalan ali štetan sadržaj, uključujući govor mržnje, kako bi se dale preporuke o merama zaštite slobode govora za veštačku inteligenciju u moderaciji sadržaja, kao i smernice za transparentnost, pristup podacima, nezavisni nadzor, pravne lekove i okvire za dužnu pažnju prema ljudskim pravima.

Rad ove ekspertske grupe i izradu ovog dela izveštaja vodio je predsedavajući **prof. Martin Šejnin (Martin Scheinin)**, a podržali izvestioci **prof. Matijas Keteman (Matthias Kettemann)** i **Marlena Visniak (Marlena Wisniak)**.

Moderacija sadržaja – govor mržnje

Alati zasnovani na veštačkoj inteligenciji koji se koriste za otkrivanje i procenu potencijalno štetnog, ali legalnog sadržaja, sa posebnim fokusom na govor mržnje onlajn i algoritamsku diskriminatornu pristrasnost

Druga radna grupa za moderaciju sadržaja bavila se stvarnim i predvidivim negativnim uticajem koji automatizovani alati zasnovani na veštačkoj inteligenciji za otkrivanje i procenu govora mržnje na mreži imaju na ljudska prava pojedinaca, sa naglaskom na pravo marginalizovanih grupa na slobodu izražavanja i mišljenja. Uticaj diskriminatorne pristrasnosti može se manifestovati kao „pristrasna cenzura“ prema sadržaju koji objavljuju članovi određenih društvenih grupa koje su često na meti izražavanja mržnje i zlostavljanja onlajn. Iako je sam govor mržnje veoma zavisan od konteksta i teško ga je automatski otkriti i ukloniti, grupe koje će verovatno biti meta onlajn zlostavljanja mogu da budu ućutkane pošto je njihova sopstvena komunikacija cenzurisana. Skupovi podataka se koriste za treniranje automatizovanih alata za identifikaciju i razlikovanje različitih kategorija sadržaja. Ako ovi skupovi podataka ne uključuju primere govora na različitim jezicima i iz različitih zajednica, ili ako određene grupe nisu predstavljene u podacima za treniranje, to

može da dovede do pogrešnih klasifikacija koje nesrazmerno utiču na marginalizovane grupe. Automatizovane alatke mogu ili da propuste sadržaj koji potencijalno izražava mržnju (lažno negativno) ili da pogrešno označi legitimne izraze kao govor mržnje (lažno pozitivno).

Zajednička preporuka o zaštiti slobode govora za veštačku inteligenciju u moderaciji sadržaja pruža smernice za transparentnost, pristup podacima, nezavisni nadzor, pravne lekove i okvire za dužnu pažnju prema ljudskim pravima. Konkretne preporuke radne grupe za „govor mržnje" imaju za cilj da omoguće identifikaciju i rešavanje sistemskih rizika, posebno za marginalizovane grupe, koji proizilaze iz sistema za moderaciju sadržaja zasnovanih na veštačkoj inteligenciji koji su angažovani da otkriju potencijalno štetan sadržaj, kao što je govor mržnje. Ove preporuke daju smernice za automatizovane alate za moderaciju sadržaja prilagođene ljudskim pravima, kao i za povećanje digitalnog učešća marginalizovanih grupa u javnom diskursu.

Rad ove ekspertske grupe i izradu ovog dela izveštaja vodila je predsedavajuća, **prof. Lorna Vuds (Lorna Woods)**, a podržali izvestioci, **Emi Bevensi (Emmi Bevensee)** i **Kejti Pentni (Katie Pentney)**.

VI u prireĎivanju i deljenju sadržaja

Deo A: PrireĎivanje i deljenje sadržaja – pluralizam medija

Alati zasnovani na veštačkoj inteligenciji dizajnirani za prireĎivanje i deljenje i personalizaciju onlajn sadržaja, sa fokusom na sisteme za preporuke sadržaja i njihov uticaj na medijski pluralizam

U prvom delu odeljka o prireĎivanju i deljenju sadržaja analizira se negativan uticaj algoritamskih sistema za preporuku sadržaja na ljudska prava pojedinaca, sa naglaskom na apsolutnom pravu na slobodu mišljenja, kao i na medijski pluralizam i slobodu medija. On se bavi: povećanjem potencijalno štetnog sadržaja, kao što je obmanjujući, polarizujući sadržaj ili sadržaj koji izaziva mržnju; uticajem sistema preporuka na različitost

mišljenja i ideja; uticajem algoritamskog prireĎivanja i deljenja na pravo na formiranje mišljenja i medijski pluralizam; i rizikom od polarizacije društava. Algoritamski odabir sadržaja zasnovan je na politikama posrednika, koje prate svoje interne ekonomske interese i interese oglašivača, a ne fokusiraju se na tačnost, raznovrsnost ili javni interes (kao što je vrednost vesti). Ovakav pristup utiče na javnu komunikaciju i slobodan protok informacija, a istovremeno vrši pritisak na profesionalno novinarstvo kanalisanjem novca od reklama posrednicima. Pored toga, vestima se pristupa ređe nego paketu celokupne ponude pojedinačnog informativnog sadržaja, tako da se svaki pojedinačni post bori za pažnju u news feed-u, što podstiče upotrebu klikbejta (mamca za klikove) za angažovanje korisnika. Iako ovaj model olakšava oglašavanje i ostvaruje profit za posrednike, on predstavlja izazov za medijski pluralizam.

Nakon opisa izazova, ovaj deo iznosi skup preporuka o politikama za države članice OEBS-a kako bi se osigurala značajna transparentnost internet posrednika, povećalo individualno posredovanje i kontrola, zajedno sa preporukama za promovisanje raznovrsnosti glasova, informacija od javnog interesa i medijskog pluralizma.

Rad ove ekspertske grupe i izradu ovog dela izveštaja vodila je predsedavajuća, prof. **Kristina Rozgoni (Krisztina Rozgonyi)**, a podržali izvestioci, **Lucien Haic (Lucien Heitz)** i **Bojana Kostić**.

Deo B: PrireĎivanje i deljenje sadržaja – praćenje

Alati zasnovani na veštačkoj inteligenciji koji se koriste u oglašavanju zasnovanom na praćenju i njihova veza sa prireĎivanjem i deljenjem sadržaja kroz profilisanje pojedinaca i predviĎanje budućih ponašanja

Drugi deo odeljka o prireĎivanju i deljenju sadržaja fokusira se na vezu između prireĎivanja i deljenja sadržaja i oglašavanja. VI u targetiranom oglašavanju odnosi se na praksu usmeravanja specifičnih reklama na pojedince na osnovu upotrebe automatizovane statistike – npr. mašinsko učenje, obrada prirodnog jezika (NLP), prepoznavanje govora i prepoznavanje slike. Različiti oblici eksploatacije podataka, uključujući psihološko profilisanje i nano-targetiranje, omogućeni su obradom

podataka, ekstrakcijom signala i automatizovanom analizom širokog spektra različitih tipova podataka – kao što su sadržaj koji generiše korisnik, podaci o lokaciji, obrasci ponašanja, psihografija, informacije o rasi, ekonomskom statusu, polu, starosti, generaciji, nivou obrazovanja, nivou prihoda i zaposlenju korisnika. Kratkoročni i dugoročni, kao i direktni i indirektni efekti ovog oglašavanja zasnovanog na praćenju na ljudsko ponašanje, blagostanje i društvo u celini nisu poznati, ali sistemi zasnovani na veštačkoj inteligenciji su više puta proizvodili pristrasne i pogrešne rezultate.

Ovaj deo izveštaja analizira dalekosežne uticaje koje automatizovani procesi zasnovani na veštačkoj inteligenciji koji se koriste za oglašavanje zasnovano na praćenju imaju na lične interakcije pojedinaca, komunikaciju i učešće u demokratskim debatama. Od kršenja privatnosti do fragmentacije informativnog prostora, oglašavanje zasnovano na praćenju može ozbiljno da naškodi pravu na slobodno formiranje i zadržavanje mišljenja, kao i na traženje, primanje i prenošenje informacija. Radna grupa koja je izradila ovaj deo izveštaja bavi se pitanjima kao što su: inherentni nedostatak objašnjivosti i transparentnosti algoritamskih sistema koji se hrane ličnim podacima i podacima o ponašanju pojedinaca; manipulativne marketinške prakse koje iskorišćavaju konkretne karakteristike i ranjivosti korisnika kako bi se povećala ubedljivost poruke; diskriminacija uzrokovana algoritmima koji optimizuju oglašavanje; i pojačavanje potencijalno štetnog sadržaja u cilju povećanja angažovanja korisnika, radi povećanja profita.

Preporuke zasnovane na ovoj analizi uključuju mere namenjene povećanju transparentnosti i sprečavanju i ublažavanju rizika po ljudska prava koji proizilaze iz praksi kao što su nametljivo targetiranje i personalizacija sadržaja. Preporuke takođe naglašavaju potrebu da se pozabave poslovnim modelima nekoliko dominantnih internet posrednika, koji se zasnivaju na praćenju. Političke preporuke državama članicama OEBS-a uključuju osnaživanje pojedinaca da vrše kontrolu nad svojim podacima i informacijama koje primaju i saopštavaju, kao i bolju zaštitu slobode mišljenja u digitalnom ekosistemu.

Rad ove ekspertske grupe i izradu ovog dela izveštaja vodio je predsedavajući, **prof. Vladan Joler**, a podržali izvestioci, **Holi Sardžent (Holli Sargeant)** i **Julia Haas**.

VI u moderaciji sadržaja

VI u moderaciji sadržaja sa posebnim fokusom na bezbednosne pretnje i govor mržnje

Ovaj deo izveštaja fokusira se na upotrebu veštačke inteligencije u moderaciji sadržaja i implikacije na ljudska prava koje proizilaze iz upotrebe alata veštačke inteligencije za targetiranje određenih kategorija sadržaja koji generišu korisnici. Ističe nedostatke VI u moderaciji sadržaja u kontekstu kako očigledno nezakonitog sadržaja, kao što je teroristički ili ekstremistički sadržaj, tako i potencijalno štetnog, ali legalnog sadržaja, kao što je govor mržnje, posebno iz perspektive marginalizovanih zajednica. Završava se pružanjem operativnih i tehničkih preporuka usmerenih na ljudska prava za države članice OEBS-a. Ove preporuke treba da se bave postojećim negativnim uticajem alata veštačke inteligencije za moderaciju sadržaja na pravo na traženje, primanje i prenošenje informacija i ideja svih vrsta.

1. Definisanje obima moderacije sadržaja

Dve stručne radionice fokusirale su se na upotrebu VI alata u moderaciji sadržaja, prvenstveno na dve kategorije onlajn sadržaja koji generišu korisnici: nezakonit sadržaj i potencijalno štetan, ali legalan sadržaj, sa posebnim naglaskom na govor mržnje. Sledeći odeljci objašnjavaju obim rada ekspertskih grupa u svakoj oblasti.

1.1 Bezbednosne pretnje i nezakonit sadržaj onlajn

Automatizovani alati za otkrivanje usmereni na potencijalno nezakonit sadržaj onlajn – koji se takođe nazivaju proaktivnim merama – u centru su akademske i političke debate. Privatni akteri i kreatori politika često predstavljaju veštačku inteligenciju kao čudotvorno rešenje koje će na kraju biti u stanju da reši veoma složena pitanja oko deljenja nelegalnog sadržaja, uključujući širenje terorističke propagande. Međutim, ovaj pogled na tehnologiju, koji je predstavljen kao opravdanje za povećanje „usvajanja veštačke inteligencije u privredi, kako od strane privatnog tako i od javnog sektora“[5] zanemaruje sistemske rizike uključene u proaktivnu

5 Evropska komisija, Annex to the Communication from the Commission to the European Parliament, the European Council, the Council, the European economic and Social Committee and the Committee of the Regions, “Coordinated Plan on Artificial Intelligence” (7

identifikaciju, otkrivanje i uklanjanje sadržaja koji generiše korisnik. Iako je rešavanje bezbednosnih pretnji legitimno i neophodno, odgovori ne smeju da budu na štetu ljudskih prava. Rizici proizilaze iz automatizovanih sistema za donošenje odluka koje primenjuju onlajn platforme, a često ih nameću države, bilo direktno, kroz pravno obavezujuće zakonodavne okvire, ili indirektno, kroz povećan pritisak na platforme da „učine više".

Bez obzira na konkretan tehnološki metod koji se koristi, takvi automatizovani alati mogu nametnuti prethodna ograničenja prava na slobodu izražavanja i informisanja. U praksi, to znači da oni a priori mogu da isključe određene osobe, grupe, ideje ili sredstva izražavanja iz javnog diskursa. Postoje strogi zahtevi za opravdavanje prethodnih ograničenja slobode izražavanja u međunarodnom okviru ljudskih prava i u različitim ustavnim zakonima. Ti zahtevi proizilaze iz zabrinutosti u pogledu preteranog ograničavanja slobodnog protoka informacija. U tom smislu, alati veštačke inteligencije su posebno zabrinjavajući, jer su ti sistemi zaštićeni od bilo kakvog uvida javnosti, slepi su za kontekst i funkcionišu na veoma netransparentan način koji sprečava bilo kakvu mogućnost delotvornog pravnog leka i pravne zaštite. Dok je prethodna provera sadržaja radi ograničavanja širenja zlonamernog softvera i seksualnog zlostavljanja dece široko prihvaćena kao pozitivna upotreba automatizacije, moramo da budemo oprezni u primeni iste logike na druge tipove govora koji spadaju u širu oblast upravljanja sadržajem.[6]

S obzirom da je veliki broj zakonskih predloga za regulisanje potencijalno nelegalnih onlajn sadržaja nedavno predstavljen širom regiona OEBS-a, rad ekspertske grupe je posebno značajan. Izveštaj o ishodu grupe pruža preporuke usredsređene na ljudska prava za bolju regulaciju alata veštačke inteligencije u moderaciji sadržaja. Namera mu je da pomogne u identifikaciji regulatornih odgovora koji poštuju prava, na širenje i deljenje nelegalnog sadržaja na mreži.

December 2018, COM(2018) 795 final, strana 4, na <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56017>.

6 Emma Llanso, "No amount of 'AI' in content moderation will solve filtering's prior-restraint problem" Big Data & Society 7(1), strana 1-2, na <https://journals.sagepub.com/doi/pdf/10.1177/2053951720920686>.

Iako ovaj izveštaj o rezultatima ne definiše šta predstavlja potencijalno nezakonit sadržaj, njegove preporuke vezane za bezbednost usmerene su na proaktivne metode za otkrivanje i procenu:

- Sadržaj koji je nezakonit, bez obzira na kontekst
  Tipičan primer takvog sadržaja je seksualno zlostavljanje dece, koje je zabranjeno brojnim međunarodnim pravnim instrumentima, kao što su Budimpeštanska konvencija Saveta Evrope, Lanzarote Konvencija, Konvencija 182 Međunarodne organizacije rada, Konvencija UN o pravima deteta, i dr. Međutim, čak i za ovu kategoriju sadržaja nacionalni zakoni ne pružaju jedinstven odgovor.

- Sadržaj koji je deo šireg krivičnog dela
  Na primer, u slučaju video snimaka odsecanja glave koji postaju viralni, najmanje jedan nasilni zločin se dogodio u „stvarnom životu". Svaka inicijativa za moderaciju sadržaja koja ne uzima u obzir oflajn elemente krivičnog dela rizikuje da žrtve ostanu bez pravne zaštite. Osim toga, takav onlajn sadržaj, kao i njegovo uklanjanje, može da utiče na istrage (kao dokaz) i dokumentovanje kršenja ljudskih prava.

- Pravni sadržaj koji je nezakonit zbog konteksta
  Ovo se odnosi na sadržaj koji sam po sebi nije nezakonit, ali način na koji postaje dostupan onlajn može da predstavlja krivično delo. Tipičan primer takvog sadržaja je prikaz golotinje bez pristanka ili neovlašćenog objavljivanja ličnih podataka.

- Sadržaj koji je nezakonit uglavnom zbog svoje namere i efekta
  Ova kategorija uključuje podsticanje na nasilje ili podsticanje na terorizam. Obično, krivično delo ne predstavlja sam sadržaj, već (subjektivna) namera koja stoji iza njegovog objavljivanja, zajedno sa (objektivnim) rizikom da će neki primaoci biti podstaknuti na nasilje. Ova kategorija takođe uključuje, na primer, ksenofobiju, podsticanje na diskriminaciju i podsticanje na mržnju.

1.2 Govor mržnje onlajn

Veliki broj internet posrednika različitih oblika i veličina stvorio je globalno tržište ideja, omogućavajući pojedincima širom sveta da dele i primaju informacije i ideje. Međutim, istovremeno je omogućio i širenje i pojačavanje govora mržnje.[7] Države moraju da se hvataju u koštac sa konkurentnim interesima zaštite slobode govora pojedinaca, istovremeno podržavajući prava i slobode meta i primalaca govora mržnje, kao i javnosti u celini. Konkretno, ostvarivanje ljudskih prava može da bude ograničeno za marginalizovane grupe, koje su podložne diskriminatornoj pristrasnosti i često su ućutkane društvenim pojavama kao što je govor mržnje. Ova manifestacija mržnje nije jedinstvena za onlajn kontekst. Naprotiv, postoji u „stvarnom svetu", u raznim društvima i kroz istoriju, i targetira pojedince i grupe na osnovu prepoznatljivih karakteristika, kao što su rasa, pol/rod, veroispovest i seksualna orijentacija. Međutim, onlajn dimenzija predstavlja nove izazove u pogledu obima, dosega i uticaja govora mržnje. Na primer, Facebook je uklonio više od 20 miliona sadržaja govora mržnje samo u poslednjem kvartalu 2020,[8] dok je Google u istom periodu uklonio skoro 100.000 video snimaka sa YouTube-a.[9] Metama (ili široj javnosti) takođe može da bude teško da izbegnu ili utišaju govor mržnje na mrežama, jer govornici mogu da dopru do tradicionalno bezbednih prostora, uključujući i do domova ljudi, često pod velom anonimnosti, a u nekim slučajevima i kroz koordinisane klevetničke kampanje.[10]

Kao odgovor na ovaj sve prisutniji fenomen i kao posledica zabrinutosti zbog njegovih društvenih uticaja, napori za delotvorniju borbu protiv govora mržnje značajno su porasli u poslednjih nekoliko godina. Fokus je na pitanjima kako se najbolje boriti protiv govora mržnje u onlajn okruženju,

7 Videti, npr. Evropska komisija, Countering illegal hate speech online: 5th evaluation of the Code of Conduct (June 2020) na <https://ec.europa.eu/info/sites/default/files/codeof-conduct_2020_factsheet_12.pdf>.
8 Facebook Transparency, Community Standards Enforcement Report na <https://transparency.facebook.com/community-standards-enforcement#hate-speech>.
9 Google Transparency Report, "Featured policies: Hate Speech" (Oct 2020 —Dec 2020) na <https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech?hl=en>.
10 Matthew Williams i Mischon de Reya, "Hatred Behind the Screens: A Report on the Rise of Online Hate Speech" (2019) strana 18 na <https://hatelab.net/wp-content/uploads/2019/11/Hatred-Behind-the-Screens.pdf>.

diferenciranim ulogama i odgovornostima država i privatnih aktera u moderaciji govora mržnje onlajn i ulozi koju će imati automatizovani sistemi za donošenje odluka u otkrivanju i uklanjanju takvog sadržaja.

Istovremeno, ne postoji univerzalno prihvaćena definicija govora mržnje na međunarodnom nivou. Takav nedostatak definicije ostavio je prostor sudovima i tribunalima da odrede granice dozvoljenog i nedozvoljenog izražavanja. Opseg govora mržnje obuhvata širok spektar izražavanja: od nelegalnog govora mržnje, kao što je podsticanje na genocid, na najokrutnijem kraju spektra; preko potencijalno nezakonitog govora mržnje, kao što su pretnje nasiljem i uznemiravanje; do govora koji ne dostiže prag nelegalnosti ali je ipak štetan i uvredljiv.[11] Proliferacijom onlajn platformi i sadržaja koji generišu korisnici, zadatak definisanja i regulisanja govora mržnje se sve više delegira privatnim kompanijama. Međutim, države zadržavaju krajnju odgovornost za zaštitu ljudskih prava – uključujući slobodu izražavanja, nediskriminaciju i pristup odgovarajućim pravnim lekovima. Od najveće je važnosti da se obezbede odgovarajuće smernice i omogući nadzor kada privatne korporacije intervenišu na digitalnom tržištu ideja.

Sloboda izražavanja zahteva zaštitu ne samo informacija i ideja koje su pozitivno primljene, već i onih koje vređaju, šokiraju ili uznemiravaju.[12] Svako ograničenje prava mora da bude legitimno, proporcionalno i u skladu sa međunarodnim pravom. Dok slučajevi uklanjanja nelegalnog sadržaja mogu biti jasni; položaj je složeniji za sadržaj koji ne prelazi granicu nezakonitosti, ali je štetan i može da utiče na prava drugih. Izazov je definisati ovu drugu kategoriju sadržaja i identifikovati odgovarajuće odgovore na nju. Istorijski marginalizovane grupe u društvu, čiji se glas često ne čuje i koje možda nisu zastupljene u kuloarima moći, česte su mete govora mržnje. Stoga je od ključne važnosti da se obezbedi veće učešće marginalizovanih pojedinaca i grupa u donošenju odluka o ovim fundamentalnim pitanjima, uključujući diskusije o tome kako delotvorno rešavati pitanje govora mržnje. U praktičnom smislu, odsustvo participativnog i reprezentativnog donošenja odluka dovelo je do

11 UN Strategy and Plan of Action on Hate Speech (2020), Table 1, strana 16, dostupno na <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate%20Speech_Guidance%20on%20Addressing%20in%20field.pdf>.
12 Handyside v United Kingdom, App no 5493/72 (ECHR, 7 December 1976) [49]; UN Strategy and Plan of Action, strana 14.

preširokih i nedovoljno inkluzivnih pristupa govoru mržnje, posebno u onlajn okruženju.[13]

Nedostatak razumevanja i inkluzije može dovesti do situacija u kojima se slobodno izražavanje marginalizovanih zajednica neprikladno označava kao govor mržnje, omogućavajući automatizovanom odlučivanju da delotvorno ućutka pojedince i grupe. Do toga može da dođe zbog nerazumevanja konteksta, uključujući dinamiku unutar i van grupe. Na primer, termini „kvir" i „gej" mogu se koristiti kao homofobične ili transfobične uvrede, definisane i regulisane kao govor mržnje; ali podjednako mogu da budu reklamacije od strane članova LGBTQ+ zajednice ili korišćene za „pro-društvene funkcije", kao što su izgradnja zajednica i grupa, i pomoć pojedincima da se bolje pripreme za neprijateljstvo i nose sa njim.[14] Pokazalo se da slična nerazumevanja konteksta i namere dovode do prekomernog uklanjanja sadržaja rasnih manjina na mreži.[15] Regulisanje govora mržnje je nužno kontekstualna vežba – od namere govornika, preko verovatnih efekata govora, do posebnog značenja reči ili slika u datom društveno-političkom kontekstu. Studije su pokazale da automatizovano donošenje odluka jednostavno nije sposobno za ovu kontekstualnu vežbu. Generalizovani ili preterano inkluzivni pristupi, stoga, mogu da dovedu do cenzure pripadnika marginalizovanih grupa, kršenjem njihove slobode izražavanja.[16] Ovaj efekat utišavanja trebalo bi da bude glavna briga i za države i za internet posrednike.

13 Videti npr. Molly K. Land i Rebecca J. Hamilton, "Beyond Takedown: Expanding the Toolkit for Responding to Online Hate" in Predrag Dojcinovic (ed.) Propaganda, War Crimes Trials and International Law: From Cognition to Criminality 143 (Routledge, 2020), strana 2, dostupno na <https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3514234_code858831.pdf?abstractid=3514234&mirid=1>.
14 Thiago Dias Oliva, "Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online" (2021) Sexuality & Culture 25, strana 705-7 dostupno na <https://www.researchgate.net/publication/345501707_Fighting_Hate_Speech_Silencing_Drag_Queens_Artificial_Intelligence_in_Content_Moderation_and_Risks_to_LGBTQ_Voices_Online>.
15 Thomas Davidson, Debasmita Bhattacharya i Ingmar Weber, "Racial Bias in Hate Speech and Abusive Language Detection Datasets" (2019) na <https://www.aclweb.org/anthology/W19-3504.pdf>; Maarten Sap et al, "The Risk of Racial Bias in Hate Speech Detection" (2019) na <https://homes.cs.washington.edu/~msap/pdfs/sap2019risk.pdf>.
16 Ibid.

Tamo gde su politike govora mržnje nedovoljno inkluzivne – to jest, gde ne uspevaju da se bave govorom koji je zakonit, ali štetan – onlajn prostori mogu da postanu nesigurno ili nepoželjno okruženje za pripadnike marginalizovanih grupa, koje ih efektivno potiskuje. Ovo je posebno problematično u svetlu važne uloge koju ova onlajn okruženja igraju na našem novom (digitalnom) tržištu ideja, kojem se pojedinci sve više okreću da dele ideje, konzumiraju vesti i učestvuju u javnoj debati. Rezultat može da bude „demokratski deficit", pri čemu pojedinci iz marginalizovanih grupa – žene i nebinarne osobe, rasne i etničke manjine, pripadnici LGBTQ+ zajednice, itd. – ne mogu ili ne žele da u potpunosti učestvuju u demokratskom diskursu.[17] Pored toga, politike mogu da budu nedovoljno inkluzivne jer ne uzimaju u obzir intersekcionalnost – to jest, govor mržnje usmeren na pojedince ili grupe na osnovu dva ili više identifikacionih faktora.[18]

Što se tiče preporuka usmerenih na govor mržnje, treba imati na umu sledeće komentare koji se tiču obuhvata:

- Iako govor mržnje nije definisan, ovaj izveštaj se fokusira na zakonit, ali štetan govor mržnje u onlajn okruženju. Preporuke su prilagođene regulisanju i moderaciji takvog govora na način koji je u skladu sa ljudskim pravima.

- U svetlu nesrazmernih uticaja moderacije govora mržnje na marginalizovane zajednice, izveštaj daje prilagođene preporuke državama članicama OEBS-a da obezbede zaštitu marginalizovanim pojedincima i grupama – uključujući njihovo pravo na slobodu izražavanja, nediskriminaciju i pristup adekvatnim pravnim lekovima.

- Iako se moderacija sadržaja dešava na različitim nivoima – što je detaljnije objašnjeno u nastavku – ovaj izveštaj se prvenstveno fokusira na opsežnu ili „industrijsku" moderaciju govora mržnje, kako bi se prikazao stepen i obim njegovog uticaja na slobodu izražavanja.

17 Nani Jansen Reventlow, "The power of social media platforms: who gets to have their say online?" Lilith (4 February 2021) na <https://www.lilithmag.nl/blog/2021/2/3/the-power-of-social-media-platforms-who-gets-to-have-their-say-online>.
18 UN Strategy and Plan of Action, strana 28.

2. Uputstvo o moderaciji sadržaja

Otkrivanje i moderacija sadržaja koji je nezakonit, ili potencijalno štetan, ali zakonit, težak je zadatak od početka do kraja. Iako je od ključne važnosti da se internet posrednici pozivaju na odgovornost, razumevanje ograničenja same tehnologije (kao i uključenih poslovnih modela) pomaže javnim organima da preduzmu efektnije korake u vezi sa korporacijama u sektoru društvenih mreža. Ovaj odeljak daje pregled algoritamskih alata i tehnika za moderaciju sadržaja pre nego što identifikuje nekoliko osetljivih tačaka u odgovarajućoj moderaciji sadržaja.

Kada se koriste VI alati za moderaciju sadržaja, često nedostaje adekvatno opravdanje i obrazloženje odluka. To znači da korisnici često ne znaju zašto je doneta automatizovana odluka i koje su specifične informacije unete, koje su dovele mašinu do donošenja određene odluke.

Vrste moderacije sadržaja

Postoje tri osnovna modela moderacije sadržaja, kako ih je definisala Robin Kaplan (Robyn Caplan):

- **Interni**
  Moderacija sadržaja od strane malih internih timova moderatora pojedinaca.

- **Zavisan od zajednice**
  Moderacija sadržaja se pretežno oslanja na volontere-moderatore zajednice iz različitih podsekcija posrednika, kao što čine Wikipedia ili Reddit.

- **Industrijski**
  Moderacija sadržaja koje uključuje obiman angažman spoljnih saradnika za moderaciju koje vrše ljudi, u kombinaciji sa vlasničkom automatizovanom detekcijom mašinskog učenja.[19]

19 Robyn Caplan, Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches (2018), na <https://datasociety.net/library/content-or-context-moderation/>.

Ovaj izveštaj se prvenstveno fokusira na uticaje i razmatranja trećeg tipa moderacije, „industrijske".

Alati za moderaciju sadržaja mogu se podeliti u sledeće kategorije:

- **Otkrivanje**
  Lociranje i prepoznavanje sadržaja koji može da ugrozi politike internet posrednika.

- **Odlučivanje**
  Utvrđivanje da li otkriveni sadržaj zaista krši politike posrednika.

- **Sprovođenje**
  Delovanje prema sadržaju na osnovu posledica navedenih u politikama posrednika.

- **Žalba**
  Povratak na fazu donošenja odluke ako korisnik ospori ili uloži žalbu na odluku posrednika.

- **Politika**
  Skup principa, pravila ili smernica koji određuju koji sadržaj je prihvatljiv na platformi posrednika. U praksi, ove smernice se revidiraju i ažuriraju na osnovu drugih komponenti procesa moderacije sadržaja.[20]

Kada se utvrdi da je sadržaj nelegalan ili da krši uslove pružanja usluge posrednika – ili kada se predviđa da spada u jednu od ovih kategorija – postoji nekoliko mogućih ishoda. Najčešći su označavanje sadržaja (flagging) ili brisanje. U slučaju brisanja, sadržaj se odmah uklanja i ponekad se sprečava da se ponovo učita. Osim uklanjanja sadržaja, postoji niz ex post alata koji su dostupni za rešavanje „problematičnog" sadržaja.

20 Meedan, Content Moderation Toolkit: Toolkit for Civil Society and Moderation Inventory, na <https://meedan.com/reports/toolkit-for-civil-society-and-moderation-inventory/>.

Neki primeri uključuju:

- **Demonetizacija sadržaja:**
  Na platformama kao što su YouTube i Twitch, gde kreatori mogu da profitiraju od popularnosti svog sadržaja, Uslovi korišćenja usluge se mogu primeniti na način da onemoguće korisnicima da ostvaruju profit od određenih vrsta sadržaja. Iako takva demonetizacija može da ima prednosti, ona se često nesrazmerno primenjuje protiv marginalizovanih ljudi, bilo zbog prethodno navedenih izazova algoritma ili iz razloga namernog ućutkivanja.

- **Uklanjanje prioriteta i derangiranje sadržaja**
  Ono što korisnik vidi na onlajn platformama generalno kontroliše niz privatnih algoritama dizajniranih da povećaju angažovanje. Internet posrednici mogu da odbace ili uklone istaknutost uvredljivih ili štetnih naloga. Iako ovo može da bude korisno za borbu sa dometom štetnog sadržaja, kao što su govor mržnje ili dezinformacije, često se pretvara u podizanje ranga već popularnog sadržaja, kao što su konvencionalne (mainstream) vesti, potencijalno na račun marginalizovanih glasova.

- **Suspenzija naloga ili ograničenje funkcija**
  Privremene suspenzije destimulišu korisnike da krše smernice zajednice, bez trajne zabrane. Iako takve suspenzije mogu da spreče dodatni govor mržnje, mogu se koristiti i protiv marginalizovanih ljudi koji pokušavaju da kreiraju bezbedan prostor u svom uglu interneta.

- **Uklanjanje naloga**
  Uklanjanje celog naloga može da poremeti sposobnost korisnika da održi veliku bazu pratilaca i stoga može da bude posebno delotvoran odgovor za serijske prestupnike. Međutim, kako su nedavna uklanjanja na Telegramu pokazala, ekstremističke zajednice se brzo prilagođavaju i ponovo kreiraju kanale i pratioce nakon uklanjanja.

- **Blokiranje/Isključivanje zvuka/Uklanjanje prijatelja**
  Ove opcije predstavljaju vid subjektivne moderacije koji omogućava korisnicima da odaberu koji sadržaj ne žele da vide na svom ličnom feed-u. Na platformama društvenih mreža povezanih sa Secure-Scuttlebutt, blokiranja su transparentna na način koji šalje poruku poverenja i nepoverenja kao alat za ograničavanje širenja neželjenih poruka.[21]

21 Za više detalja, pogledajte društvenu mrežu Scuttlebutt, decentralizovanu platformu, na <https://scuttlebutt.nz/>.

Industrijska algoritamska moderacija sadržaja

Algoritamska moderacija sadržaja uključuje niz tehnika iz statistike i računarstva koje se razlikuju po složenosti i efikasnosti. Sve ove tehnike su dizajnirane da identifikuju, upare, predvide ili klasifikuju sadržaj koji generiše korisnik na osnovu njegovih konkretnih svojstava ili opštih karakteristika. Internet posrednici primenjuju automatizovane alate za čuvanje sadržaja u velikom broju slučajeva, uključujući terorizam, eksplicitno nasilje, „toksični govor", golotinju bez pristanka, zlostavljanje dece i otkrivanje spem sadržaja (spam). Pretežno se koriste dve vrste algoritamske moderacije sadržaja, ali ne isključivo, za borbu protiv potencijalno nelegalnog sadržaja onlajn: analiza teksta i analiza slike. Kada alati veštačke inteligencije označe određeni deo sadržaja kao potencijalno nelegalan, on se obično stavlja na čekanje, ili mu se daje prioritet da ga pregleda ljudski „stručni" moderator. Zatim se može izbrisati ili rešavati pomoću jednog od prethodno navedenih ex post alata.

Sistemi za mašinsko učenje koji sprovode analizu teksta redovno primenjuju obradu prirodnog jezika (NLP). NLP sistemi raščlanjuju tekst na sveobuhvatan način, pokušavajući da analizu približe ljudskom razumevanju teksta o kome je reč. NLP alati su obučeni da predvide da li određeni tekst prenosi pozitivne ili negativne emocije (tzv. analiza sentimenata), a samim tim i da klasifikuju da li pripada ili ne pripada određenoj kategoriji sadržaja koji generiše korisnik. NLP je dizajniran da predviđa ishode na osnovu označenih instanci, na primer „uvredljivo" ili „neuvredljivo". Najpoznatiji primer NLP alata je Google/Jigsaw's Perspective API, komplet alata otvorenog koda koji omogućava operatorima vebsajtova, istraživačima i drugima da koriste svoje modele mašinskog učenja za procenu „toksičnosti" posta ili komentara.

Automatsko otkrivanje i identifikacija slika i video zapisa, s druge strane, često uključuje otkrivanje sadržaja koji je ranije identifikovan kao nelegalan, uz istovremeno otkrivanje novih sadržaja koji bi se mogli dodati u kategoriju nelegalnih. Tehnologije detekcije i identifikacije slike koriste takozvane heš vrednosti. Heš je jedinstvena numerička vrednost, koja se takođe naziva „digitalni otisak", koja se generiše specifičnim algoritmom pokrenutim na datoteci slike. Jednostavna tehnologija heširanja procenjuje dimenziju slike ili vrednosti boje piksela. Jednostavna izmena piksela slike u potpunosti menja heš takve datoteke, što znači da je alatku lako zaobići. Iznijansiraniji alati koriste perceptivno heširanje koje uključuje otiske

slika i video zapisa pomešanih sa drugim karakteristikama sadržaja, kao što je, na primer, herc-frekvencija tokom vremena u zvučnom zapisu. Perceptivni hešovi su robusniji i mogu da identifikuju slike i video zapise čak i nakon njihove izmene. Tipičan primer perceptivnog heširanja je PhotoDNA, koji je razvio Microsoft i koristi se za borbu protiv zlostavljanja dece onlajn.

Nakon zločina u Krajstčerču na Novom Zelandu 2019. godine, Facebook, Google, Twitter i Microsoft su kreirali Globalni internet forum za borbu protiv terorizma (GIFCT), organizaciju osnovanu kao deo njihove posvećenosti povećanju dobrovoljnog poštovanja EU Kodeksa ponašanja u borbi protiv nelegalnog govora mržnje onlajn. Unutar GIFCT okvira, četiri kompanije dele najbolje prakse za razvoj svojih algoritamskih alata za moderaciju sadržaja. One takođe vode veoma tajnu i netransparentnu heš bazu podataka terorističkog sadržaja, u kojoj jedne sa drugima dele digitalne otiske „nedozvoljenog sadržaja", uključujući slike, video, audio i tekstualne zapise. Na primer, u roku od nekoliko sati od napada u Krajstčerču, Facebook je otpremio hešove približno 800 različitih verzija video snimka napadača. U teoriji, svaki pojedinačni video zapis koji su preneli korisnici Facebook, YouTube i Twitter naloga sada se može heširati i proveriti u bazi podataka. Sadržaj koji odgovara nekom unosu u bazi podataka biće odmah blokiran. Bazom podataka upravljaju isključivo privatni akteri i van je bilo kakvog javnog nadzora, što dovodi do ozbiljnih izazova za novinarski i umetnički sadržaj.

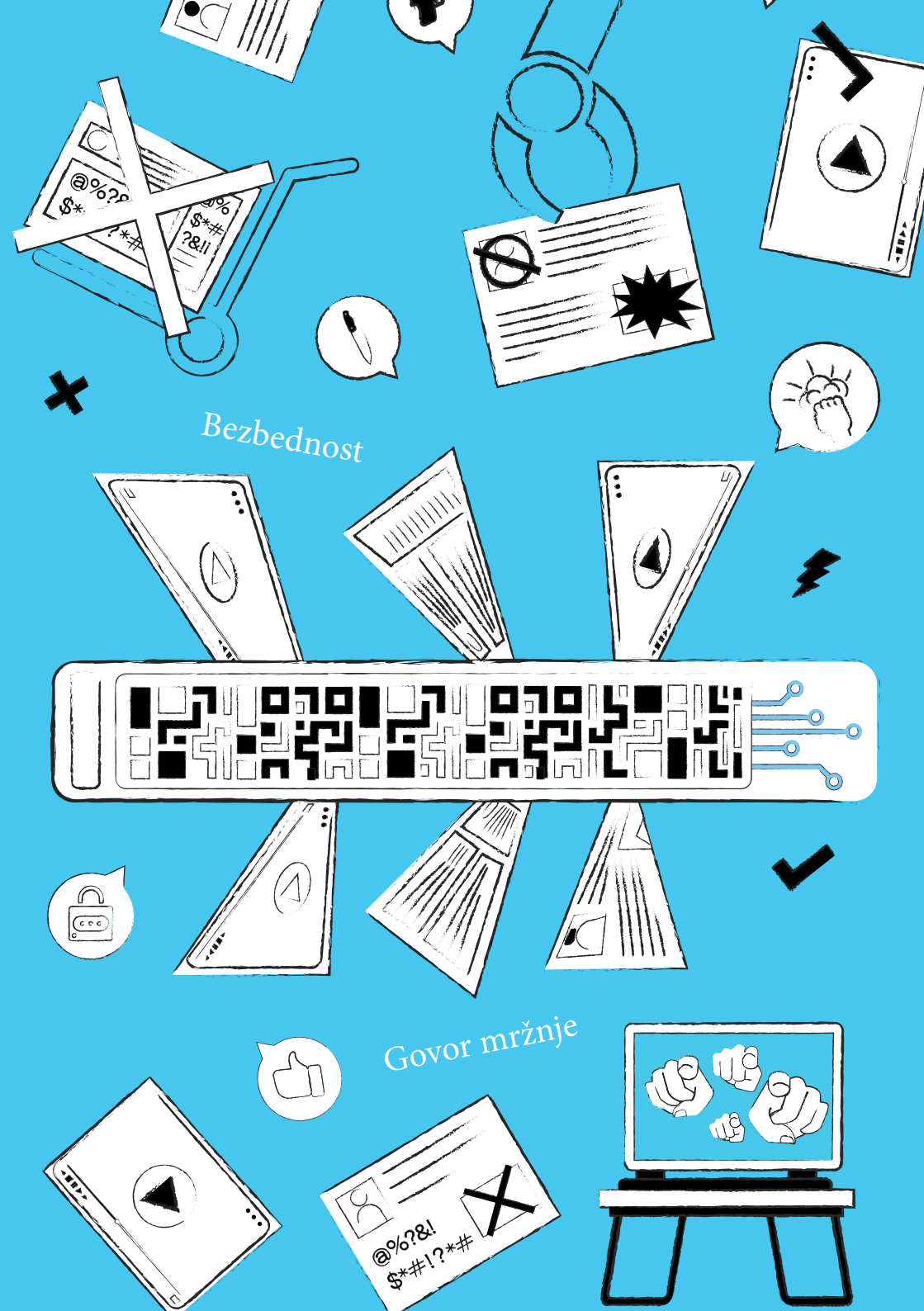
Nedostaci algoritamske moderacije sadržaja

Postoji nekoliko osetljivih tačaka u dizajnu i razvoju algoritamskih alata za moderaciju sadržaja – kao i u poslovnim modelima internet posrednika koji koriste te alate – i to treba da imaju na umu kreatori politika. Jedna velika ranjivost u dizajnu algoritma za mašinsko učenje javlja se kada tim ljudi odlučuje o pravilima za označavanje trening podataka koji će se koristiti u modelu mašinskog učenja. Ovaj korak je presudan jer je VI u osnovi samo mašina za kopiranje. Sistemi veštačke inteligencije uče ono što ih ljudi nauče da uče – pa čak i tada, može da dođe do odstupanja. Predrasude ljudi koji su uključeni u postupak i one koje su ugrađene u same podatke će se replicirati tokom životnog ciklusa VI sistema.

Na primer, kada se sistem koji pokušava da otkrije zločin onlajn oslanja na strukturno rasističke podatke, on će još dublje ukoreniti rasističke rezultate. Dodatni izazovi proizilaze iz suptilnosti govora, ili gore pomenute upotrebe datog termina u grupi i van grupe, što bi moglo da dovede do sistematskog pogrešnog označavanja pojmova, nanoseći štetu onima koje sistem treba da zaštiti. Ne postoji jedno jednostavno VI rešenje za mržnju onlajn, jer je svaki tip mržnje i kontekst zasnovan na identitetu drugačiji i zato što se pejzaž stalno menja kako se protivnici prilagođavaju. Okruženje u kojem se kreiraju i primenjuju sistemi za mašinsko učenje, posebno za nešto tako delikatno kao što je otkrivanje govora mržnje, duboko je dinamično i kontekstualno.

Dok bi transparentnost i participativni procesi za rešavanje ovih izazova mogli značajno da umanje njihove rizike, internet posrednici imaju privatne interese koji su u suprotnosti sa vrstom transparentnosti koja je potrebna. Na primer, algoritam mašinskog učenja za otkrivanje govora mržnje je sam po sebi roba koja se može prodati. Kao takvo, posrednici će verovatno zakonom zaštititi svoje vlasništvo. Osim toga, mnogi posrednici tvrde, bez obzira na stepen istine, da bi deljenje takvih algoritama omogućilo protivnicima da ih zloupotrebe. Uz to, može da bude teško da se objasne specifične vrste odluka za koje se koristi VI alat , ili da se osigura da se VI alat generalizuje kako bi se bavio novim oblicima problema za čije rešavanje je korišćen.

Jasan okvir za moderaciju sadržaja treba da obezbedi da odgovori na nezakonit ili potencijalno štetan sadržaj budu proporcionalni i tačni, istovremeno nastojeći da reši tehničke i društveno-političke probleme u vezi sa moderacijom. Dok države generalno imaju indirektnu kontrolu nad praksama moderacije korporacija, puna svest o postojećim politikama i praksama, kao i alternativnim mogućnostima, pomaže u informisanju i usmeravanju kreiranja politika.


Bezbednost
Govor mržnje
@%?&!
$*#!?*#

3. Preporuke o korišćenju veštačke inteligencije u moderaciji sadržaja, usmerene na ljudska prava

3.1 Preporuke o transparentnosti

Preporuke za algoritamsku transparentnost

- **Države treba da obavežu internet posrednike da obezbede dokumentaciju o VI alatima koje primenjuju za moderaciju sadržaja.** Svako otkrivanje treba da bude razumljivo i dostupno svim korisnicima. Platforme treba da otkriju koji su zaključci izvedeni o ličnim zaštićenim karakteristikama korisnika (tj. starost, rasa, pol, invaliditet) ili njihovom društvenom krugu, članstvu u zajednici i posrednicima (proxies). Platforme treba da dele informacije u vezi sa:

    - Trening podacima: sadržaj i poreklo skupova podataka koji se koriste za algoritme za obuku; metode za obuku VI modela; varijable/funkcije/karakteristike koje utiču na algoritamsko prireĎivanje sadržaja; sistemi preporuka i/ili rangiranja (npr. starost korisnika, pol, itd.) i kolika je kontrola korisnika nad tim varijablama; i procesi oko upravljanja trening podacima (npr. prikupljanje, skladištenje, prethodna obrada/obrada, prenos, zadržavanje).

    - Uslugama obogaćivanja podataka: priprema i čišćenje podataka (označavanje podataka, analiza sentimenata, prepoznavanje slike, validacija govora u tekst, itd.) i zadaci koncepta "human-in-the-loop" koji podrazumeva ljudsku intervenciju (ljudska moderacija sadržaja, razvijanje kontinuirane povratne sprege, validacija algoritamskih izlaza i modela, itd.),uključujući navoĎenje osobe/a koje sprovode uslugu obogaćivanja podataka i informacije o obuci.

    - Procesima i rezultatima testiranja, evaluacije i validacije ovih modela, uključujući merenja kvaliteta i tačnosti.

- **Države treba da nalože dokumentovanje modela specifičnih za sadržaj od strane internet posrednika.** Posrednici bi trebalo da imaju zakonsku obavezu da otkriju kriterijume, parametre i

karakteristike koji se koriste za modele mašinskog učenja namenjene za prireĎivanje i deljenje sadržaja, moderaciju sadržaja i bilo koju drugu analizu podataka ili prepoznavanje obrazaca. To bi trebalo da obuhvata razvrstavanje podataka za modele mašinskog učenja koji su dizajnirani za skidanje i uklanjanje sadržaja koji su generisali korisnici i za modele dizajnirane da pojačaju i ponovo smanje „skrivenu zabranu" (shadow banning) i derangiranje sadržaja. Svako otkrivanje treba da bude razumljivo i dostupno svim korisnicima, uz obezbeĎivanje njihove privatnosti i zaštite podataka.

- **Države treba da obezbede različite skupove podataka, zasnovane na različitim atributima, pošto samo atributi koji se mere i beleže mogu da budu uključeni u trening podatke ili podatke za evaluaciju za algoritam.** Mnogi široko dostupni skupovi podataka se fokusiraju na nepromenljive karakteristike (kao što su etničke grupe) ili karakteristike koje beleže i regulišu vlade (kao što su zakonski rod, novčani prihod ili profesija). Nasuprot tome, karakteristike kao što su seksualna orijentacija i rodni identitet često nisu vidljive. Ovo je ozbiljan izazov za borbu protiv intersekcionalne diskriminatorne pristrasnosti svojstvene nekim algoritamskim sistemima.

- **Države treba da sprovedu transparentnu upotrebu sistema veštačke inteligencije u javnom sektoru, usredsreĎenu na ljudska prava, uključujući korišćenje VI alata za moderaciju sadržaja.** Države treba da uspostave mehanizme za povećanu kontrolu i zahteve za transparentnošću kada javni sektor koristi sisteme veštačke inteligencije za analizu sadržaja – kao što su tehnologije za prepoznavanje lica i praćenje sadržaja koji se deli na onlajn platformama.

- **Države treba da obavežu internet posrednike da obaveste korisnike kada su podvrgnuti automatizovanim procesima i kada se koriste automatizovani sistemi za moderaciju sadržaja trećih strana, a platforme treba da objasne kako takvi mehanizmi funkcionišu.** Platforme treba da pruže detaljne informacije korisnicima o osnovama za uklanjanje, sa posebnim osvrtom na pravilo koje je prekršeno i objašnjenjem mogućnosti da se traži ljudska provera.

- **Države treba da obelodane sve zahteve poslate internet posrednicima i odgovore koje su primile,** i treba da nalože platformama da otkriju da li je bilo koji zahtev države doveo do podešavanja ili promena modela mašinskog učenja koji se koristi za moderaciju potencijalno nelegalnog sadržaja.

- **Države treba da zahtevaju da internet posrednici omoguće istraživačima i organizacijama civilnog društva pristup skupovima podataka i modelima,** kako bi mogli da ih procene i pruže činjeničnu osnovu za istraživanja zasnovana na javnom interesu. Ako je potrebno, mogu se uspostaviti institucionalni odbori za proveru i nezavisni proces akreditacije.

- **Države treba da zahtevaju dokaz o korisnosti upotrebljenih alata za praćenje.** Na primer, od njih bi se moglo zatražiti da detaljno navedu slučajeve upotrebe u kojima je nezakonit sadržaj tačno identifikovan – i gde neautomatizovana sredstva ne bi dovela do istog stepena uspeha. Dokaz o korisnosti je od suštinskog značaja za odlučivanje o neophodnosti intervencije. Na kraju krajeva, samo dokazana korisnosti se može proceniti prema srazmernosti sa nanetom štetom za ljudska prava.

Preporuke za transparentnost usmerenu na korisnika

- **Države treba da obezbede da internet posrednici na odgovarajući način obelodane da je na korisnika uticalo ili će uticati algoritamsko donošenje odluka, uključujući moderacija sadržaja, i da korisnici mogu barem da odustanu od automatizovanog donošenja odluka.** Korisnici moraju da budu u mogućnosti da vrše kontrolu nad alatima za otkrivanje moderacije sadržaja, koji bi u idealnom slučaju trebalo da budu podrazumevano obezbeđeni mehanizmom „opt-in" (sa mogućnošću prihvatanja). Smislena svesnost omogućava pojedinačnim korisnicima da odaberu/isključe automatizovano donošenje odluka, ako to žele. Internet posrednici treba da osmisle saglasnost i politiku privatnosti na način koji olakšava informisan izbor korisnika, u skladu sa zakonima o zaštiti podataka.

- **Države treba da obezbede da korisnici imaju pristup podacima o profilisanju[22] koje internet posrednici imaju o njima, uključujući sve zaključke koji se o njima izvode.** Ovi podaci bi trebalo da budu dostupni korisnicima na zahtev u razumljivom i dostupnom formatu. Korisnici takođe treba da budu u mogućnosti da isprave i izbrišu svoj profil. Iako Opšta uredba o zaštiti podataka o ličnosti (GDPR) u velikoj meri obezbeđuje ovo pravo u Evropskoj uniji, postoji potreba za delotvornim i pristupačnim procedurama ili interfejsima koji omogućavaju pojedincima da lako dobiju ove informacije. Prema tome, države širom regiona OEBS-a trebalo bi da nalažu minimalne standarde za obaveze transparentnosti usmerene na korisnika, kao što je navedeno u članovima 13(2)(f) i 14(2)(g) GDPR-a.

- **Države bi trebalo da zakonski obavežu internet posrednike da daju objašnjenja u vezi sa korišćenim modelima, ulaznim podacima, metrikom učinka i testiranjem njihovog modela mašinskog učenja, na konkretnom, razumljivom jeziku prilagođenom uzrastu.** Takvo objašnjenje će omogućiti korisnicima da osporavaju algoritamsko donošenje odluka i/ili da odustanu. Pravo na protivljenje upotrebi automatizovanih sistema za donošenje odluka trebalo bi da važi čak i ako je čovek uključen u proces.

- **Države bi trebalo da obavežu internet posrednike da pravilno objasne korisnicima algoritamsko donošenje odluka.** Objašnjenje određene odluke trebalo bi da bude dostupno korisnicima kao minimalan uslov, da bi se obezbedila mogućnost osporavanja automatizovanih odluka u moderaciji sadržaja. Objašnjenje treba da bude na razumljivom jeziku i treba da sadrži statističke podatke koji su korišćeni i detaljno objašnjenje posredničke politike koja stoji iza odluke.

22 GDPR definiše profilisanje kao automatizovanu obradu podataka radi analize ili predviđanja o pojedincima; što znači da se „jednostavna procena ili klasifikovanje pojedinaca na osnovu karakteristika" može smatrati profilisanjem, sa ili bez svrhe predviđanja.

Preporuke za zahteve transparentnosti, neophodne za efikasan pristup pravnom leku i pravnoj zaštiti za one koji su meta govora mržnje

- **Države treba da obavežu internet posrednike da daju obrazložene odluke koje objašnjavaju proces i konkretne odluke koje su donete u vezi sa sadržajem koji se smatra govorom mržnje.** Obrazložena odluka da se reaguje na govor mržnje treba da bude dostavljena svim pogođenim korisnicima, uz obrazloženje prava svake zainteresovane strane i jasno formulisana uputstva kako da se žale na odluku. Isto pravilo treba da važi i za obaveštenja o protivtužbama, bilo da su odbijene ili postoji nalaz u korist pružaoca sadržaja (content provider).

- **Države treba da obavežu internet posrednike da čuvaju sve podatke o uklanjanju sadržaja, u skladu sa standardima zaštite podataka.** To uključuje, između ostalog, informacije o tome koja uklanjanja nisu proverili ljudi, da li su korisnici pokušali da ulože žalbu na uklanjanje i slučajeve u kojima je sadržaj prijavljen, ali nije reagovano. Pored toga, tamo gde je to izvodljivo, internet posrednici treba da uključe u svoje izveštaje o transparentnosti informacije i statistike o vrstama govora mržnje na koje su reagovali (na primer, koje zaštićene karakteristike su povređene), proporciji i stopi uspešnih žalbi i odobrenim pravnim lekovima.

- **Države treba da osiguraju da zahtevi za transparentnost za internet posrednike obezbede očuvanje svih sadržaja klasifikovanih kao govor mržnje koji se automatski blokiraju ili uklanjaju,** uključujući pojedinačne postove, video zapise, slike i čitave naloge. U skladu sa zahtevima zaštite podataka i privatnosti, ovaj sadržaj bi trebalo da bude dostupan istraživačima na zahtev, kako bi se obezbedio dodatni nadzor nad mehanizmima pravne zaštite i pravičnosti i delotvornosti žalbenih mehanizama, posebno za marginalizovane grupe.

Preporuke za zahteve transparentnosti neophodne za delotvoran javni nadzor

- **Države treba da prepoznaju i ovlaste imenovana nadzorna tela, sa ekspertizom u oblastima jednakosti i nediskriminacije, da prate**

i rešavaju neujednačene ili diskriminatorne efekte automatskog donošenja odluka na marginalizovane grupe. Ova tela mogu uključivati nacionalne institute za ljudska prava, ombudsmane ili poverenike za informacije i privatnost, i mogu da dopunjavaju rad domaćih sudova. Za javne organe je od ključnog značaja da omoguće i osnaže ova nadzorna tela u ispunjavanju ove uloge dajući im adekvatna i značajna zakonska ovlašćenja, kao i sigurna i dovoljna sredstva.

- **Tela za ravnopravnost treba da su u mogućnosti da vode strateške parnice kako bi osporila diskriminatorne ishode automatizovanih mera.** Ova tela treba da imaju dovoljno sredstava i da imaju tim zaposlenih koji je posvećen ovoj konkretnoj temi i koji radi na povećanju transparentnosti u korišćenju automatizovanih mera.

- **Države treba da obezbede da se obavezni uslovi transparentnosti izveštavanja za internet posrednike fokusiraju na kvalitet,** a ne na kvantitet. Same brojke služe isključivo za poređenje; ne pružaju dragocene informacije o tome kako se internet posrednici bave sadržajem koji generišu korisnici. Zbog toga bi trebalo da se zahteva da izveštaji o transparentnosti internet posrednika sadrže: broj svih primljenih obaveštenja; vrstu subjekata koji su ih izdali, uključujući privatna lica, administrativna tela ili sudove; razloge za utvrđivanje zakonitosti sadržaja ili kako on krši uslove korišćenja internet posrednika; i da li je sadržaj obeležen od strane privatnih lica, automatizovanih alata ili pouzdanih lica za obeležavanje.

- **Države treba da obezbede da zakonski propisano izveštavanje o transparentnosti jasno ukaže koji je metod moderacije sadržaja primenjen:** uklanjanje sadržaja, demonetizacija sadržaja, uklanjanje prioriteta sadržaja, suspenzija naloga, uklanjanje naloga ili bilo koja druga radnja protiv označenog sadržaja ili naloga korisnika.

- **Države treba da postave minimalne zahteve za izveštavanje o transparentnosti, uključujući izveštavanje o:** konkretnim vremenskim okvirima za obaveštavanje pružaoca sadržaja pre preduzimanja bilo kakve radnje; konkretnim vremenskim okvirima za podnošenje protivtužbi; tačnom vremenu koje će proći pre nego

što se sadržaj ograniči; vremenskom okviru za žalbeni postupak; i broju primljenih žalbi i način na koji su rešene.

- **Posebno u vezi sa govorom mržnje, države treba da obavežu internet posrednike da objave broj prijava uvredljivog ili štetnog ponašanja koje dobijaju na godišnjem nivou.** Ovo bi trebalo da obuhvati i podatak koliko od ovih prijava se odnosi na izražavanje mržnje usmereno prema zaštićenim karakteristikama, kao što su rasa, etnička pripadnost, vera ili pol. Posebnu pažnju treba posvetiti intersekcionalnim razmatranjima o načinima na koje se rasa, klasa, pol i druge individualne karakteristike mogu kombinovati u različite načine diskriminatornog postupanja.

- **Države treba da obavežu internet posrednike da objave zbirne podatke o tome koliko moderatora sadržaja zapošljavaju po regionu, kao i o jeziku na kojem moderatori rade.** Trebalo bi da pruže konkretne informacije o tome kako se moderatori obučavaju da identifikuju rodne i druge potencijalno štetne sadržaje zasnovane na identitetu, kao i kako se moderatori obučavaju o međunarodnim standardima ljudskih prava.

- **Države treba da obezbede smisleno transparentno izveštavanje o radnjama koje su preduzele kao odgovor na širenje potencijalno zakonitog, ali štetnog sadržaja.** Javni organi treba da redovno stavljaju na raspolaganje javnosti sledeće sveobuhvatne informacije: broj, prirodu i pravnu osnovu svih zahteva za ograničenje sadržaja koji se šalju internet posrednicima; radnje preduzete kao rezultat tih zahteva; i ograničenja sadržaja na osnovu ugovora o uzajamnoj pravnoj pomoći.

Preporuke o okvirima za pristup podacima za nezavisne relevantno stručne zainteresovane strane

- **Države treba da uspostave obavezno eksterno izveštavanje za internet posrednike** koje treba da bude dostupno svim relevantnim nezavisnim zainteresovanim stranama i javnim organima, uključujući istraživače i organizacije civilnog društva. Internet posrednici bi trebalo da budu u obavezi da omoguće nezavisnu

eksternu reviziju bilo kog automatizovanog modela, štiteći pri tom poslovnu tajnu i privatnost/bezbednost podataka.

- **Države treba da uspostave obavezne modalitete pristupa podacima i eksterno izveštavanje za onlajn platforme.** Takvo izveštavanje treba da bude dostupno svim relevantnim nezavisnim zainteresovanim stranama i javnim organima, uključujući istraživače, organizacije civilnog društva i pogođene korisnike. Platforme bi trebalo da budu u obavezi da omoguće nezavisnu eksternu reviziju bilo kog njihovog modela koji pokreće algoritam, štiteći pri tom poslovnu tajnu i privatnost/bezbednost podataka. Države treba da utvrde kriterijume za obezbeđivanje nezavisnosti i kompetentnosti revizora.

- **Svako zakonodavstvo ili politika upravljanja sadržajem koju su pokrenule države mora da bude zasnovana na dokazima i istraživanju.** Javnim organima se mora odobriti pristup podacima koje čuvaju internet posrednici, u skladu sa adekvatnim okvirima zaštite podataka, tako da javni organi mogu da razviju politike zasnovane na dokazima i obezbede adekvatan nezavisni javni nadzor. Države bi stoga trebalo da uspostave uslove za pristup podacima za treća lica, sa jasnim odrednicama ko može da pristupi podacima, kojim podacima se može pristupiti i kako će se ti podaci prikupljati i proveravati pre obelodanjivanja, kao i ko će to raditi.

- **Države treba da utvrde kriterijume za obezbeđivanje nezavisnosti i kompetentnosti revizora.** Internet posrednici treba da se podvrgavaju redovnim nezavisnim, sveobuhvatnim i delotvornim revizijama. Opis potencijalnih pravnih ili drugih efekata sistema treba da bude dostupan za reviziju od strane nezavisnih tela sa neophodnim kompetencijama. Bez obzira na to, takve procene rizika uvek treba da budu sekundarna mera. Ex ante procene uticaja na ljudska prava koje se sprovode pod javnim nadzorom treba da budu primarni korak.

- **Organizacije civilnog društva, akademski istraživači koji sprovode istraživanja u javnom interesu i novinari treba da budu u mogućnosti da sprovode smisleno praćenje i revizije**

automatizovanih sistema za donošenje odluka. Nezavisne zainteresovane strane koje vrše reviziju trećeg lica trebalo bi da budu u mogućnosti da pristupe svim informacijama koje su im potrebne, kao što su izvorni kod, kriterijumi podataka i metrika učinka, kako bi izvršili suštinski nadzor nad samoregulacijom internet posrednika. Obezbeđene informacije treba da omoguće trećim licima da vrše reviziju i izveštavaju o funkcionisanju, delotvornosti i greškama automatizovanih odluka iza određenih uklanjanja sadržaja – kao i sadržaja koji je ostavljen na internet posredniku.

3.2 Preporuke za poštovanje ljudskih prava u upravljanju sadržajem

- **Države treba da razviju politiku ljudskih prava sa naglaskom na vidna pitanja ljudskih prava, uključujući slobodu izražavanja, slobodu medija, privatnost, nediskriminaciju i pravo na život, slobodu i bezbednost.** Države su nosioci dužnosti prema međunarodnom pravu o ljudskim pravima i imaju pozitivnu obavezu da štite ljudska prava od uplitanja drugih, uključujući privatne aktere ili pojedince. Stoga bi trebalo da se obavežu da će poštovati međunarodni zakon o ljudskim pravima i da obezbede da nacionalni zakoni koji regulišu platforme i upravljanje sadržajem budu u potpunosti u skladu sa međunarodnim okvirom ljudskih prava koji predstavlja pravnu obavezu za države.

- **Države treba da se uzdrže od zakonskog zahtevanja od onlajn platformi da primenjuju automatizovane alate za otkrivanje i identifikaciju potencijalno nelegalnog ili štetnog sadržaja**, što se u nekim jurisdikcijama naziva „proaktivna mera".

- **Države treba da obezbede jasne smernice o tome šta se smatra nezakonitim sadržajem prema važećem zakonodavnom okviru.** Nezavisni pravosudni organi treba da pruže detaljnu procenu šta je nezakonit sadržaj i da razlikuju različite vrste/kategorije nelegalnog sadržaja. Države bi trebalo da zahtevaju da platforme otkriju kako se različiti automatizovani alati koriste za određene kategorije nelegalnog sadržaja i koji su nameravani ciljevi (otkrivanje, identifikacija, uklanjanje, upravljanje pristupom saobraćaju, pojačavanje/ smanjenje, „skrivena zabrana" (shadow banning), itd.).

- **Države treba da zaštite i zakonski nalože poštovanje ljudskih prava za algoritamsku moderaciju sadržaja** uspostavljanjem mehanizama za ublažavanje štetnih uticaja korišćenja VI sistema od strane kompanija za moderaciju i prireĎivanje i deljenje sadržaja koji generišu korisnici, uključujući govor mržnje i nezakonit govor. Ovo se može postići obavezom internet posrednika da sprovode dužnu pažnju sistema veštačke inteligencije u pogledu ljudskih prava za otkrivanje, identifikaciju i adresiranje potencijalno štetnog sadržaja. Od posrednika bi trebalo da se zahteva da procene tačnost sistema i stope grešaka i potencijalnu štetu od takozvanih lažnih negativnih i lažnih pozitivnih rezultata, dok sveukupno rade na sprečavanju i ublažavanju diskriminatornih ishoda VI sistema, sa naglaskom na slobodi izražavanja i slobodi medija. Različiti skupovi podataka, kao i poznavanje i razumevanje lokalnog konteksta, lingvističkih nijansi i kodiranog jezika su od suštinskog značaja.

- **Države treba da zakonski nalože poštovanje ljudskih prava poslovnih modela prikupljanja podataka.** Poslovni modeli prikupljanja podataka mogu da povećaju negativne uticaje na ljudska prava podsticanjem potencijalno legalnog, ali štetnog sadržaja onlajn. Internet posrednici čiji poslovni modeli zavise od targetiranog oglašavanja i masovnog prikupljanja i analize korisničkih podataka treba da rade po principu opt-in, gde korisnici proaktivno pristaju na prikupljanje podataka i personalizovanu moderaciju i prireĎivanje i deljenje sadržaja. U najmanju ruku, takvi internet posrednici treba da ponude mogućnost da se odustane od prikupljanja podataka i/ili algoritamske moderacije sadržaja, istovremeno obezbeĎujući alternativna sredstva za osiguranje bezbednosti korisnika onlajn.

    - Zaštita „podrazumevanog prihvatanja“ („opt-in by default“) za algoritamske sisteme za moderaciju sadržaja bila bi poželjan mehanizam jer nudi veću zaštitu za korisnike koji su možda manje svesni kako ovi sistemi funkcionišu. Internet posrednici treba da osmisle saglasnost i politiku privatnosti na način koji olakšava informisani izbor korisnika i u skladu je sa zakonima o zaštiti podataka. Mehanizam „opt-in“ treba da omogući korisnicima da ostvare bar neki definisani minimalni stepen kontrole nad sistemima preporuka.

Preporuke za obaveznu procenu uticaja na ljudska prava

- **Države treba da nalože transparentne, nezavisne i inkluzivne ex ante procene uticaja na ljudska prava (HRIA), u okviru jasnog regulatornog okvira i uz nadzor regulatorne agencije ili nezavisnih zainteresovanih strana sa relevantnom stručnošću.** Procene treba da obuhvate pregled posredničkih proizvoda, usluga i sistema i njihovog uticaja na ljudska prava, sa naglaskom na pravo korisnika na slobodno izražavanje i zabrinutosti u vezi sa pluralizmom medija. HRIA treba da se sprovodi što je moguće otvorenije i transparentnije, uz aktivno angažovanje pojedinaca i grupa pogođenih govorom mržnje i nelegalnim govorom. Ove ex ante HRIA procene treba da se zasnivaju na doprinosima pogođenih zajednica i grupa zainteresovanih strana, uključujući civilno društvo i marginalizovane grupe. Rezultati HRIA treba da budu dostupni javnosti i da budu pristupačni i lako razumljivi.

- **Države treba da nalože da kompanije sprovode stalne procene uticaja na ljudska prava** algoritamskih modela moderacije sadržaja svojih platformi, tokom životnog ciklusa svojih sistema veštačke inteligencije. Kompanije treba da angažuju spoljne zainteresovane strane sa relevantnom ekspertizom u oblasti ljudskih prava i dizajna, razvoja i primene sistema koji moderiraju nelegalni sadržaj onlajn. Ovo angažovanje treba da naglasi inkluzivne i participativne pristupe koji se daju marginalizovanim i ranjivim grupama. Postavka, metodologija i rezultati procene uticaja na ljudska prava treba da prate opšte priznate najbolje prakse i da budu javno dostupni. Procene uticaja takođe moraju da omoguće rešavanje pitanja proporcionalnosti. Ovo zauzvrat zahteva da se proceni i korisnost/neophodnost intervencije i rezultirajuća povreda ljudskih prava.

- **Države treba da obezbede da internet posrednici razviju interne procese koji im omogućavaju da otkriju i spreče rizike za ljudska prava.** Dok će struktura i obim ovih mehanizama zavisiti od veličine posrednika, svi posrednici, bez obzira na njihovu veličinu, treba da uspostave odgovarajuće interne mehanizme, uključujući interne revizije. Posebno u slučaju govora mržnje, kriterijumi za procenu rizika moraju da pomognu da se utvrdi da li su pojedinci ili grupe iz marginalizovanih zajednica neproporcionalno pogođeni, i ako jesu, kako. Posebnu pažnju

treba posvetiti intersekcionalnim razmatranjima o načinima na koje se rasa, klasa, pol i druge individualne karakteristike mogu kombinovati u različite načine diskriminatornog tretmana.

3.3 Preporuke o pristupu efikasnom pravnom leku i pravnoj zaštiti

- **Države treba da zahtevaju da internet posrednici uspostave operativne mehanizme za žalbe.** Prvo, pogođeni korisnik mora da ima mogućnost da zatraži dodatne informacije o ishodu alata za moderaciju sadržaja vođenog algoritmom, posebno ako rezultati dovode do uklanjanja sadržaja. Drugo, korisnik mora da ima mogućnost da zatraži ljudski pregled. Treće, korisnici moraju da imaju pristup svim neophodnim informacijama za žalbu na odluku, uključujući i sudske sporove. To obuhvata, između ostalog, informacije koje se odnose na svrhu alata za moderaciju sadržaja vođenog algoritmom, uslove primene, metriku evaluacije (lažno pozitivne/lažno negativne) itd.

- **Da bi se osiguralo da pojedinci imaju pristup delotvornom pravnom leku, države treba da zahtevaju da se navedu konkretni razlozi za odluke o upravljanju sadržajem**, bez obzira da li su donete putem ljudske ili automatizovane provere. Korisnici moraju da budu obavešteni o odlukama o moderaciji sadržaja koje ih se tiču, uključujući uklanjanje sadržaja, demonetizaciju i suspenzije i uklanjanja naloga.

- **Države moraju da osiguraju da internet posrednici pružaju smislenu priliku korisnicima da se žale na odluke.** Ovo je od posebnog značaja kada se sadržaj uklanja i nalozi suspenduju. Žalbeni procesi moraju da budu pristupačni i blagovremeni, i da obezbede efikasne pravne lekove, koji mogu uključivati vraćanje uklonjenog sadržaja ili poništavanje suspenzije naloga. Kada se početne odluke o sadržaju donose automatizovanim sredstvima, proces žalbe mora da uključi procenu čoveka. Trebalo bi navesti jasne razloge u slučajevima kada je žalba koju je pokrenuo korisnik neuspešna, kako bi korisnik mogao da razume tu odluku. Korisnici bi trebalo da budu u mogućnosti da pruže dodatne dokaze kada se žale na uklanjanje njihovog sadržaja ili suspenziju njihovog naloga. Žalbeni postupak na nivou internet posrednika može da obezbedi pravne lekove, kao što su ispravka,

izvinjenje, detaljan odgovor, objašnjenje, korekcije, vraćanje naloga ili kombinacije nekoliko oblika pravnog leka u jednom. Međutim, ovaj oblik pravnog leka ne bi trebalo da zameni delotvoran sudski lek i pravosudnu zaštitu. Sve u svemu, onlajn platforme bi trebalo da obezbede dodatnu procenu od strane ljudi – vodeći računa o tome da uključe ljudsku intervenciju („human-in-the-loop").

- **Države treba da podstiču politike i istraživačke inicijative koje se bave uticajem dizajna interfejsa na ponašanje korisnika, kao i rešavanjem problema kao što su obmanjujući interfejsi poznati kao „tamni obrasci".** Pored automatizovanih alata za otkrivanje, internet posrednici i dalje se oslanjaju na korisničke izveštaje o zloupotrebi i uznemiravanju na svojim sajtovima. Korisnici imaju pravo na odgovarajuću naknadu za govor mržnje koji ih targetira.

- **Internet posrednici treba da poboljšaju karakteristike dizajna interfejsa mehanizama za prijavu zloupotrebe, tako da budu pristupačni, efikasni, prilagođeni uzrastu i usmereni na korisnika.** Internet posrednici treba redovno da prikupljaju povratne informacije i traže doprinos od korisnika i civilnog društva, posebno onih koji predstavljaju istorijski marginalizovane i rizične grupe, kako bi poboljšali efikasnost i dostupnost mehanizama za izveštavanje. Pored toga, korisnici koji označe sadržaj treba da budu obavešteni o donetim odlukama i ishodima u vezi sa sadržajem koji su prijavili.

- **Države treba da obezbede da standardi zajednice i uslovi za korišćenje usluge koji čine osnovu odluka o moderaciji sadržaja budu jasno formulisani i dostupni.** Moraju da budu dostupna razumljiva pravila i smernice o dozvoljenom i nedozvoljenom korišćenju usluge internet posrednika, kao i posledicama kršenja uslova korišćenja usluge. Ova transparentnost je neophodna za pojedinačne korisnike, kao i za smisleni nadzor civilnog društva i vlade. Internet posrednici treba da redovno obaveštavaju sve korisnike o promenama u pogledu usluge na sveobuhvatan i jasan način.

- **Države treba da stvore povoljno okruženje za učešće u javnoj debati, uključujući slobodu medija.** Države treba da preduzmu preventivne i proaktivne napore kako bi se pozabavile strukturalnim i institucionalizovanim oblicima mržnje i širenjem govora mržnje na internetu. To uključuje pokretanje i podržavanje kampanja podizanja

svesti za edukaciju javnosti – posebno korisnika platformi društvenih mreža – o šteti nanesenoj onima koji su na meti uznemiravanja i zlostavljanja onlajn, uključujući društvene uticaje i efekte zastrašivanja na marginalizovane grupe. Takvi proaktivni napori takođe treba da obuhvataju: ulaganja u istraživanje potencijalnih pozitivnih upotreba veštačke inteligencije za stvaranje bezbednijih onlajn prostora koje vodi zajednica; inicijative za zaustavljanje plime govora mržnje koje prevazilaze uklanjanje ili suspenziju naloga; i stvaranje mogućnosti i foruma za dijalog između internet posrednika, civilnog društva i marginalizovanih grupa kako bi se poboljšalo otkrivanje i moderacija govora onlajn.

3.4 Preporuke o pozitivnoj upotrebi veštačke inteligencije za stvaranje bezbednih prostora za marginalizovane grupe koje vodi zajednica

- **Države treba da podstaknu internet posrednike da marginalizovanim zajednicama daju važnu moć donošenja odluka u procesu dizajniranja i implementacije novih proizvoda veštačke inteligencije.** Od trening podataka do primene sistema veštačke inteligencije, mnogi ljudi imaju stručnost i proživljena iskustva da vode ovaj proces na način koji maksimizira pozitivan uticaj i minimizira eksterne efekte na istorijski marginalizovane i rizične zajednice.

- **Države treba da podrže postojeće VI inicijative od strane i za marginalizovane zajednice, i napore da se osnaže i obrazuju zajednice da razumeju i koriste potencijalno korisnu veštačku inteligenciju.** Neke grupe već kreiraju vodiče za zajednicu i radionice o tome kako da koriste veštačku inteligenciju za osnaživanje, umesto za nadzor i dodatno marginalizovanje zajednice.[23]

- **Države treba da podrže široku raznovrsnost pristupa, a ne rešenje „jedna veličina za sve".** Primeri uključuju dodatak za pretraživač „Odustanak" („Opt-out" browser add-on), koji nastoji da otkrije mizoginiju i ukloni sadržaj, kao što je blokator oglasa. U prostoru

23 Na primer, videti https://alliedmedia.org/wp-content/uploads/2020/09/peoples-guide-ai.pdf.

P2P tehnologije, neki zagovaraju upotrebu „naočara“ ili subjektivnu moderaciju.[24] U ovom slučaju, može da postoji niz algoritama za moderaciju VI, a pojedinačni korisnik može da izabere da aktivira jedan ili više sistema u isto vreme. Takvi algoritamski vođeni alati ne bi cenzurisali ceo razgovor, već bi jednostavno promenili sadržaj kojem su korisnici pojedinačno izloženi. Raznolikost strategija stvara jasna rešenja kroz primenjeno eksperimentisanje.

- **Države treba da promovišu otvoreni kod postojećih vlasničkih modela gde god je to moguće i da omoguće povratne informacije zajednice u njihovoj primeni.** Dok modeli mogu da budi veoma profitabilni kada su zatvorenog koda, projekti kao što je Hugging Face pokazuju da svet veštačke inteligencije može biti jasan, profitabilan i otvoren.[25]

3.5 Preporuke za formalizovanje saradnje sa organima za sprovođenje zakona

- **Države treba da adekvatno sprovode zaštitne mere za zabranu obaveznog prenosa podataka, posebno organima za sprovođenje zakona**, i u tom smislu, preduzmu posebne mere za zaštitu marginalizovanih i ranjivih grupa.

- **Kada nadgledaju i/ili prate sadržaj onlajn, države treba da se pridržavaju međunarodnog zakona o ljudskim pravima, uključujući tripartitni test za bilo kakva ograničenja slobode izražavanja.** Kada naređuju nadgledanje i praćenje sadržaja – ili nalažu platformama da uklone sadržaj identičan ili sličan sadržaju koji je ranije ocenjen kao nezakonit – države će obezbediti da su sve mere propisane zakonom, da imaju legitiman cilj, da su neophodne i da koriste najmanje nametljiva sredstva da efikasno postignu svoj cilj. Konkretno, treba jasno identifikovati legitimni cilj bilo koje državne mere koja rezultira upotrebom alata veštačke inteligencije za upravljanje sadržajem, a koristi moraju da budu eksplicitno prikazane, tako da proporcionalnost između tih dokazanih koristi i rezultirajućih povreda ljudskih prava može da se demonstrira.

24 Emmi Bevensee, The Decentralised Web of Hate: White supremacist are starting to use Peer-to-Peer technologies. Are we prepared? na <https://rebelliousdata.com/p2p/>.
25 Za dodatne detalje konsultovati https://huggingface.co/.

4. Zaključak

Proliferacija potencijalno nelegalnog i štetnog sadržaja na mreži i uticaji algoritamskog donošenja odluka ostaju složena i nijansirana pitanja. Skoro pet decenija nakon potpisivanja Završnog akta iz Helsinkija, saradnja među državama članicama OEBS-a je i dalje neophodna, kako bi se uhvatili u koštac sa novim izazovima moderacije onlajn sadržaja i širenja nelegalnog sadržaja, kao i zakonitog, ali štetnog govora mržnje, onlajn. Ovo je naročito vidno sa porastom moćnih internet posrednika, koji deluju kao čuvari i moderatori izražavanja na ovom novom i sve važnijem (digitalnom) tržištu ideja.

Ovaj deo izveštaja ima za cilj da predstavi principijelan pristup regulisanju nelegalnog sadržaja i zakonitog, ali štetnog govora mržnje onlajn, sa posebnim fokusom na uticaje govora mržnje i algoritamskog donošenja odluka na marginalizovane grupe. Države su prvenstveno odgovorne za poštovanje, promovisanje i sprovođenje ljudskih prava, uključujući slobodu izražavanja, slobodu medija i zaštitu od diskriminacije. Ova odgovornost uključuje efikasnu regulaciju internet posrednika u svim fazama procesa – od dizajna i razvoja algoritamskih modela do pravnih lekova koji moraju biti dostupni pogođenim pojedincima i grupama.

Ovaj deo izveštaja navodi niz proaktivnih, preventivnih i odgovarajućih preporuka, koje imaju za cilj da usmere države članice OEBS-a u ovom zadatku. Preporuke se odnose na različite relevantne aspekte, kao što je osiguranje algoritamske transparentnosti; preduzimanje odgovarajuće analize ljudskih prava; obezbeđivanje pristupa delotvornom pravnom leku i pravnoj zaštiti; smisleno uključivanje civilnog društva i pogođenih zajednica u svim fazama životnog ciklusa alata vođenog algoritmom; i promovisanje pozitivne upotrebe veštačke inteligencije za stvaranje bezbednih prostora za marginalizovane grupe, koje vodi zajednica.

VI u priređivanju i deljenju sadržaja

Priređivanje i deljenje sadržaja i medijski pluralizam

Ovaj deo se fokusira na korišćenje **VI u priređivanju i deljenju sadržaja**, baveći se pitanjima uticaja koje sistemi za preporuku sadržaja zasnovanog na podacima imaju na raznovrsnost i medijski pluralizam. Ovaj i naredni deo u kome se ističu nedostaci priređivanja i deljenja sadržaja zasnovanog na VI i targetiranog oglašavanja predočavaju preporuke koje su usredsređene na ljudska prava kako bi se sprečili negativni uticaji VI alata u priređivanju i deljenju sadržaja na pravo na slobodu mišljenja i izražavanja.

1. Definisanje obima uticaja priređivanja i deljenja sadržaja na medijski pluralizam

1.1 Relevantnost algoritma za priređivanje i deljenje sadržaja i sistema za preporuke baziranih na podacima za medijski pluralizam i raznovrsnost

Raznovrsnost i medijski pluralizam predstavljaju okosnicu demokratskih principa na čiji kvalitet utiče porast dominantnih internet posrednika i njihov uticaj na javni diskurs. Internet posrednici, pre svega društvene mreže, postali su značajan izvor, tačka pristupa i ključni distributer informacija, uključujući i sadržaja vesti. Širenje informacija, i, sve veća, agregacija se prevashodno dešava preko algoritma za priređivanje i deljenje sadržaja[26] i sistema za preporuke. Korišćenjem optimizacije i analize ljudskih i neljudskih agenata, ovi sistemi „dostavljaju" personalizovan sadržaj upodobljen pojedinačnim profilima, što dovodi do vrste i količine sadržaja kome je svaki pojedinac izložen. Sistemi za preporuku sadržaja, koji rangiraju sadržaj da utvrde šta se predstavlja pojedinačnim korisnicima, utiče na slobodu pojedinaca da traže i šire informacije, kao i na opšti informativni pejzaž i medijske slobode. Dizajn sistema za preporuke

26 Priređivanje i deljenje postojećeg sadržaja treba da se shvati kao set algoritamskih procesa i procesa kojima upravlja čovek, a koji podržavaju distribuiranje sadržaja publici, poput rangiranja sadržaja ili uređivačke analize podataka. Videti: B. Bukovska et al, Spotlight on Artificial Intelligence and Freedom of Expression #SAIFE (2020), strana 19.

značajno utiče na ono što se vidi onlajn, i šta ostaje skriveno - i za koga. Proces algoritamskog priređivanja i deljenja temelji se[27] na vrednostima i ciljevima kreatora algoritma,[28] sociološko-tehničkim faktorima, samo-regulaciji (uslovi za pružanje usluga, na primer) i regulaciji na nivou države. S obzirom na sveprisutnost onlajn sadržaja, i njegov značaj za kreiranje mišljenja i donošenja odluka, postavlja se ključno pitanje: Gde leži odgovornost u definisanju i primeni politika za uspostavljanje prioriteta i kodifikaciju medijskog pluralizma i raznovrsnosti[29] u eri digitalnog informisanja?

Ovaj deo izveštaja predočava konceptualni rezime ključnih algoritamskih procesa za priređivanje i deljenje, i njihov transformativan uticaj na medijski pluralizam. On dalje predočava niz preporuka za zemlje članice OEBS-a u vezi sa pristupom algoritamskom priređivanju i deljenju sadržaja zasnovanom na ljudskim pravima. Kao takav, ovaj deo se fokusira na uticaj algoritamskog priređivanja i deljenja sadržaja i sistema za preporuke vođenih podacima na medijski pluralizam i raznovrsnost u demokratskim društvima - i ulogu države da deluje kao najviši garant ljudskog prava slobode izražavanja i da osigura okruženje za njegovo izražavanje.

1.2 Nepodudarnost algoritamskog priređivanja i deljenja sadržaja i slobode izražavanja

Mogućnost da se filtrira, uspostavi prioritet i angažuje sa onlajn sadržajem na osnovu ličnih afiniteta i interesa često je u sukobu sa individualnim posredovanjem u traženju, primanju i deljenju raznovrsnih informacija.[30]

27 K. Klonick, The New Governors: The People, Rules, and Processes Governing Online Speech, The Harvard Law Review, strana 1664.
28 Radsch, Courtney. "Digital Information Access." In A New Global Agenda: Priorities, Practices, and Pathways of the International Community, uredio D. Ayton-Shenker, 72–83. Rowman & Littlefield Publishers, 2018. https://books.google.com/books?id=tyJJ-DwAAQBAJ.
29 Za temu koja se odnosi na izloženost, videti: Philip M Napoli, "Rethinking Program Diversity Assessment: An Audience-Centered Approach" (1997) 10 Journal of Media Economics 59-74.; N. Helberger & M. Wojcieszak (2018). Exposure Diversity. In STR. M. Napoli (Ed.), Mediated Communication (p. 535-560). (Handbooks of Communication Science; Vol. 7). De Gruyter Mouton. https://doi.org/10.1515/9783110481129-029. https://doi.org/10.1515/9783110481129-029.
30 P. Leersen, The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems, European Journal of Law and Technology (2020), strana 12.

Internet posrednici, kao osnovni princip, uglavnom uspostavljaju prioritete i prikazuju sadržaje pojedincu na osnovu sistema predikcije da će se pojedinac verovatno angažovati sa sadržajem. Slično sistemima za personalizovano oglašavanje zasnovano na ponašanju, sistemi za preporuku sadržaja opsežno prikupljaju podatke korisnika (i onih koji to nisu) da bi kreirali digitalne profile, procenili sličnosti i doneli zaključke na osnovu tih podataka.

Mnogi poslovni modeli onlajn platformi,[31] koji prioritizuju angažovanje i profit nad pristupom baziranim na ljudskim pravima, mogu i dovode do praksi eksploatišućih i intruzivnih i zadiranja u podatke, širenja pogrešnih informacija/dezinformacija i algoritamskih povratnih sprega.[32] Utvrđeno je da imaju negativan uticaj na pluralizam sadržaja, naročito kada je reč o sadržajima koje kreiraju marginalizovane zajednice ili koji je kreiran za njih. Model ovekovečava praznine u informisanju[33] i predstavlja prepreku za zagovaranje, čime se ponovo stvara i jača strukturna društvena nejednakost. Postoje i dokazi koju ukazuju da proces moderacije sadržaja koristi onim grupama koje već dominiraju onlajn prostorima i narativima nad marginalizovanim grupama, informacijama i narativima.[34] Osim toga, utvrđeno je da otkrivanje sadržaja vođeno algoritmom (npr. pretraživači) pojačava rasizam tako što sugeriše diskriminatorne fraze za pretraživanje i nepodudarnosti, naročito na rasnoj, jezičkoj i rodnoj osnovi, u prikazima pripadnika marginalizovanih zajednica.[35]

Uglavnom, algoritamski sistemi za prireďivanje i deljenje postojećeg sadržaja i sistemi za preporuke zasnovani su na ličnim (internim) pravilima, interesima i pretpostavkama posrednika, pre nego na

31 Videti: Ranking Digital Rights, It's the Business Model: How Big Tech's Profit Machine is Distorting the Public Sphere and Threatening Democracy (2020).
32 Bodó, B., Helberger, N., Eskens, S., & Möller, J, Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. Digital Journalism (2019), strana 219.
33 A. Causevic and A. Sengupta, Whose Knowledge Is Online? Practices of Epistemic Justice for a Digital New Deal, IT for Change (2020), preuzeto sa: https://itforchange.net/digital-new-deal/2020/10/30/whose-knowledge-is-online-practices-of-epistemic-justice-for-a-digital-new-deal/.
34 B. Marshall, Algorithmic misogynoir in content moderation practice, Heinrich-Böll-Stiftung (2021), str. 7,11. Vidi takođe: M. E, Mazzoli and D. Tambini, Prioritisation uncovered: The Discoverability of Public Interest Content Online. Council of Europe (2020), str. 44.
35 Safiya Umoja Noble, Algorithms of Oppression How Search Engines Reinforce Racism, NYU Press (2018).

demokratskim vrednostima ili javnom interesu.[36] Preporuka sadržaja od ključnog je značaja za rast i dominaciju velikih internet posrednika, i u središtu je njihovih poslovnih modela. Kako su sistemi za preporuku „ključna logika koja upravlja tokovima informacija od kojih zavisimo“,[37] internet posrednicima je omogućeno da deluju kao čuvari informacija i znanja. Ovo ima šire implikacije po javni interes, predstavljanje i (ne) jednakost moći, kako onlajn tako i oflajn.[38] Posrednički sistemi za preporuku značajno su rekonfigurisali logiku javnih komunikacija, uključujući pristup vestima, ključnim informacijama i opštem sadržaju od javnog interesa. Tako njihovi sistemi za preporuku značajno ograničavaju jednak pristup novinarima i medijima, dok vrše pritisak na profesionalno novinarstvo zbog odliva novca od reklama posrednicima. Nalazi skorašnjih istraživanja o algoritamskoj prioritizaciji, koji su definisani kao „dijapazon dizajniranih i algoritamskih odluka koje dovode do isticanja i mogućnosti otkrivanja sadržaja“[39] otkrivaju potencijal za polarizaciju onlajn mišljenja i stavova. Na primer, važan faktor u procesu uspostavljanja prioriteta kod sadržaja jeste individualna politička predispozicija i/ili pripadnost. Uspostavljanje prioriteta može, stoga, da učvrsti i produbi polarizaciju onlajn mišljenja i stavova, naročito među onim korisnicima koji su na krajnjim pozicijama političkog spektra koji verovatno već predominantno konzumiraju pripadajuće sadržaje.[40] Pokazalo se takođe da su „određene grupe u društvu sklonije selektivnoj izloženosti u odnosu na druge“.[41]

36 C. Radsch. “Digital Information Access.” In A New Global Agenda: Priorities, Practices, and Pathways of the International Community, uredio D. Ayton-Shenker, 72–83. Rowman & Littlefield Publishers, 2018. https://books.google.com/books?id=tyJJDwAAQBAJ. J. Haas, Freedom of the media and artificial intelligence, https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/human_rights-droits_homme/policy-orientation-ai-ia.aspx?lang=eng.

37 T. Gillespie (2018). Custodians of the internet. https://www.researchgate.net/publication/327186182_Custodians_of_the_internet_Platforms_content_moderation_and_the_hidden_decisions_that_shape_social_media.

38 P. Leerssen, The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems. preuzeto sa file:///Users/eliskapirkova/Downloads/Leerssen%20EJLT_corr.pdf.

39 M.E. Mazzoli and D. Tambini. Prioritisation uncovered: The Discoverability of Public Interest Content Online. Council of Europe (2020), strana 12.

40 B. Stark, D. Stegmann, Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse. preuzeto sa https://algorithmwatch.org/wp-content/uploads/2020/05/Governing-Platforms-communications-study-Stark-May-2020-AlgorithmWatch.pdf.

41 B. Bodó, N. Helberger, S. Eskens & J. Möller, Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. Digital Journalism (2019), strana 15.

Iako aktivna personalizacija zasnovana na inputima korisnika ima tendenciju da proizvede veću raznovrsnost informacija, pasivna personalizacija zasnovana na algoritamskom izboru sadržaja ima tendenciju da pogorša takozvani efekat filter balona.[42]

Pristrasnost i diskriminacija, uključujući i rodno zasnovanu diskriminaciju, u algoritamskom donošenju odluka zasnovanom na podacima, mogu da se jave iz nekoliko razloga i na mnogo nivoa u sistemima za priređivanje i deljenje sadržaja, i može biti teško da se one detektuju i ublaže. Predočeno je da isključivanje senzitivnih informacija/informacija zasnovanih na identitetu dovoljno štiti protiv diskriminacije. Ipak, diskriminacija može i stvarno se dešava, uprkos tim „zaštitama", s obzirom na sve veći broj i raznovrsnost informacija koje su sadržane u algoritamski-utemeljenim skupovima podataka. Pristrasnost u algoritmima može da proistekne iz dizajna i implementacije, uključujući i nereprezentativnim ili nepotpunim trening podacima, ili oslanjanje na individualne, eksperimentalne ili vrednosno-utemeljene podatke koji reflektuju istorijske/strukturne nejednakosti. Algoritamska pristrasnost može da ima kolektivan, različit uticaj na zajednice, posebno na marginalizovane grupe, čak i kada ne postoji namera da se one diskriminišu. Stoga je neophodno ispitivanje kako željenih, tako i neželjenih posledica algoritama. Trenutne javne politike mogu da se pokažu kao nedovoljne za utvrđivanje, ublažavanje i rešavanje uticaja na pojedince ili društvo u celini. Pored namernih napora da se uobliči pažnja pojedinca (direktna manipulacija), postoji opasnost od uvođenja neželjene i indirektne pristrasnosti u algoritam kroz uključivanje velikih podataka (big data) na različitim nivoima sistema za priređivanje i deljenje sadržaja. Kako direktna, tako i indirektna diskriminacija koja nastaje usled algoritama koji koristi velike podatke (big data) predstavljaju jednu od najurgentnijih opasnosti algoritamski vođenih procesa priređivanja i deljenja sadržaja.

Kombinacija posledica filtriranja i personalizacije sadržaja u suštini kreira slojeve restrikcija u smislu mogućnosti za dalje otkrivanje, i samim tim, pristupačnost različitih medijskih sadržaja. Sva gore navedena pitanja imaju ozbiljne implikacije po medijski pluralizam, shvaćen kao pluralizam izvora informacija (eksterni pluralizam) i sadržaja (interni pluralizam)[43]

42 D. Wagner, Artificial Intelligence and Disinformation as a Multilateral Policy Challenge https://www.osce.org/files/f/documents/d/0/506702.pdf.

43 Za temu koja se odnosi na izloženost, videti: P. M. Napoli, "Rethinking Program Diversity Assessment: An Audience-Centered Approach" (1997) 10 Journal of Media Economics 59-74.; Helberger,

Konkretnije, i u kontekstu ovog izveštaja, medijski pluralizam se takođe odnosi na raspodelu komunikativne moći (ili „glasa“) u društvu. Pravedna raspodela „glasa“, kao preduslov, zahteva dekoncentraciju moći i decentralizaciju resursa unutar informacionog ekosistema,[44] kao i podršku alternativnim modelima koji nude raznovrsnost narativa i sadržaja. Jasno je da algoritamski vođeni procesi priređivanja i deljenja sadržaja transformišu pojmove medijskog pluralizma i raznovrsnosti, koji su neophodni za demokratsku, javnu debatu i inkluzivna društva.

U tom kontekstu, predlaže se državnim i nedržavnim akterima, prvenstveno internet posrednicima i medijskim organizacijama, ali i međunarodnim i regionalnim organizacijama, predstavnicima civilnog društva i akademske zajednice, da usvoje politike koje doprinose stvaranju okruženja za medijski pluralizam. To znači omogućavanje pristupa, dostupnosti, mogućnosti otkrivanja i potrošnje različitih vrsta (medijskih) sadržaja kroz različite medije i preko više kanala.

2. Algoritamsko priređivanje i deljenje sadržaja i sistemi za preporuku vođeni podacima: uticaj na medijski pluralizam

2.1 Tipologija

Izrazit izvor uticaja, a time i komunikativne moći, internet posrednika i kompanija društvenih medija leži u njihovim sistemima za preporuku sadržaja, koji takođe „daju značaj njihovoj ulozi u demokratskoj kulturi“.[45] U suštini, „sistemi za preporuku“ uključuju različite tehnologije koje filtriraju, preuzimaju i organizuju informacije za pojedince. Faktori za rangiranje mogu da uključe nivo angažovanja sa određenim sadržajem, tip sadržaja, kada je prvi put podeljen ili način na koji su korisnici ranije imali interakciju sa sličnim sadržajem. Rangiranjem sadržaja, ovi sistemi imaju mogućnost da oblikuju i utiču na sposobnost pojedinaca da formiraju mišljenja.

N., & Wojcieszak, M. (2018). Exposure Diversity. In STR. M. Napoli (Ed.), Mediated Communication (p. 535-560). (Handbooks of Communication Science; Vol. 7). De Gruyter Mouton. https://doi.org/10.1515/9783110481129-029.

44 M. Moore i D. Tambini (eds) (2018) Digital Dominance: The Power of Google, Amazon, Facebook, and Apple. New York: Oxford University Press.

45 K. Klonick, The New Governors: The People, Rules, and Processes Governing Online Speech, The Harvard Law Review, strana 1663.

Osnovna svrha sistema za preporuku sadržaja jeste da filtriraju velike količine informacija onlajn. Ovaj algoritamski vođen proces funkcioniše na različite načine:

- **Filtriranje zasnovano na sadržaju:** pojedinci dobijaju preporuke sadržaja na osnovu njihovih navedenih ili podrazumevanih preferencija. Na primer, ako neko voli klasičnu muziku ili vesti o omiljenom sportskom timu, onda će sistem preporuka dati prioritet tim stavkama koje su u skladu sa njihovim interesovanjima i verovatno će podstaći angažovanje.

- **Kolaborativno filtriranje:** pojedinci dobijaju preporuke sadržaja na osnovu ljudi sa kojima su blisko povezani ili sa kojima dele sličnosti (u demografskoj kategoriji, preferencijama sadržaja itd.). Na primer, kada čitate vesti, sistem preporučuje članke koje je prijatelj podelio/pročitao, ili, kada obavljate onlajn kupovinu, sistem preporučuje artikle koje su kupili ljudi sa sličnom istorijom kupovine.

- **Hibridno filtriranje:** kombinacija gore navedenih metoda filtriranja i prireďivanja i deljenja. Na primer, preporučivanje informativnog članka koji se dopao prijatelju, ali samo ako pokriva određenu temu od interesa za korisnika, i kombinovanje sa širokim spektrom različitih metapodataka, kao što su lokacija pojedinca, istorija korišćenja itd.

Svi ovi procesi zasnovani su na korisničkim podacima, profilima i interakcijama sa datom platformom, kao i informacijama prikupljenim iz osnovne arhitekture reklamnih tehnologija. Algoritam jasno utvrđuje precizan način na koji se prikupljaju preporuke sadržaja, koristeći filtriranje zasnovano na sadržaju i kolaborativno filtriranje. Sistem kreira strategiju preporuke o tome kako se podaci kombinuju da bi se izračunao potencijalni angažman, na osnovu preporuka korisnika, u cilju zadovoljavanja kriterijuma optimizacije. Jednostavnije rečeno, algoritamsko prireďivanje i deljenje sadržaja predstavlja strategiju koju koriste sistemi za preporuku da bi se utvrdilo kako se prikupljeni podaci mogu najbolje iskoristiti za postizanje unapred definisanih ciljeva optimizacije.[46]

46 Da bi se postigao cilj optimizacije, algoritam može da istakne različite načine kako da odredi prioritet kod prikupljenih podataka. Na primer, algoritam bi mogao da favorizuje novinske članke pregledane u skorašnje vreme. Druga strategija bi bila da se posmatra popularnost članaka kao kriterijum rangiranja - kako sortirati konačne preporuke izračunate za svakog korisnika. Tačnost je još jedan često korišćen način za prireďivanje i deljenje sadržaja. Pristup optimizovan za tačnost pokušava da reprodukuje preferencije

Svi veliki internet posrednici, a posebno društvene mreže, koriste takozvane „otvorene sisteme za preporuku sadržaja".[47] Ovi sistemi podrazumevano koriste korisnički generisani sadržaj u sistemima za preporuku, ali određene stavke sadržaja mogu biti isključene na osnovu, na primer, kršenja Uslova korišćenja usluge. Da bi optimizovali angažovanje, ovi sistemi „personalizuju" onlajn iskustvo tako što daju prioritet sadržaju za koji se pretpostavlja da je privlačan i koji se prikuplja prethodnim angažovanjem i ponašanjem svakog pojedinca. Shodno tome, video snimci, rezultati pretrage, novinski članci ili bilo koji drugi tip sadržaja koji se prikazuje korisniku je jedinstven za njihovo iskustvo i razlikuje se od onoga što drugi korisnici vide. Iz tog razloga, između ostalog, algoritamski sistemi za preporuku sadržaja imaju potencijal da potkopaju i poremete demokratske procese.[48] Oni rizikuju da sužavaju izloženost pojedinaca i pristup različitim gledištima, vrednostima i narativima, ugrožavajući time pluralizam i raznovrsnost. Ovo može da zahteva državnu intervenciju i pažnju.

Algoritamsko filtriranje i prilagođavanje onlajn sadržaja na osnovu spekulisanih ličnih preferencija i interesovanja smanjuje izloženost raznovrsnosti informacija, sa potencijalnim negativnim efektima na raznovrsnost i javni diskurs, kao i privatnost. Sistemi priređivanja i deljenja sadržaja stoga mogu duboko da utiču na izvore informacija koji čine osnovu za donošenje dobro utemeljenih mišljenja, a time i na proces mišljenja pojedinaca. Iako istraživanje još nije konačno, ovo bi moglo da potkopa sposobnost pojedinaca da formiraju svoje mišljenje i da ih učini ranjivim na manipulativno uplitanje. Pošto su sistemi izgrađeni na intruzivnim praksama zadiranja u podatake i arhitekturi ubeđivanja (velikog obima), neizbežna personalizacija sadržaja može da ima značajan uticaj na kognitivnu autonomiju pojedinaca i da omete njihovo pravo na formiranje mišljenja.

korisnika što je bliže moguće. Oni izračunavaju preporuke koje su u skladu sa postojećim korisničkim preferencijama. U zavisnosti od podataka koji su prikupljeni i prirode predmetne stavke, postoji više načina kako da se prečisti svaka strategija.

47 Kao suprotnost zatvorenog sistema za preporuku sadržaja, koji korisnicima pruža sadržaj sa ograničene liste opcija. Te liste sastavlja vlasnik platforme.

48 N. Helberger (2019) On the Democratic Role of News Recommenders. Digital Journalism 7(8). Routledge: 993–1012. DOI: 10.1080/21670811.2019.1623700.

2.2 Prireďivanje i deljenje sadržaja i uspostavljanje prioriteta među sadržajem od javnog interesa

Metode kojima onlajn platforme prireďuju i dele sadržaj putem sistema za preporuku nisu transparentne i vrlo retko su predmet javne i/ili državne detaljne provere. Kada internet posrednici uključe raznovrsnost u sisteme za preporuku, to se odnosi na dizajnerski izbor da se aktivno uključuju korisnici i povećava profit. Raznovrsnost koji je priredila i podelila platforma prvenstveno je korišćena za optimizaciju finansijske dobiti, a ne za promovisanje demokratske debate, kroz praksu produženog angažovanja kako bi se postigao ono što se naziva cilj optimizacije – povećanje prihoda od oglasa ili veća vrednost platforme/usluga kroz povećan promet.[49] Pojednostavljeno rečeno, poslovno vođena merila za prireďivanje i deljenje sadržaja većinom su uspostavljena da bi se optimizovala ekonomska dobit i povećalo angažovanje korisnika za korporativni interes, a ne za traženje i osiguranje istinski raznovrsnih sadržaja.[50] Sistemi za preporuke sadržaja takođe mogu da imaju neželjene posledice iz perspektive širih društvenih ciljeva i mogu negativno da oblikuju i ometaju apsolutno pravo na slobodu misli i mišljenja.

Pored toga, procesi sistema za preporuku internet posrednika obično isključuju izbor, kontrolu i posredovanje pojedinačnih korisnika – preduslove za obezbeďivanje individualne autonomije u traženju i prenošenju različitih informacija i ideja. Nakon javnog obelodanjivanja brojnih slabosti sistema za preporuku,[51] došlo je do sve većeg pritiska javnosti i države da bi se osiguralo da njihovi procesi bolje i smislenije daju prioritet „raznovrsnoj" medijskoj izloženosti. Konkretno, zabrinutost je izražena u kontekstu zakonitosti i dometa političkog govora, kao i širenja i normalizacije specifičnih sistema vrednosti kao zaštićenog govora, čak i kada sadržaj krši Uslove korišćenja usluge ili međunarodne standarde ljudskih prava. S obzirom na potpuno nepostojanje informacija o načinu na koji platforme upravljaju i daju prioritet govoru, jasno je da bi sistemi

49 Prema rečima zvaničnika Facebook-a: "Facebook je profitabilan samo onda kada dodate mnoštvo malih interakcija koje ništa ne vrede, i to odjednom vredi milijarde dolara," K. Klonick, New Governor, strana 1627.
50 K. Klonick, The New Governors, strana 1664.
51 Poznat i najviše citiran slučaj je bio "Napalm Girl", novinarske fotografije koju je Facebook uklonio u skladu sa svojom politikom o nagosti.

za preporuku i povezane logike optimizacije mogli da naruše mogućnost da svi imaju „pravednu priliku da učestvuju“ [52]. Istovremeno, treba priznati da sistemi za preporuku zasnovani na ljudskim pravima mogu pozitivno da utiču na pluralizam, na primer u kontekstu autoritarizma i zarobljenih medija (media capture).

Važno je da se definiše šta predstavlja javni interes u pogledu algoritamskog priređivanja i deljenja sadržaja i kako to utiče na određivanje prioriteta različitih tipova sadržaja. Koncept i definicija sadržaja od javnog interesa jednako su sporni kao i definicija raznovrsne izloženosti. U principu, sadržaj od javnog interesa predstavlja onu informaciju o kojoj bi „javnost imala interes da bude informisana“.[53] Drugi način razmišljanja o sadržaju od javnog interesa je kao o sadržaju koji je relevantan za dobrobit građana, život zajednice ili lokalnog stanovništva. Očigledni primeri uključuju informacije o pandemiji KOVID-19 ili informacije koje se odnose na procese demokratskog glasanja. Obim povezanog, i ne uvek pouzdanog, sadržaja podstakao je posrednike da (javno) daju prioritet tačnosti. Nekoliko platformi je, u relativno kratkom vremenskom periodu, pokazalo svoju sposobnost da ponovo konfiguriše algoritme sistema za preporuku u nastojanju da filtriraju ili označe lažne informacije i daju prioritet sadržaju koji dolazi od pouzdanih organa za javno zdravlje. Uspeh ovih napora ili logika koja stoji iza motivacije za ove promene je, međutim, i dalje predmet žestoke debate,[54] dok su nedostatak transparentnosti u korišćenim podacima i izborima moderacije sadržaja koje donose platforme i dalje misterija. Ako debate ostavimo po strani, internet posrednici i društvene mreže – suočavajući se sa sve većim zahtevima javnosti i država da snose odgovornost za svest o javnom zdravlju – pokazali su svoju sposobnost da razmišljaju i da restrukturišu načine kojim daju prioritet i rangiraju sadržaj.[55]

52 K. Klonick, The New Governors, strana 1664.
53 M.E. Mazzoli and D. Tambini, Prioritisation uncovered: The Discoverability of Public Interest Content Online. Council of Europe (2020), strana 13.
54 M. Cinelli, The COVID-19 social media infodemic, The Nature (2020), strana 10; Videti i: Global Disinformation Index, Why is tech not defunding COVID-19 disinfo sites? (2020), Preuzeto sa: https://disinformationindex.org/2020/05/why-is-tech-not-defunding-covid-19-disinfo-sites.
55 Evropska komisija, Joint communication to the European Parliament, the European Council, The Council, The European Economic and Social Committee and the Committee of the Regions, Tackling COVID-19 disinformation - Getting the facts right, JOIN(2020) 8 final, 10 June 2020, section 5.

Ovo ukazuje na potrebu za većom pažnjom javnosti i političkim pritiskom na platforme da učine transparentnijim svoje procese preporučivanja i restrukturiraju sisteme za preporuke i njihove ciljeve optimizacije, kako bi se pozabavili strukturalnim problemima našeg savremenog medijskog okruženja. Ovo pitanje prevazilazi pitanja upravljanja sadržajem, a dodatno se tiče zakona o konkurenciji, vlasništva nad medijima i pravila koncentracije.[56] Takođe naglašava hitnu potrebu da se da prioritet ciljevima politike medijskog pluralizma i raznovrsnosti i intervencijama za podsticanje boljeg digitalnog prostora.

2.3 Agregacija vesti i medijski pluralizam

Agregatori vesti (news aggregators) funkcionišu kao centralno središte distribucije onlajn vesti, usmeravajući čitaoce na vesti i drugi sadržaj koji se smatra vestima (od strane agregatora vesti). Ovaj proces uglavnom sprovode algoritmi, zbog čega se agregatori vesti ponekad nazivaju „algoritamskim čuvarima".[57]

Agregatori vesti često podrazumevaju napetost između „algoritamske logike" i „uredničke logike".[58] „Algoritamska logika" značajno utiče na raznovrsnost kao i na politički diskurs tako što daje prednost novitetima, na primer, u odnosu na druge kriterijume vrednosti vesti (npr. relevantnost za javnost, raznovrsnost, itd.). Studija procesa prireďivanja i deljenja sadržaja iza AppleNews-a, koji koristi i ljudsku moderaciju (u Najvažnijim pričama (Top Stories)) i algoritamsko prireďivanje i deljenje sadržaja (u Pričama u trendu (Trending Stories)), pokazala je da sadržaj koji moderiraju ljudi sadrži „raznovrsniju i pravedniju distribuciju izvora od algoritamski odabranih" priča.[59] Priče u trendu, prema istoj studiji, skoro su isključivo obuhvatale „meke vesti" (npr. priče o slavnim ličnostima), dok su Najvažnije vesti rezervisane za „teške vesti" (npr. politički sadržaj).[60] Utvrđeno je da ove prakse ozbiljno utiču na pluralizam

56 M. E. Mazzoli i D. Tambini, Prioritisation uncovered: The Discoverability of Public Interest Content Online. Savet Evrope (2020), strana 23.
57 Napoli 2014.
58 T. Gillespie, PJ Boczkowski, KA Foot, Media technologies: Essays on communication, materiality, and society, MIT Press (2014).
59 J. Bandy i N. Diakopoulos, Auditing News Curation Systems: A Case Study Examining Algorithmic and Editorial Logic in Apple News, Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020), strana 43.
60 Ibid.

izvora i distribuciju vesti, a samim tim i na pluralizam sadržaja, na dva načina: Prvo, agregatori su stvorili takozvani „efekat širenja tržišta" jer su pojedincima omogućili izloženost informativnim medijima sa nižom popularnošću ili svešću o brendu. Drugo, primena agregatora je podstakla neke korisnike da ograniče ili zaustave direktno korišćenje informativnih medija, što je dovelo do takozvanog „efekta zamene". Pošto se reakcije korisnika obično zasnivaju na prvim utiscima, mamac za klikove (clickbait) se koristi u news feed-u kako bi se privukla pažnja i angažovali korisnici, čime se olakšava oglašavanje koje generiše profit. Ovaj pristup dodatno dovodi u pitanje održivost medija, a samim tim i nezavisnost i pluralizam, dodajući sveukupnom pritisku i finansijskim ograničenjima sa kojima se suočavaju konvencionalni mediji zbog koncentrisanog oglašavanja i modela eksploatacije podataka internet posrednika.

Postoji sve veća neravnoteža između dometa i uticaja komunikacija konvencionalnih medija i onlajn platformi, i kreiranja sadržaja u odnosu na priređivanje i distribuciju sadržaja. S obzirom da je tradicionalni poslovni model zasnovan na pretplati u opadanju, konvencionalni mediji se bore za održivost. Sve veći broj ljudi dobija vesti isključivo iz drugih izvora, sa većom verovatnoćom da će članci biti dostupni „besplatno". Spremnost da se plati za kvalitetne vesti je opala, dok je upotreba „besplatnih" agregatora vesti i društvenih mreža porasla. Kao rezultat toga, mnoge onlajn novinske kuće nemaju drugog izbora osim da traže nove tokove prihoda. Oni su primorani da usvoje mnoge prakse koje koriste velike platforme, a koje imaju značajnu moć nad logikom digitalne adtech industrije (npr. korišćenje targetiranog oglašavanja, objavljivanje sponzorisanog sadržaja ili prikupljanje i prodaja podataka korisnika). Ovaj trend ima duboko negativne implikacije na slobodu medija širom sveta: To stvara okruženje u kojem konvencionalne medijske organizacije moraju da se takmiče sa kompanijama društvenih mreža i posrednicima za iste izvore prihoda, dok su takođe izložene posredničkim sistemima preporuka i politikama priređivanja i deljenja sadržaja. Situacija doprinosi eroziji poverenja u medije, smanjenju odgovornosti za kreiranje i širenje dezinformacija i drugih problematičnih sadržaja.

Neke konvencionalne medijske organizacije i same koriste alate vođene algoritmom, pri čemu personalizacija i optimizacija sadržaja igraju integralnu ulogu u medijskim produkcijskim procesima. Ostaju velike razlike između „logike personalizacije vesti" i „platformske logike

personalizacije“[61] pri čemu su mediji podložni sistemskom nedostatku tehnoloških i finansijskih resursa, devalvaciji tradicionalne uređivačke i profesionalne etike, prevladavanju novonastalih privatnih interesa i ekonomskih podsticaja. Algoritamski modeli za priređivanje i deljenje sadržaja i strategije preporuka koje su razvili i primenili nezavisni konvencionalni mediji, a posebno javni servisi, mogli bi da ponude alternativne modele za bolje obezbeđivanje izloženosti publike raznovrsnostima, potencijalno čak nudeći i pojedincima modele za „raznovrsnost po dizajnu“.[62]

Dokazi pokazuju da novinari cene „uredničku logiku“, kao što je „transparentnost, raznovrsnost, uređivačka autonomija, široka ponuda informacija, lična relevantnost, upotrebljivost i iznenađenje“ u odnosu na poslovno vođenu algoritamsku logiku sistema za preporuku.[63]

Algoritamski vođeni procesi i prakse priređivanja i deljenja sadržaja i preporuka predstavljaju pretnju medijskom pluralizmu i raznovrsnosti medija i izazivaju zabrinutost u vezi sa punim uživanjem prava na slobodu izražavanja. Glavni faktori koji doprinose ovim pretnjama su:

- Finansijska nestabilnost i fiskalni pritisak na konvencionalne medije: Onlajn platforme su stekle ogromnu ekonomsku moć, prvenstveno kroz prihode od oglašavanja, i one koriste ovu polugu kako bi diktirale uslove za priređivanje i deljenje svih onlajn sadržaja, uključujući uređivačke medije i vesti. Ova neravnoteža moći uključuje neravnotežu „snage mišljenja“[64] i moć da se

61 B. Bodó, Selling News to Audiences – A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media, Digital Journalism (2019), strana 17-18.
62 Vidite više o ovom sadržaju u N.Helberger, Diversity by design — Diversity of content in the digital age, Government of Canada (2020), str. 8; Natali Helberger, Kari Karppinen & Lucia D'Acunto (2018) Exposure diversity as a design principle for recommender systems, Information, Communication & Society, 21:2, 191-207, DOI: 10.1080/1369118X.2016.1271900.
63 Ova studija uključuje novinske agencije iz Holandije i Švajcarke; M. Bastian, N.Helberger & M. Makhortykh, Safeguarding the Journalistic DNA: Attitudes towards the Role of Professional Values in Algorithmic News Recommender Designs, Digital Journalism (2021), strana 21.
64 N. Helberger, The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power. Digital Journalism, 8(6), strana 842-854 (2020).

„utiče na procese formiranja individualnog i javnog mnjenja", što zauzvrat omogućava da „ove platforme promene samu strukturu i ravnotežu medijskog tržišta, a samim tim direktno i trajno utiču na pluralističku javnu sferu."[65]

- Konvencionalni mediji, primorani da usvoje slične poslovne modele i „logiku" društvenih mreža, propuštaju priliku da promene „pravila novih komunikacionih poredaka",[66] i da doprinesu raznovrsnijem medijskom pejzažu. Ipak, postoje alternativni modeli algoritamskog priređivanja i deljenja koji su usredsređeni na sadržaj od javnog interesa i profesionalne novinarske prakse – koje obično uspostavljaju konvencionalne i javne medijske organizacije – i oni nude alternativne modele za ublažavanje potencijalnih problema uzrokovanih nedostatkom davanja prioriteta sadržaju od javnog interesa.[67]

- Poboljšanje algoritamskog priređivanja i deljenja sadržaja sa ciljem povećanja raznovrsnosti medijskih izvora predstavlja nekoliko izazova:

    - Čak i ako internet posrednici i kompanije društvenih mreža „treniraju i igraju se" algoritmima „sa dobrim ciljem" – da izlože heterogenu publiku heterogenom sadržaju – ovim praksama nedostaje značajna transparentnost, a pojedinci nemaju nikakav uticaj u pogledu dizajna i logike koja upravlja ovim sistemima. Ovo predstavlja značajan i sistemski rizik za uživanje slobode izražavanja.

    - Dok bi personalizovani sadržaj i procesi optimizacije mogli da doprinesu ispunjavanju različitih individualnih, grupnih i društvenih potreba – i da generišu potencijal za raznovrsnost – takav potencijal treba da bude vođen ciljevima javne politike i odgovarajućim intervencijama.

65 Ibid, strana 846.
66 Za detaljniju diskusiju o ovom problemu, vidite: N. Helberger,The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power. Digital Journalism, 8(6),strana 842-854 (2020).
67 M.E,Mazzoli i D.Tambini,Prioritisation uncovered: The Discoverability of Public Interest Content Online. Savet Evrope (2020).

- Malo je ili uopšte nema informacija o tome kako sadržaj proizveden od strane i za marginalizovane zajednice kruži onlajn, i kako se preporuke odnose prema takvim sadržajima. Studije sugerišu[68] da se određeni sadržaj i govor različito tretiraju, što dovodi do zabrinutosti da sadržaj nije jednako dostupan i da nisu razvijene ili primenjene mere zaštite za sprečavanje diskriminatornih algoritamskih ishoda i obezbeđivanje pravičnog i ravnopravnog učešća javnosti i odlučivanja.

- Posledice priređivanja i deljenja sadržaja vođenog algoritmom mogu da povećaju zloupotrebe ljudskih prava i kršenja vladavine prava kada se pojačaju u određenim nacionalnim kontekstima, posebno u vezi sa sistemskim (državnim i privatnim) zarobljavanjem medija i monopolizovanom kontrolom javnog dijaloga. U ovim okolnostima, dodatni slojevi algoritamski-vođenih ograničenja nad i za medijski pluralizam i raznovrsnost još više pojačavaju zbirni iznos individualnog gubitka prava na slobodu izražavanja.

68 Vidite, na primer: A. Chinmayi, Facebook's Faces, Forthcoming Harvard Law Review Forum Volume 135 (2021) i K. Klonick, The New Governors: The People, Rules, and Processes Governing Online Speech, The Harvard Law Review (2018); C. O'Neil, Facebook's VIP "Whitelist" Reveals Two Big Problems, Bloomberg Opinion (2021), https://www.bloomberg.com/opinion/articles/2021-09-15/facebook-s-xcheck-vip-whitelist-reveals-two-big-problems.

Medijski pluralizam

3. Preporuke zasnovane na ljudskim pravima o upotrebi VI u priređivanju i deljenju sadržaja

Države su prvenstveno garanti pluralizma medija u skladu sa međunarodnim okvirom zaštite ljudskih prava. One treba da deluju kao krajnji garanti za uživanje ljudskih prava, uključujući odgovornost za podsticajno okruženje za prava na slobodu izražavanja i slobodu medija. Sledeće preporuke za države članice OEBS-a, nastale tokom radionice, fokusiraju se na: jačanje pluralističkog medijskog pejzaža i pluralizma glasova (3.1); negovanje podsticajnog okruženja za raznovrsnost medijskog sadržaja i individualnu izloženost različitim medijima (3.2); i omogućavanje individualnog delovanja i kontrole (3.3).

3.1 Preporuke za jačanje pluralističkog medijskog pejzaža i mnoštva glasova

Ovaj deo izveštaja nastoji da državama članicama ponudi normativnu agendu koja neguje pluralističko medijsko okruženje i „koegzistenciju raznovrsnih i konkurentnih interesa – što predstavlja osnovu za demokratski ekvilibrijum".[69] Ova agenda je ograničena smanjenim mogućnostima za demokratski vođen medijski prostor, neuravnoteženom dominacijom digitalne platforme i preteranom koncentracijom na tržištu. Države članice treba da obezbede uslove za medijske inovacije, nezavisnost i održivost, posebno medije vođene javnim interesom, i da primenjuju modele kreiranja, priređivanja i deljenja i distribucije sadržaja koji neguju ove uslove.

- **Države treba da obezbede, kroz regulatorne inicijative, jednake uslove za sve medijske aktere, uklanjanjem prepreka za obezbeđivanje pravičnih i delotvornih tržišnih uslova.** Nastali tržišni uslovi trebalo bi da omoguće svim medijima pristup i korišćenje novih tehnologija i razvoj alternativnih poslovnih modela – uključujući alternativne modele za algoritamsko priređivanje i deljenje sadržaja koji neguju raznovrstan medijski pejzaž i širenje sadržaja od javnog interesa.

69 A. Roksa-Zubcevic et al, Media Regulatory Authorities and media pluralism, Regional Publication. Savet Evrope (2021), strana 12-14.

- **Države treba da analiziraju na koji se način postojeća i buduća politika vezana za pluralizam medija bavi pitanjem sadržaja od javnog interesa,** posebno u svetlu značaja onlajn platformi u distribuciji informacija o javnom zdravlju tokom pandemije KOVID-19.

- **Javno-privatna partnerstva između država i kompanija društvenih mreža i drugih posrednika treba da budu strogo transparentna** i podložna nadzoru građana i javnosti. Ovo bi trebalo da uključi regulatorni okvir medijskog pluralizma.

- **Države treba da primenjuju javne politike i zakone kako bi sprečile neuravnoteženu i monopolizovanu tržišnu moć koja u ovom momentu postoji,** posebno u pogledu internet posrednika i distribucije sadržaja koju kontroliše država. Svaka intervencija države mora da osigura prodemokratski režim koji je istinski nezavisan i nudi strukturna rešenja za jačanje pluralizma.

- **Države bi trebalo da promovišu pluralizam i tehnološke i medijske inovacije** finansiranjem holističkog nezavisnog istraživanja koje pomaže medijskim akterima, institucijama javnog nadzora i akademskoj zajednici da shvate trenutnu raspodelu moći – posebno u pogledu efekata sistema za preporuku, analize logike preporuka i rezultirajućeg uticaja na medijski pluralizam i raznovrsnost.

3.2. Preporuke za osnaživanje okruženja koje podstiče raznovrsnost medijskih sadržaja i individualnu izloženost pluralističkom informisanju

Ovaj deo izveštaja govori o tome da li i kako internet posrednici treba da obezbede pojedincima jednak pristup javnim prostorima i učešće u njima, ispitujući raznovrsnost kao normativni koncept.

- **Regulatorne intervencije i javne politike država treba da očuvaju i neguju internet kao prostor za demokratsko učešće i predstavljanje.** Svako nacionalno uređenje digitalnog prostora

treba da ima jasno definisan obim koji je neophodan i srazmeran transparentnom cilju, u punoj saglasnosti sa međunarodnim okvirom ljudskih prava.

- **Države bi trebalo da se angažuju i podrže međusektorski dijalog kako bi prikupile najaktuelnije i najrelevantnije podatke o uticajima algoritamskog priređivanja i deljenja sadržaja,** kao što su polarizacija, nedostaci u informacijama, itd. Nezavisan nadzor i transparentnost praćenja raznovrsnosti zahteva **multidisciplinarni pristup, predvođen akademskim institucijama ili organizacijama civilnog društva, uz podršku države.** Međusektorsko praćenje raznovrsnosti trebalo bi da se koristi za utvrđivanje sadržaja i publike koji su u opasnosti od isključivanja iz učešća i/ili predstavljanja javnosti ili su se u prošlosti suočavali sa tim.

- **Države treba da usvoje inkluzivan pristup i obezbede učešće više zainteresovanih strana, kao i vlasništvo nad algoritamskim priređivanjem i deljenjem sadržaja.** Demokratije nisu sistemi koji samostalno održavaju kontinuitet. Da bi demokratije napredovale, građani moraju da imaju sposobnost da donose utemeljene odluke. Kroz obezbeđivanje otvorenog dijaloga i međusektorske saradnje sa internet posrednicima, države bi mogle – zajedno sa organizacijama civilnog društva, marginalizovanim zajednicama, medijskim organizacijama, novinarima i njihovim predstavnicima – da podstiču održivu međusektorsku saradnju, uključujući i saradnju između državnih i nedržavnih aktera. Osim toga, države treba da se zalažu za raznovrsnost u timovima programera koji kreiraju algoritamske sisteme za priređivanje i deljenje sadržaja, tako da različiti interesi i perspektive budu zastupljeni u dizajnu i implementaciji algoritama.

- **Države treba da obezbede podršku i resurse postojećim nezavisnim medijskim regulatornim telima** koja koriste proces zajedničkog stvaranja i uključivanja svih nacionalnih medijskih aktera i stručnjaka radi pružanja podrške ekonomskom, pravnom i političkom okruženju u kome se raznovrsnost neguje kao osnovni demokratski cilj.

- **Države treba da razviju zakonodavni okvir zasnovan na dokazima i istraživanjima kako bi osigurale odgovornost internet posrednika, uključujući i obavezu detaljne provere ljudskih prava.** Procene

uticaja na ljudska prava treba da budu deo svake strategije za smanjenje rizika ili bilo koje spoljne revizije kako bi se obezbedio javni nadzor.

- **Države treba da ojačaju nezavisna regulatorna tela za medije i druge nadležne institucije i uključe ih u javni nadzor i istraživanje.** Na primer, ova tela bi trebalo da se uključe u procene uticaja na ljudska prava kako bi se pozabavili rizicima koje internet posrednici predstavljaju za raznovrsnost i pluralizam, uključujući rizike koje predstavljaju za marginalizovane zajednice. Takve procene bi trebalo da budu praćene mehanizmima odgovornosti, a trebalo bi da postoji transparentno obelodanjivanje i objavljivanje procena, revizija i sl.

- **Države treba da povećaju javno finansiranje nezavisnog, kvalitetnog novinarstva i/ili da obezbede finansijska sredstva nezavisnim akterima** sa relevantnom stručnošću i dokazanim iskustvom na polju ljudskih prava. Ovi nezavisni akteri mogu da ponude alternative postojećim poslovnim modelima orijentisanim na prihod i zasnovanim na podacima, na taj način podstičući decentralizovane tehnološke algoritamske sisteme za priređivanje i deljenje koji promovišu javne vrednosti, kao što su raznovrsnost medija, inkluzivnost i tolerancija.

- **Države treba da obezbede da svaka potencijalna intervencija u ovoj oblasti ne ograničava pozitivnu funkcionalnost personalizacije ili nezavisnosti medija,** dok u isto vreme pruža podršku za omogućavanje da raznovrsnost sadržaja i da sadržaji od javnog interesa budu u fokusu dizajna. Personalizacija može da bude dragocena za pojedince kada se koristi za preciziranje pretraga i ubrzanje pronalaženja informacija.

- **Države treba da uspostave i zaštite adekvatne okvire za pristup podacima koji omogućavaju proverenim istraživačima, organizacijama civilnog društva i drugim nezavisnim zainteresovanim stranama, kao što su mediji, da pristupe podacima koje poseduju internet posrednici.** U isto vreme, zloupotreba takvog okvira treba da se spreči kroz korišćenje etičkih smernica ili uspostavljanje nezavisnog organa sa nadzornom funkcijom.

3.3. Preporuka o omogućavanju individualnog posredovanja i kontrole

Osnaživanje pojedinaca treba da bude ugrađeno u algoritamski dizajn, a slično tome, javni interes i dizajn usredsređen na ljudska prava treba da budu u prvom planu algoritamske primene.

- **Države treba da podrže inicijative za samoregulaciju i koregulaciju, i da stvore uslove koji kodifikuju individualnu kontrolu nad onim što se vidi onlajn.** Ovo bi moglo da se postigne obaveznim opcijama koje su propisane zakonom, kao što je podrazumevano prihvatanje sistema (opt-in by default) za preporuke sadržaja i laka identifikacija i izbor za definisanje uredničke i neuredničke personalizacije.

- **Države bi trebalo da nalože transparentnost i objašnjivost unapred odabrane personalizacije vesti i obrade podataka.** Ovo bi trebalo da uključi transparentnost kriterijuma, principa i tipova aranžmana koji utiču na odluku o određivanju prioriteta sadržaja (da bi se podstaklo poverenje javnosti i omogućilo javnosti da razume da li se razmatraju ciljevi od komercijalnog ili javnog interesa). Države bi takođe trebalo da zahtevaju od posrednika da uvedu odgovarajuće procesne mere. Na primer, kada posrednici ograniče news feed-ove, oni bi trebalo da informišu pogođene pojedince o njihovim politikama i da obezbede delotvorne mehanizme pravne zaštite. Slično tome, **države treba da podrže inicijative za samoregulaciju i koregulaciju koje obezbeđuju posredničku transparentnost** u procesima koji dovode do odluka o određivanju prioriteta sadržaja.

- **Države treba da uspostave održive programe medijske i digitalne pismenosti za sve društvene grupe.** Pojedinci često nisu svesni, i/ili ne shvataju, implikacije algoritamskog prireďivanja i deljenja sadržaja na njihovo uživanje ljudskih prava i osnovnih sloboda.

- **Države treba da stave poseban fokus na negovanje prava na pristup, traženje i prenošenje mišljenja i ideja svih vrsta među svim starosnim grupama. Države posebno treba da osnaže proces formiranja individualnog mišljenja,** uključujući i mlade ljude, koji su redovno lišeni odgovarajućeg pristupa konvencionalnim medijskim sadržajima.

4. Zaključak

Priređivanje i deljenje sadržaja je, barem u kontekstu medijskog pluralizma i medijske raznovrsnosti, uglavnom gurnuto na margine širih diskusija o upravljanju sadržajem. Ovaj propust, dopunjen nedostatkom razumevanja važnosti raznovrsnosti medija i informisanja za heterogenu publiku, dovodi do štete koja je podjednako zabrinjavajuća kao i rizici koji proizilaze iz nezakonitog sadržaja i dezinformacija. Svi algoritamski vođeni procesi priređivanja i deljenja i moderacije sadržaja suštinski su povezani i moraju se kao takvi tretirati.[70] Ovi procesi su posebno važni jer algoritmi određuju ono šta pojedinci vide, koje informacije imaju prioritet i koji sadržaj je isključen. Onlajn čuvari se sve više oslanjaju na sisteme za preporuku koji sistematski analiziraju obrasce ponašanja korisnika i kreiraju profile kako bi utvrdili koje informacije će verovatnije zainteresovati datog korisnika. Drugim rečima, čuvari prikupljaju podatke kako bi odredili koji personalizovani sadržaj da ponude pojedinačnim korisnicima, kako bi podstakli njihovo angažovanje i prikupili više podataka o njima – čak i ako su donete odluke u suprotnosti sa demokratskim diskursom, raznovrsnošću informacija, medijskim pluralizmom i pravom na privatnost. Iz tog razloga, i kao što je jasno uobličeno u preporukama, interdisciplinarno istraživanje i transparentnost posredničkih politika i praksi za priređivanje i deljenje sadržaja, predstavljaju ključne preduslove za stavljanje medijskog pluralizma i raznovrsnosti u dizajnu i primeni algoritama u fokus.

Ovaj izveštaj o ishodu naglašava problematičnu prirodu personalizovanih sistema za preporuke sadržaja koje koriste internet posrednici, a posebno društvene mreže. On ističe zabrinjavajuće implikacije koje ovi sistemi imaju na društvenu koheziju, raznovrsnost, kvalitet informacija u okviru javnog diskursa i na privatnost. Na individualnom nivou, onlajn iskustvo je strateški obojeno odlukama donetim radi profita, implementiranim putem algoritama bez svesti o pogođenim pojedincima ili nadzora javnih organa, i zasnovano na nametljivom prikupljanju i analizi podataka koji su dizajnirani da zaobiđu zakone o privatnost i zaštiti podataka – sa velikim negativnim uticajem na raznovrsnost informacija, medijski pluralizam i pravo na privatnost.

70 Jednostavnim rečima, preterano uklanjanje „legitimnog“ sadržaja je u stvari i rizik za medijsku raznolikost.

Postoji dovoljno dokaza da uticaj onlajn platformi na mišljenja ima sposobnost da upravlja i pojača određene javne narative i tipove diskursa u odnosu na druge. Za zemlje sa krhkim ili opresivnim političkim sistemima, ova moć mišljenja, zajedno sa algoritamskim pojačavanjem, može da ima katastrofalne posledice po individualno uživanje ljudskih prava. Rangiranjem i razlikovanjem sadržaja i rezultata preporuka, internet posrednici rekonfigurišu javnu debatu na način koji osnažuje one koji su već na privilegovanim pozicijama. Cena za to je smanjenje različitosti uopšte, a posebno je nepovoljna za istorijski marginalizovane grupe, koje su i dalje gurnute na margine javnih rasprava u procesu koji ponovo stvara i jača nejednakost i nepravdu. Dok algoritamsko prireďivanje i deljenje sadržaja ima moć da ograniči učešće, stvori podelu i ograniči širenje informacija, medijska raznovrsnost podstiče društvenu koheziju, tolerantnost i distribuciju komunikacijske moći. Na državama je, pre svega, ali i nedržavnim akterima, prvenstveno posrednicima i medijskim organizacijama, da obezbede da pluralizam medija, jednak pristup i puno uživanje ljudskih prava budu osnova za pravila koja utiču na informacioni prostor onlajn, jer su to sastavni delovi za istinski demokratska digitalna društva.

Veštačka inteligencija u priređivanju i deljenju sadržaja i oglašavanju zasnovanom na praćenju

Ovaj deo se fokusira na upotrebu veštačke inteligencije u priređivanju i deljenju sadržaja, uz akcenat na vezi između kapitalizma praćenja i targetiranog oglašavanja, i na shodni uticaj na slobodu mišljenja i izražavanja. Ističe nedostatke priređivanja i deljenja sadržaja zasnovanog na veštačkoj inteligenciji i targetiranom oglašavanju i pruža preporuke koje se odnose na ljudska prava za države članice OEBS-a, da se pozabave negativnim uticajem koji alati veštačke inteligencije u priređivanju i deljenju sadržaja imaju na pravo na slobodu mišljenja i izražavanja.

1. Definisanje obima uticaja poslovnih modela zasnovanih na praćenju u njihovom korišćenju za priređivanje i deljenje sadržaja

1.1 Uticaj automatskog donošenja odluka na pravo na slobodu mišljenja

Međunarodni okvir ljudskih prava pravi razliku između unutrašnje i spoljašnje dimenzije prava na slobodu mišljenja. Dok spoljašnja dimenzija ovog prava može da bude podložna legitimnim, proporcionalnim i nediskriminatorskim ograničenjima koja su neophodna u demokratskom društvu, unutrašnja dimenzija slobode mišljenja, takozvani forum internum, je apsolutna i od nje se ne može odstupiti.[71] Član 19. Univerzalne deklaracije o ljudskim pravima, kao i Međunarodni pakt o građanskim i političkim pravima štite ovo apsolutno pravo od bilo kakvog ograničenja ili uplitanja. Prema rečima specijalnog izvestioca UN-a za slobodu izražavanja i mišljenja, „zabranjeno je svako nedobrovoljno otkrivanje mišljenja i mentalna autonomija je afirmisana."[72]

71 Kancelarija visokog komesara Ujedinjenih nacija za ljudska prava, CCPR General Comment No. 22: Article 188 (Freedom of Thought, Conscience and Religion),dostupno na https://www.refworld.org/docid/453883fb22.html, 1993.
72 Izveštaj Specijalnog izvestioca za promociju i zaštitu prava na slobodu mišljenja i izražavanja, Irene Khan, Disinformation and freedom of opinion and expression, https://undocs.org/A/HRC/47/25, 2021.

Poslovni modeli velikih onlajn platformi za prikupljanje podataka omogućavaju industriji oglašavanja da razvije ili da se osloni na strategije targetiranja zasnovane na podacima. Kroz ovaj pristup, kompanije identifikuju i iskorišćavaju obrasce ponašanja i karakteristike ljudi ili zajednica. Krovni termin koji pokriva ove manipulativne tehnike je „oglašavanje zasnovano na praćenju", shvaćeno kao opšti termin za digitalno oglašavanje koje targetira pojedince ili grupe, uglavnom kroz praćenje i profilisanje na osnovu ličnih podataka. Kontekst gde se određeni oglas postavlja može biti nasumičan, jer, pošto targetira pojedince, on može da ih prati u različitim kontekstima.[73] U većini slučajeva, oglašavanje zasnovano na praćenju predstavlja deo automatizovanog procesa, kojim se svaki pojedinačni oglas bira i postavlja za samo nekoliko milisekundi. To znači da ni onaj koji objavljuje oglas (npr. vlasnik vebsajtaili aplikacije) niti onaj koji oglašava (npr. vlasnik brenda koji se promoviše) ne biraju koje će oglase kome prikazati ili gde će ih prikazati. O tome automatski odlučuju tehnološki sistemi koje često kontrolišu treći posrednici (takozvane „adtech" kompanije).[74]

Oglašavanje zasnovano na praćenju značajno je doprinelo iskorišćavanju posebnih karakteristika ljudi kako bi se povećala ubedljivost same poruke, čime se neopravdano ometala njihova apsolutna sloboda da formiraju mišljenje i da uživaju u nezavisnim misaonim procesima. Ljudi koji koriste usluge platformi su izmanipulisani da razmišljaju ili da donose odluke koje inače možda nikada ne bi doneli. Oglasi zasnovani na praćenju iskorišćavaju ranjivosti pojedinaca čak i ako ne identifikuju direktno te ranjivosti. Korišćenjem takozvane „slične publike", oglašivači mogu da dupliraju grupe ljudi sa određenim karakteristikama kako bi došli do novih pojedinaca koji dele iste karakteristike. Automatizovani alati i dominacija nekoliko onlajn platformi omogućili su veću manipulaciju jer svaki pojedinac koji koristi njihovu uslugu može biti targetiran sve vreme i u bilo kom trenutku.

73 Norwegian Consumer Council, Time to ban surveillance-based advertising: The case against commercial surveillance online, dostupno na: https://www.forbrukerradet.no/wp-content/uploads/2021/06/20210622-final-report-time-to-ban-surveillance-based-advertising.pdf, 2021.

74 Norwegian Consumer Council, Out of control: How consumers are exploited by the online advertising industry, dostupno na: https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf, 2020.

Sve je veći zahtev za zabranom praksi koje negativno utiču na apsolutno pravo ljudi na slobodu mišljenja i slobodu misli, posebno pošto su sloboda misli i mišljenja pojedinaca pod velikim uticajem bez njihovog znanja ili pristanka. Ovaj fenomen posebno uključuje targetirano praćenje ponašanja i pojedinačno praćenje kroz više vebsajtova/i preko više uređaja. Takvo stalno invazivno korporativno praćenje predstavlja rizik od sistematske manipulacije pojedincima, mimo tradicionalnih oblika reklamnog uticaja. Oglašavanje zasnovano na praćenju targetira pojedince na netransparentne načine[75] i može da iskoristiti ranjivosti, otvarajući nove mogućnosti za manipulaciju. Konkretno, kada se kombinuje sa algoritmima za maksimiziranje prihoda od priređivanja i deljenja sadržaja, oglašavanje zasnovano na praćenju može da utiče na način na koji pojedinci govore i kako se ponašaju, time utičući na raznovrsnost informacija, pogleda i mišljenja.

Uprkos tvrdnjama onlajn platformi da nema povratka sa oglašavanja zasnovanog na praćenju, internet nije izgrađen na poslovnom modelu „prikradajućih oglasa"(„creepy ad"). U stvari, situacija je sasvim drugačija. Države moraju da izbegavaju da direktno ili indirektno štite poslovne modele koji se baziraju na reklamama zasnovanim na praćenju i kršenju međunarodnog prava o ljudskim pravima. Ukidanje zloupotrebljavajućih modela takođe znači otvaranje vrata alternativama koje su u skladu sa ljudskim pravima, kao i inovativnim oblicima kontekstualnog oglašavanja koji se oslanjaju na minimalnu personalizaciju bez individualnog targetiranja.[76] Ovo će takođe omogućiti da novi igrači uđu na tržište.

Oglašavanje zasnovano na praćenju ima dalekosežne uticaje na lične interakcije ljudi, njihove izbore i učešće u demokratskim debatama. Mere namenjene povećanju transparentnosti mogu da pomognu da se bolje razumeju razmere problema, ali one nisu dovoljne da spreče i ublaže povrede ljudskih prava koje postoje. Individualne i društvene štete nastale nametljivim targetiranjem i personalizacijom zahtevaju sistematski odgovor. Od upada u privatnost do priređivanja i deljenja sadržaja, invazivno praćenje ugrožava pravo na slobodu mišljenja na vidan način.

75 Napori civilnog društva da se zadobije više transparentnosti naišli su na prepreke. https://algorithmwatch.org/en/defend-public-interest-research-on-platforms/.
76 Natasha Lomas, Data from Dutch public broadcaster shows the value of ditching creepy ads, dostupno na: https://techcrunch.com/2020/07/24/data-from-dutch-public-broadcaster-shows-the-value-of-ditching-creepy-ads/?guccounter=1, 2020.

Pozitivna je obaveza država da zaštite ovo apsolutno pravo od takvih mešanja stvaranjem adekvatnog regulatornog okvira koji uspostavlja i sprovodi snažne mere zaštite ljudskih prava.

1.2 Smernice o onlajn targetiranju

Dok mnogi posrednici targetiraju svoje korisnike (kao i one koji nisu njihovi korisnici) putem bihevioralnog profilisanja i međulokacijskog praćenja, onlajn čuvari sa pristupom velikim količinama podataka korisnika, bez presedana, vode adtech industriju. U praksi, oglašavanje zasnovano na praćenju počinje od onoga koji objavljuje oglas, koji upravlja vebsajtom ili mobilnom aplikacijom koja pruža uslugu ili sadržaj. Oni pružaju prostor za postavljanje oglasa na svojim platformama i/ili pristup podacima o svojim korisnicima. Njihovi trgovinski partneri su prodavci na tržištu, kompanije koje žele da prodaju svoje proizvode najvrednijim kupcima. Ali prodavci posrednici i onlajn razmene oglasa nalaze se između ovih aktera. Oni funkcionišu u senci, nemaju direktan odnos sa korisnicima, odlučuju koji će oglasi biti postavljeni na kojim vebsajtovima i dobijaju deo transakcije. Ova komplikovana mreža za oglašavanje prikuplja, analizira i spaja velike količine ličnih podataka bez znanja ljudi. Ni oni koji objavljuju ni trgovci nisu u mogućnosti da u potpunosti, niti čak delimično kontrolišu ovaj proces.

Glavni internet posrednici potvrđuju dominaciju na tržištu u celom reklamnom ekosistemu tako što igraju sve tri uloge istovremeno – deluju kao oni koji objavljuju oglase, prodavci na tržištu i prodavci posrednici. Njihova dominacija dodatno je ojačana njihovim praktično neograničenim pristupom podacima, uključujući podatke iz sopstvenih usluga i podatke trećih strana. Ovo stvara ogromnu neravnotežu moći koja podstiče nelojalnu konkurenciju na digitalnom tržištu i predstavlja rizik od sistemske zloupotrebe ljudskih prava.

Glavni fokus ovog dela izveštaja su internet posrednici čiji se poslovni modeli u velikoj meri oslanjaju na onlajn targetiranje. Dok mnogi posrednici targetiraju svoje korisnike putem bihevioralnih podataka i međulokacijskog praćenja, što per se predstavlja prakse koje ugrožavaju ljudska prava, ovi onlajn čuvari sa neviđenim pristupom velikim količinama podataka korisnika takođe su lideri u adtech industriji. Na primer, velike

društvene mreže kao što je Facebook razvile su veoma granularne sisteme za svoj interfejs za oglašavanje zahvaljujući kojima kontrolišu velike prihode od oglašavanja na globalnom nivou. Organizacija za digitalna prava Panoptykon je mapirala i detaljno opisala adtech ekosistem koji je razvio Facebook čuvar, kao i njegov uticaj na ljudska prava. Panoptykon ističe da Facebook nije samo pasivni posrednik između oglašivača i korisnika.[77] On omogućava oglašivačima da izaberu kriterijume koje zatim tumači Facebook-ov algoritam kako bi postigli željene ciljeve oglašivača.

Ovaj izveštaj koristi Facebook kao praktičan primer da pokaže kako oglašavanje zasnovano na praćenju funkcioniše u praksi. Prvo, oglašivači mogu da izaberu svoju ciljnu publiku na osnovu kriterijuma targetiranja koje određuje posrednik. Postoji niz kriterijuma koje oglašivači mogu da koriste. Između ostalog,[78] oglašivači mogu da izaberu da targetiraju prilagođenu publiku ili takozvane sličnosti. Facebook je uveo oba kriterijuma poslednjih godina i oni se često opisuju na sledeći način:

- **Kriterijum prilagođene publike** zasniva se na informacijama oglašivača koje oni imaju o svojim korisnicima i koje mogu da otpreme posredniku. Shodno tome, posrednički algoritam povezuje ove informacije sa sopstvenim podacima o korisnicima – bez otkrivanja profila korisnika oglašivačima.

- **Kriterijum slične publike** omogućava oglašivačima da targetiraju grupu korisnika koja je slična prvobitno željenoj. U praksi, posrednik predviđa koja publika ima zajedničke karakteristike sa prvobitnom targetiranom grupom – takozvanom „početnom publikom". [79] Slične publike se identifikuju posredničkim algoritmom za podudaranje.

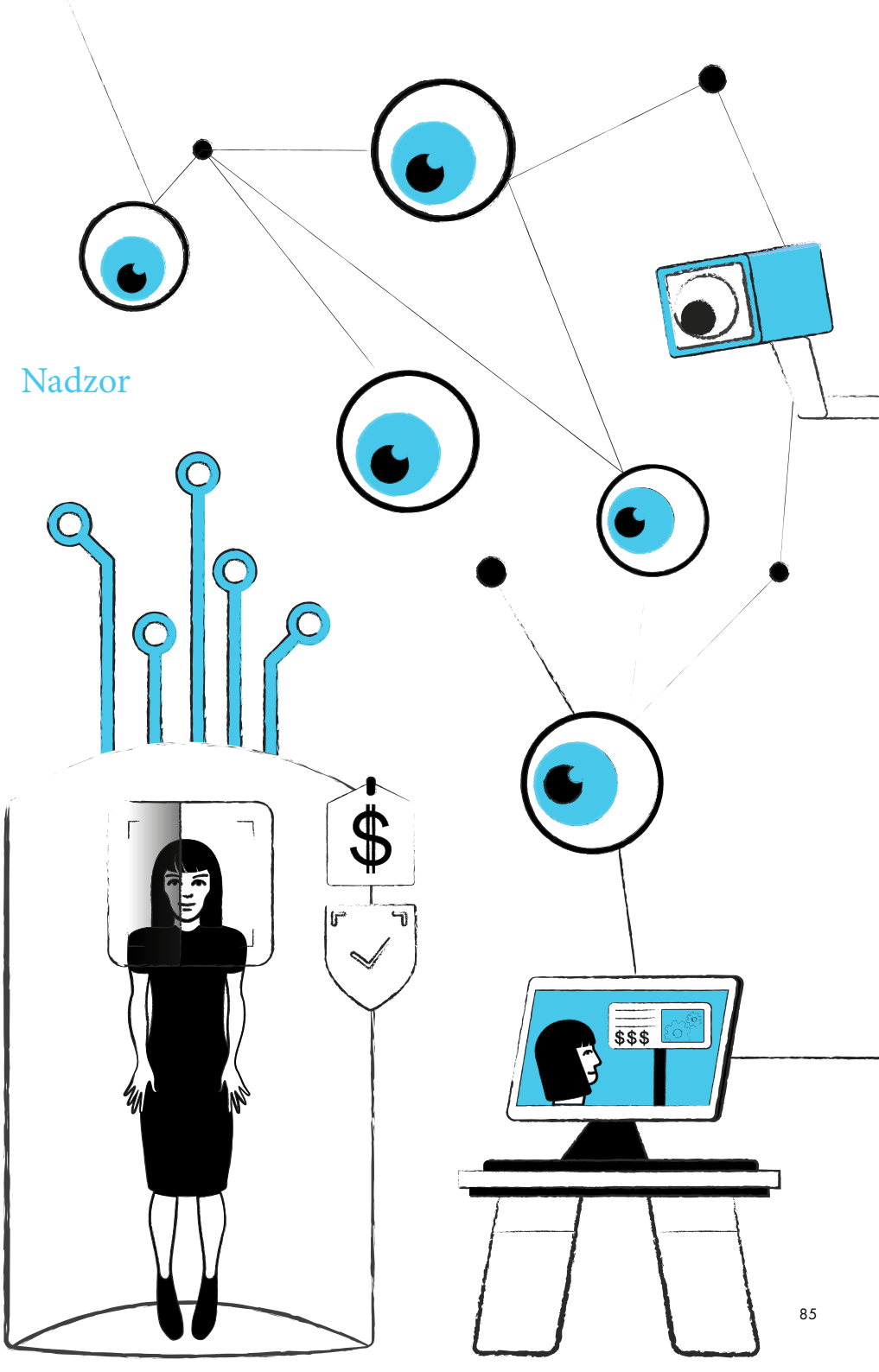
Onlajn platforme su u stanju da targetiraju pojedince sa velikom preciznošću jer poseduju podatke i znanje o pojedinačnim korisnicima i onima koji nisu korisnici. Analiza velikih podataka (big data) omogućava im da predvide ponašanje pojedinaca, koristeći podatke koje direktno pružaju korisnici ili dobijaju posmatranjem onlajn aktivnosti

77 Panoptykon Foundation, Who (really) targets you? Facebook in Polish election campaigns, 2020.
78 Ibid.
79 Ibid. Pogledati, Norwegian Consumer Council, Out of control: How consumers are exploited by the online advertising industry, 2020.

i bihevioralnih obrazaca korisnika i drugih. Visoko osetljivi algoritmi kreiraju profile na osnovu bihevioralnih podataka – navika, preferenci, nesviđanja (dislajkova) i interakcije sa korisnicima. Ovi profili mogu čak da uključe zaključke izvučene iz vremena kada su korisnici najaktivniji onlajn. I kreiranje i naknadno korišćenje profila zadiru u privatnost i uključuju pretpostavke, i mogu da dovedu do diskriminacije. Algoritmi takođe mogu da izvuku dodatne informacije o pojedincima koje targetirani pojedinac ne namerava da otkrije. Obrazloženje iza ove tehnike je ideja da što više kompanije znaju o svojim korisnicima, veća je verovatnoća da mogu uspešno da predvide i potencijalno manipulišu njima. Ove informacije se zatim koriste za isporuku specifičnog sadržaja i reklamiranja „u pravo vreme i u pravom kontekstu“, da podstaknu korisnike da kupe određene proizvode ili usluge ili da gledaju određene video snimke.[80]

80 V. Joler, The Human Fabric of the Facebook Pyramid, SHARE Lab Foundation, 2017.

Nadzor



2. Preporuke o regulisanju oglašavanja zasnovanog na praćenju usredsređene na ljudska prava

2.1 Preporuke za dodatno osnaživanje korisnika i ličnog posredovanja u onlajn ekosistemu

- **Države treba da zauzmu pristup usredsređen na ljude i korisnike u cilju dodatnog osnaživanja korisnika, individualnog posredovanja i kontrole nad njihovim podacima.** Postoji značajan rizik koji stvara naša nemogućnost da znamo da li smo profilisani ili identifikovani, kako smo profilisani ili identifikovani i kojim algoritmom. Ipak, sama transparentnost nije dovoljna da ljudi kontrolišu upotrebu sopstvenih informacija. Transparentnost bi trebalo da bude u kombinaciji sa snažnim, delotvornim pravima da se takva postupanja odbace.

- **Države treba da obezbede da osnaživanje korisnika i povećavanje mogućnosti njihovog individualnog delovanja i stvaranje komplementarnih sistema spoljnog nadzora i istrage ne budu međusobno isključivi.** Važno je dati prioritet korisničkom posredovanju i kontroli u trenutnom društveno-političkom okruženju.

- **Države bi trebalo da ulažu u istraživanje** radi razvoja empirijske osnove za utvrđivanje i razumevanje efekata oglašavanja zasnovanog na praćenju na autonomiju i delovanje korisnika. Bez daljih empirijskih studija, može doći do prevelikog pojednostavljenja korisničkih iskustava na osnovu oskudnih podataka i istraživanja koja su fokusirana na određene onlajn zajednice.

- **Države treba da promovišu regulatorni okvir za poboljšanje informacija koje se distribuiraju korisnicima,** kako bi omogućile korisnicima da vrše slobodan izbor u oglasima koje gledaju i na koje odgovaraju. Okvir takođe treba da osigura da korisnici budu svesniji podataka koji se prikupljaju o njima i načina na koji se oni koriste (uključujući razloge zašto je određeni korisnik targetiran za određeni oglas).

- **Države treba da razjasne gde se postojeća regulativa o medijima i sadržaju primenjuje na virtuelni sadržaj.** Tamo gde se identifikuju nedostaci, države treba da preispitaju i izrade politike i preporuke za moderaciju onlajn sadržaja u kontekstu nepristupačnosti (ili „stavljanja u crnu kutiju“) onlajn ekosistema i platformi.

- **Države treba da promovišu poslovnu praksu koja predstavlja alternativu trenutnom oglašavanju zasnovanom na praćenju.** Trenutne poslovne prakse internet posrednika stvaraju problematičnu koncentraciju moći koja negativno utiče na ličnu autonomiju i posredovanje korisnika.

- **Države treba da obezbede da se privatni akteri ponašaju u skladu sa Vodećim principima UN-a o poslovanju i ljudskim pravima** tako da korporativne vrednosti i strukture upravljanja ne daju prioritet maksimalnom uvećanju profita na štetu ljudskih prava i demokratskih vrednosti. Javnost sve više zahteva da preduzeća ne rade u komercijalnom vakuumu, već da odražavaju demokratske vrednosti i prioritete.

- **Države treba da podstaknu privatni sektor da traži nepravne puteve za promovisanje veće transparentnosti i odgovornosti.** Privatne inicijative o etičkim kodeksima igraju ključnu ulogu u korporativnoj društvenoj odgovornosti. Međutim, ako su izolovani, takvi samoregulatorni pristupi sami po sebi ne mogu da obezbede efektivnu zaštitu od mogućnosti da oglašavanje zasnovano na praćenju ugrozi apsolutno pravo na slobodu mišljenja.

Preporuke o promovisanju i podizanju svesti o reklamama zasnovanim na praćenju u opštoj javnosti

- **Države treba da promovišu svest i digitalnu pismenost** kako bi pojedinci znali kako da upravljaju sopstvenom medijskom potrošnjom i upotrebom internet posrednika. Korisnici bi trebalo da imaju smisleno razumevanje zašto im se prikazuje targetiran sadržaj, kao i kako se njihovi lični podaci obrađuju i na koji način im se pristupa.

Za korisnike je važno da razumeju ne samo količinu svojih ličnih podataka koja se obrađuje, već i vrstu informacija kojima se može pristupiti, ko im može pristupiti i način na koji se određene informacije mogu povezati sa zaštićenim karakteristikama. Povećana digitalna pismenost važna je za osnaživanje korisnika i jačanje otpornosti na industriju koja se stalno prilagođava.

- **Države treba da prepoznaju kako oglašavanje zasnovano na praćenju može da utiče na prava na jednakost i nediskriminaciju u kombinaciji sa pravom na slobodu mišljenja i izražavanja.** Oglašavanje zasnovano na praćenju stvara različita i potencijalno diskriminišuća iskustva, kako unutar tako i između grupa ljudi koji dele određene karakteristike. Ovo se može pogoršati jer određene grupe nisu digitalno pismene i to može da uveća negativna iskustva sa modelima zasnovanim na praćenju.

- **Države treba da podstaknu privatne aktere da razmotre koncept „društvene dozvole“,** koji nastoji da obezbedi da privatni i javni pružaoci usluga deluju odgovorno i etički u najboljem interesu zajednice.

- **Države treba da ulažu u istraživanje** kako bi razvile snažnu empirijsku osnovu koja može da obezbedi inicijative za promovisanje i podizanje svesti koje se bave praktičnim pitanjima o tome kako javnost reaguje na onlajn manipulacije i strategije targetiranog oglašavanja.

Preporuke za zakonski propisanu smislenu transparentnost: različiti slojevi transparentnosti

- **Države treba da obezbede smislenu transparentnost oglašavanja zasnovanog na praćenju.** Personalizacija oglašavanja zasnovanog na praćenju znači da različiti pojedinci vide različite oglase na osnovu brojnih faktora, uključujući doba dana, kontekst, demografiju, lične karakteristike i obrasce ponašanja. Ipak, algoritamski sistemi koji se popunjavaju korisničkim podacima su duboko netransparentni, često se opisuju kao „skriveni u crnoj kutiji“. Stoga su odluke koje stoje iza oglašavanja zasnovanog na praćenju skoro nerazumljive korisnicima (ili regulatorima). Kao posledica toga, korisnici nemaju nikakvo smisleno razumevanje zbog čega im se prikazuje određeni oglas u

određenom trenutku i kako se njihovi lični podaci dele i koriste u tom procesu.

- **Imenovana nadzorna tela sa ekspertizom u oblastima ravnopravnosti i nediskriminacije treba da budu ovlašćena da prate i rešavaju nejednake/neravnopravne ili diskriminatorne posledice koje oglašavanje zasnovano na praćenju ima na marginalizovane grupe.** Države treba da razmotre različite pristupe odgovornosti za štetno oglašavanje zasnovano na praćenju. Države treba da razmotre prednosti modela samoregulacije i koregulacije, korporativne odgovornosti i upravljanja, mehanizama parnica ili alternativnih e-sudova u uspostavljanju odgovornosti za štetne posledice.

- **Države treba da obezbede da tela za ravnopravnost imaju ovlašćenja da preduzmu strateške parnice** kako bi osporila diskriminatorne ishode automatizovanih mera.

- **Države bi trebalo da sarađuju sa akademskom zajednicom, civilnim društvom i nezavisnim zainteresovanim stranama** u cilju preusmeravanja napora ka ostvarivanju većeg pristupa razvrstanim podacima velikih razmera koji mogu da omoguće istraživanje i razumevanje profilisanja i oglašavanja zasnovanog na podacima. Transparentnost je neophodna da bi države i javnost znali kako se primenjuje oglašavanje zasnovano na praćenju. To će da omogući smislena istraživanja i preispitivanje problematičnih procesa. Pristup podacima za smislena istraživanja javnog interesa treba da se zasniva na zakonskom okviru.

Preporuke koje se bave interakcijom između privatnosti pojedinca i privatnosti grupe

- Dok međunarodno pravo o ljudskim pravima definiše individualna prava, onlajn profilisanje ima kolektivne aspekte i uticaje. Digitalni profili zasnovani su na zaključcima i pretpostavkama o složenoj mreži podataka i mreža. Algoritamsko profilisanje može da poveže karakteristike i veze da bi profilisalo pojedince iz marginalizovanih grupa. **Države stoga treba da obezbede da internet posrednici poštuju pravo na slobodu mišljenja i da budu svesni njegove važne veze sa pravom na slobodu udruživanja i izražavanja.**

- **Države treba da razmotre ograničenja postojećih zakonskih mehanizama za sprovođenje kolektivnih prava u kontekstu oglašavanja zasnovanog na praćenju.** Sadašnji pravni sistemi odražavaju individualna prava, a jedino proširenje na pitanja grupa postoji kada pojedinac pripada određenoj grupi. U ovim slučajevima, čak i kada pojedinac donese utemeljenu odluku da odustane od deljenja svojih ličnih podataka, i dalje može da bude profilisan kao deo šire grupe koja je targetirana ili kategorisana VI sistemima. Države treba da obezbede zaštitu i regulisanje korišćenja ličnih podataka, uključujući i metapodatke ili demografski identifikacione podatke, koji se smatraju izuzetno relevantnim i vrednim kada su u pitanju metode oglašavanja.

2.2 Preporuke za razvoj regulatornih i ko-regulatornih rešenja koja mogu delotvorno da reše negativan uticaj na ljudska prava koji proizilazi iz oglašavanja zasnovanog na praćenju

Preporuke za očuvanje apsolutnog prava slobode mišljenja

- **U skladu sa međunarodnim standardima ljudskih prava, države treba da poštuju i promovišu apsolutno pravo na slobodu mišljenja** – što uključuje pravo da svako čuva privatnost svojih misli i mišljenja, i da se ne manipuliše ničijim mislima i mišljenjima, i da ne budu kažnjeni za svoje misli i mišljenja.

- **Države treba da naglase da svako uživa pravo da ima svoje mišljenje bez ičijeg uplitanja; pravo na traženje, primanje i deljenje informacija i ideja putem bilo kog medija, bez obzira na fizičke granice; i pravo da ne bude podvrgnut nezakonitom ili proizvoljnom uplitanju u njegovu/njenu privatnost.** Države treba da obezbede da pojedinci mogu da formiraju svoje mišljenje, a da su istovremeno zaštićeni od manipulacije netransparentnim metodama profilisanja koje određuju kada je korisnik najpodložniji bihevioralnom uticaju kako bi se iskoristile ranjivosti korisnika. Neopravdani uticaj može da proistekne iz praksi kao što su: netransparentno ili neproverljivo

targetirano oglašavanje u velikom obimu; tehnike praćenja i posmatranja ponašanja ili prikrivanje karakteristika dizajna („tamni obrasci"); ili korišćenje neravnoteže moći da se utiče na misli (brzina, razmera, nepristupačnost, „crna kutija" i sistematski netransparentni uticaj). Takve tehnike praćenja i targetiranja mogu da dovedu do autocenzure i konformističkih efekata, koji bi mogli da budu rasprostranjeniji među određenim segmentima stanovništva. Države treba da uspostave jasne politike i kriterijume za granicu između legitimnog uticaja i nelegitimne manipulacije zasnovane na algoritamskim tehnologijama, za šta bi države trebalo da razmotre moratorijume ili zabrane.

- **Države bi trebalo da razmotre pravnu mogućnost za korisnike,** da se pozabave kaznama koje nameću internet posrednici zbog nedostataka u sistemu ili putem klasifikacije korisnika, što utiče na onlajn iskustvo pojedinca bez obzira na tačnost klasifikacije.

- **Države bi trebalo zakonski da obezbede anonimnost i enkripciju,** a to bi trebalo da uključi i osiguranje da se mišljenja ne otkrivaju nehotice.

- **Države bi trebalo da ulažu u digitalnu i medijsku pismenost i obrazovne kampanje kako bi informisale javnost o tome kako targetirano oglašavanje zasnovano na profilisanju i metodama praćenja utiče na onlajn iskustvo pojedinaca i kako oglašavanje zasnovano na praćenju ugrožava slobodu mišljenja.** Stalno praćenje nije u skladu sa ljudskim pravima i stvara se rizik od kreiranja zastrašujućih posledica na slobodu mišljenja i izražavanja. Države treba da obezbede da pojedinci imaju dovoljno alata i raznovrsnosti informacija da slobodno formiraju svoje mišljenje i da uživaju u pozitivnim aspektima slobode mišljenja.

- **Države treba da se pozabave oglašavanjem zasnovanim na praćenju u sociotehničkom kontekstu moderacije sadržaja i prireĎivanja i deljenja sadržaja,** i treba da identifikuju načine za rešavanje centralizacije moći, uključujući i zbog povezivanja više usluga.

Preporuke za smislenu transparentnost i regulatorne mere onlajn targetiranja

- **Države treba da razviju politiku ljudskih prava sa naglaskom na bitna pitanja ljudskih prava –** kao što su sloboda izražavanja, sloboda medija, privatnost i sloboda od diskriminacije. Države su nosioci obaveza prema međunarodnom pravu o ljudskim pravima i imaju pozitivnu obavezu da štite ljudska prava od uplitanja drugih, uključujući i od strane privatnih aktera ili pojedinaca. Države bi stoga trebalo da se obavežu da će poštovati međunarodne zakone o ljudskim pravima i treba da obezbede da nacionalni zakoni i politike koje regulišu internet posrednike i reklamnu industriju budu u potpunosti u skladu sa međunarodnim okvirom ljudskih prava.

- **Države treba da obezbede da privatni akteri deluju u skladu sa zakonitim postupanjem i standardima zakonitosti, legitimnosti i prihvatanja nadzora od strane nezavisnog i nepristrasnog pravosudnog tela, u skladu sa Vodećim principima UN-a o poslovanju i ljudskim pravima.** Države treba da uspostave regulatorni okvir za kompanije koje će pokazati da su u potpunosti realizovale svoje odgovornosti u skladu sa Vodećim principima.

- **Države treba da efikasno sprovode postojeće zakone o zaštiti podataka i privatnosti.** U ovom kontekstu, države treba da obezbede principe kao što su minimizacija podataka i ograničenje svrhe. Države takođe treba da efikasno sprovode zakone o konkurenciji i antimonopolsko pravo, kao i druge propise koji imaju za cilj jačanje ljudske autonomije.

- **Države bi trebalo da uslovljavaju korporativno praćenje – uključujući targetirano oglašavanje koje koristi praćenje i profilisanje – poštovanjem ljudskih prava** i evidencijama usklađenosti sa Vodećim principima UN-a o poslovanju i ljudskim pravima.

- **Države treba da obavežu internet posrednike da obezbede dokumentaciju o metodama praćenja i profilisanja zasnovanim na veštačkoj inteligenciji koje primenjuju u reklamne svrhe.** Države bi trebalo da zahtevaju od internet posrednika da daju objašnjenja u vezi sa modelima koji se koriste, koji podaci se prikupljaju i u koju svrhu – kao i metrike učinka i rezultate testiranja za korišćene modele. Države

treba da nalože da internet posrednici moraju pravilno da objasne kako funkcionišu njihovi modeli oglašavanja i poslovanja, kako je uključeno algoritamsko donošenje odluka i kako takvi automatizovani sistemi donose odluke koje utiču na korisnika. Svako otkrivanje treba da bude na način koji je razumljiv i pristupačan korisnicima. Treba da se obuhvate i informacije o prikupljanju ličnih zaštićenih karakteristika, ili njihovih posrednika. Dodatne informacije treba deliti sa istraživačima i regulatorima, na način koji štiti privatnost.

- **Za bilo koje poslovne modele zasnovane na prikupljanju podataka i oglašavanju, države bi trebalo da zahtevaju ex ante procene uticaja na ljudska prava** koje su deo jasnog regulatornog okvira, i koje su transparentne, nezavisne i inkluzivne (uključujući smislene konsultacije sa potencijalno pogođenim grupama i drugim zainteresovanim stranama). Proces treba da uključi nadzor koji vrši regulatorna agencija ili nezavisne zainteresovane strane sa relevantnom stručnošću, kako bi se osiguralo ublažavanje negativnih uticaja modela oglašavanja na sprečavanje diskriminacije i očuvanje slobode mišljenja i izražavanja.

- **Države treba da usvoje nova ograničenja ili da primene postojeća, koja ograničavaju koje vrste podataka mogu da se prikupljaju i kako mogu da se koriste, i koje vrste podataka mogu da budu otkrivene oglašivačima, brokerima podataka ili trećim licima.**

- **Države treba da jasno definišu način na koji metode oglašavanja nanose „štetu"** (pojedinačno kao i kolektivno/demokratskim procesima), na osnovu principa predostrožnosti, kako bi mogle da identifikuju prag za zabranu štetnih praksi oglašavanja zasnovanih na praćenju. Takve zabrane treba da obuhvataju, na primer, zabranu naoružavanja sofisticiranim tehnikama za uticaj na osnovu psiholoških modela koji pretpostavljaju psihološku ranjivost i mogućnost manipulisanja. Za prikupljanje podataka (data harvesting) za targetirano oglašavanje koje je unutar praga, države bi trebalo da obezbede stroge metode transparentnosti – na primer u vezi sa plasmanom proizvoda – i poštovanje ljudskih prava, što stavlja najbolji interes pojedinca u centar.

- **Države treba da zabrane neselektivno masovno prikupljanje i analizu korisničkih podataka za targetirano oglašavanje koje šteti korisnicima pojedinačno ili kolektivno, ili ometa njihovo pravo na slobodu mišljenja.** To uključuje, na primer, targetirano oglašavanje zasnovano na sveobuhvatnom praćenju ranjivosti korisnika ili kategorija zaštićenih karakteristika, kao što su etnička pripadnost, pol, verska uverenja ili seksualna orijentacija. Zabrane i ograničenja reklamiranja zasnovanog na praćenju mogu da primene model zabrane lažnog i podsvesnog oglašavanja ili ograničenja reklamiranja alkohola, duvana, kockanja ili materijala opasnih po životnu sredinu. Posebnu zaštitu treba razmotriti za ranjive/osetljive grupe, kao što su deca i mladi ljudi.

- **Države treba da obezbede da personalizovano oglašavanje korišćenjem prikupljanja ličnih podataka funkcioniše na osnovu informisane saglasnosti i na osnovu prihvatanja (opt-in).** Države treba da obezbede korisnicima mogućnost da biraju koji podaci se prikupljaju u koju svrhu i kako žele da se uključe u onlajn debatu i da budu targetirani oglašavanjem (uključujući gledanje personalizovanog oglašavanja i praćenje radi oglašavanja na prvom mestu). Za manje invazivne modele oglašavanja, trebalo bi da bude obezbeđena barem opcija za odustajanje od prikupljanja podataka, a trebalo bi da postoji i alternativni način da se obezbedi bezbednost korisnika onlajn. Saglasnost mora biti eksplicitna, bez prinude i zasnovana na utemeljenom izboru, u skladu sa zakonima o zaštiti podataka, uz priznavanje da modeli oglašavanja mogu da utiču na ljudska prava ne samo prikupljanjem i analizom ličnih podataka, već i korišćenjem drugih informacija i metapodataka. Korisnici treba da imaju kontrolu nad tim koji se podaci prikupljaju, zadržavaju ili zaključuju i kako se koriste za oglašavanje. Države treba da promovišu privatnost namerno i automatski.

- **Države treba da nalože internet posrednicima da daju informacije o svom modelu prihoda i da obezbede mrežu za transparentnost.**

- **Države treba da obavežu internet posrednike da obaveste korisnike kada su podvrgnuti bilo kom obliku praćenja i profilisanja,** da kažu korisnicima kako takvi mehanizmi funkcionišu i da obezbede opcije za prihvatanje ili odustanak na jednostavan i lak način. Države treba

da nalože da posrednici moraju da otkriju da li se reklamni sadržaj prikazuje na osnovu istorije datog korisnika, informacija o lokaciji, aktivnosti na društvenim mrežama, demografskih karakteristika ili drugih informacija (uključujući zastupnike i „sličnu publiku" koja grupiše korisnike sa određenim karakteristikama). Od posrednika bi se takođe trebalo zahtevati da otkriju svoje parametre ciljanja i kategorije publike (na osnovu ponašanja kao i sadržaja), kao i smernice na osnovu kojih se kategorije publike procenjuju i da li algoritamski generisane kategorije pre upotrebe pregledaju osobe zadužene za to.

- **Države treba da obezbede da korisnici imaju pristup podacima o profilisanju koje internet posrednici imaju o njima,** kao i svim zaključcima o njima (uključujući metapodatke, kao što su dodeljene kategorije i lista oglašivača koji pokušavaju da utiču na njih). Ovi podaci treba da budu dostupni korisnicima na zahtev, u razumljivom i pristupačnom formatu. Korisnici bi trebalo da budu u mogućnosti da isprave i izbrišu svoj profil.

- Uvođenjem određenih zabrana i propisujući transparentnost o tome koji se podaci prikupljaju, čuvaju i analiziraju i za kakve se reklamne odluke ti podaci koriste, **države treba da se pozabave netransparentnim oglašavanjem zasnovanim na praćenju koje može da utiče na sposobnost pojedinaca da koriste usluge internet posrednika kao forume za slobodno izražavanje, pristup informacijama i angažovanje u javnom životu.**

- **Države treba da zahtevaju redovno izveštavanje o transparentnosti,** propisujući minimalne zahteve u vezi sa prikupljenim podacima, korišćenim kategorijama i uključenom automatizacijom – i kako oni utiču na sadržaj i oglašavanje. Od posrednika bi takođe trebalo da se zahteva da obezbede obavezne, funkcionalne biblioteke za oglašavanje.

- **Države treba da uspostave okvir za internet posrednike, da obelodane svoje procene uticaja na ljudska prava i obezbede spoljnu nezavisnu procenu.** Ovo bi trebalo da uključuje procene o tome kako su rizici slobode izražavanja i informisanja povezani sa njihovim politikama i praksama targetiranog oglašavanja, kao i procene rizika od diskriminacije.

- Da bi se osigurale nezavisne eksterne revizije modela oglašavanja, **države treba da zahtevaju od internet posrednika da sprovode izveštavanje u skladu sa privatnošću i zaštitom podataka i da je ono dostupno svim relevantnim javnim organima i nezavisnim zainteresovanim stranama,** uključujući i istraživače i organizacije civilnog društva.

- **Države treba da zahtevaju da internet posrednici omogućavaju istraživačima i organizacijama civilnog društva pristup njihovim podacima o oglašavanju,** kako bi mogli da procene prakse oglašavanja i njihov individualni i kolektivni uticaj i da informišu o istraživanju zasnovanom na javnom interesu.

- **Države treba da obezbede demokratsko upravljanje** i da prepoznaju i ovlaste određena nadzorna tela stručna u oblastima ravnopravnosti i nediskriminacije da prate i rešavaju neravnopravne ili diskriminatorne efekte oglašavanja zasnovanog na praćenju marginalizovanih grupa.

- **Države treba da ojačaju nezavisnost tela za zaštitu podataka** i da im obezbede dovoljnu političku podršku i finansijska sredstva i nadležnosti.

- **Države treba da podstiču koordinaciju više zainteresovanih strana,** izgradnju kapaciteta i istraživanje uticaja dizajna interfejsa na ponašanje korisnika, kao i pitanja kao što su „tamni obrasci". Države takođe treba da promovišu istraživanje marginalizacije i rodne prirode digitalnog praćenja i efekata oglašavanja – i istraživanje negativnih eksternih efekata poslovnih modela zasnovanih na prikupljanju ličnih informacija na planetarnom nivou, što omogućava mikro-targetiranje pojedinaca prilagođeno njihovim specifičnim karakteristikama, osobinama i sklonostima. Pored toga, istraživanje treba da ispita: potencijal targetiranog oglašavanja da zlonamerno utiče na ponašanje; povezanost između poslovnih modela zasnovanih na praćenju i podsticaja posrednika da daju prioritet štetnom sadržaju, omogućavajući štetu od diskriminacije koja se realizuje putem algoritamskog odlučivanja; i povezane efekte zastrašivanja.

- **Države treba da se uzdrže od proizvoljnog pristupa podacima koje prikupljaju internet posrednici.** Zahtevi za podatke treba

da budu zasnovani na legitimnosti, zakonitosti i neophodnosti, i proporcionalnosti, i trebalo bi da obezbede sudski nadzor. Države treba adekvatno da sprovode zaštitne mere za zabranu obaveznog prenosa podataka, posebno organima za sprovođenje zakona, i da preduzmu posebne mere za zaštitu marginalizovanih i ranjivih grupa.

- **Države treba da se pozabave koncentracijom moći,** koja uključuje prikupljanje podataka kao izvor tržišne moći i dodatno jačanje dominacije nekoliko dominantnih posrednika na štetu potencijalnih konkurenata i informativnih medija. Mere bi mogle da obuhvataju, na primer, obavezu interoperabilnosti, prenosivost podataka (vlasništvo podataka) kroz bezbedne mehanizme i/ili decentralizaciju moći.

- **Države treba da se pozabave uticajem koji koncentracija tržišta digitalnog oglašavanja ima na konvencionalne medije i dostupnost informacija od javnog interesa.** Države treba da ulažu u jake javne servise i nezavisno novinarstvo.

- **Države treba da ulažu u istraživanje alternativnih tokova prihoda koji se ne oslanjaju na komodifikovanje privatnog ponašanja ljudi i ne utiču ili ne oblikuju novo ponašanje.** Primeri takvih alternativa uključuju kontekstualno oglašavanje ili targetiranje u skladu sa jednostavnim kriterijumima za koje se korisnik opredeli, direktne veze između pružalaca sadržaja i oglašivača bez da posrednički sektor oglašivača monetizuje sadržaj koji uređuju drugi (uključujući medije), i napore da se podstaknu inovacije koje se odnose na ljudska prava.

- **Države treba da razmotre platforme javnih usluga** koje služe i koje su u potpunosti odgovorne javnosti, na osnovu demokratskog upravljanja.

2.3 Opšti principi za sprečavanje država da koriste poslovne modele zasnovane na praćenju

Države takođe mogu da zloupotrebe poslovne modele prikupljanja podataka koji su povezani sa oglašavanjem zasnovanim na praćenju. Javne organi se sve više oslanjaju na izvlačenje podataka (data mining) koje rade privatne kompanije, koje služe kao „rezervoari potrošačkih podataka". Vlade redovno

imaju pristup raznim podacima koje pruža privatni sektor. Poslednjih godina bilo je nekoliko slučajeva prijavljenih od strane grupa za građanska prava koji su dokumentovali kako su javne i organi sklopili sporazum sa brokerima podataka kako bi dobili pristup ličnim podacima korisnika. Na primer, Electronic Frontiers Foundation (EFF) opisala je slučaj kada je imigraciona i carinska služba SAD kupovala podatke ALPR-a[81] od Vigilant-a, koji su joj pomogli u lociranju ljudi koje agencija namerava da deportuje.[82] Ovakvi neformalni sporazumi između vlada i privatnih subjekata predstavljaju ozbiljnu pretnju za zaštitu ljudskih prava. Nepotrebno i neproporcionalno praćenje može da ugrozi onlajn bezbednost i da omete pristup informacijama i idejama.[83] Praćenje može da ima zastrašujući efekat na onlajn izražavanje pojedinaca, posebno novinara i članova civilnog društva, koji mogu da se samocenzurišu zbog straha da će biti stalno praćeni. Osim toga, praćenje ima neproporcionalan uticaj na slobodu izražavanja marginalizovanih grupa, uključujući rasne, verske, etničke, rodne i seksualne manjine, kao i novinare ili branitelje ljudskih prava.[84] Ovo podjednako važi za državno praćenje kao i za korporativno praćenje.

Konkretno, napredni razvoj VI kreirao je nove mogućnosti za masovno (en mass) državno praćenje koje se oslanja na arhitekturu poslovnih modela posrednika. Različiti oblici alata za praćenje sadržaja koje koriste države da razjasne odnose između targetiranih korisnika, ili da daju značenje ili stav njihovim objavama na društvenim mrežama putem obrade prirodnog jezika (NLP-a) i analize sentimenata, mogu da imaju ozbiljne posledice po zaštitu ljudskih prava onlajn. Kada je ovaj proces osnažen mašinskim učenjem, države mogu da otkriju povezanosti i međusobne veze koje su potencijalno nevidljive ljudskom oku. Posebno u autoritarnim režimima, zaštitnici ljudskih prava, politički aktivisti i marginalizovane grupe mogu da budu proganjani zbog svojih mišljenja i stavova, što dovodi do nesrazmernih i strogih kazni.

81 https://www.techdirt.com/articles/20190321/09165441842/vigilant-customers-are-lying-about-ices-access-to-plate-records.shtml.
82 https://www.aclunc.org/blog/documents-reveal-ice-using-driver-location-data-local-police-deportations.
83 A/HRC/23/40, https://undocs.org/en/A/HRC/23/40.
84 A/HRC/29/32, https://undocs.org/en/A/HRC/29/32.

Ovaj deo sadrži opšte principe koje države treba da poštuju, kako bi sprečile kršenje ljudskih prava u velikim razmerama:

1. **Organi javne vlasti, a posebno agencije za sprovođenje zakona, treba da imaju veoma ograničen i posebno targetiran pristup podacima, usmeren na specifične identifikatore ili specifične kategorije.**

2. **Prikupljanje podataka od strane organa za sprovođenje zakona uvek treba da se zasniva na konkretnim sumnjama.** Organi za sprovođenje zakona treba da dobiju pristup samo određenim evidencijama i sadržajima. Ne treba da se vrši masovno praćenje, niti prepoznavanje lica koje može da omogući masovni nadzor.

3. **Podaci prikupljeni korišćenjem posebnih ovlašćenja za nacionalnu bezbednost ne bi trebalo da se koriste u bilo koje druge javne svrhe, pa ni za sprovođenje zakona.** Trebalo bi da se zadrže ograničeni period i zatim obrišu kada više nisu potrebni.

4. **Metapodaci koji otkrivaju informacije, kao što su sa kim ljudi komuniciraju, gde i kada, mogu da budu podaci koji izuzetno otkrivaju stvari o životima pojedinaca i stoga bi trebalo da dobiju visok nivo pravne zaštite.**

5. **Nezakonito praćenje treba da bude kriminalizovano, uz delotvorne pravne lekove.** Nezakonito prikupljeni podaci treba da budu neprihvatljivi kao dokaz, dok uzbunjivači treba da budu zaštićeni kada otkriju nezakonito ponašanje.

3. Zaključak

Ovaj deo izveštaja ističe uticaj poslovnih modela targetiranog oglašavanja omogućenog veštačkom inteligencijom i oglašavanja zasnovanog na praćenju i prikupljanju podataka (data harvesting) na priređivanje i deljenje sadržaja, pluralizam informacija i sposobnost pojedinaca da formiraju, zadrže i izražavaju svoja mišljenja i imaju slobodan pristup informacijama.

Izveštaj analizira vezu targetiranog oglašavanja sa usponom moćnih internet posrednika, koji istovremeno deluju kao čuvari izražavanja i informisanja na digitalnom tržištu ideja. Takođe ispituje kako vrednost, a time i istaknutost onlajn sadržaja, sve više postaje zavisna od njegovog doprinosa ostvarivanju profita posrednika od reklama. Ilustrujući kako se podaci, karakteristike i ranjivosti pojedinaca iskorišćavaju za targetirano oglašavanje, izveštaj predočava uticaj razmatranja profita na upravljanje sadržajem i onlajn informacione prostore. Izveštaj istražuje kako trenutni digitalni ekosistem može da utiče na apsolutno pravo na slobodu mišljenja i pravo na traženje, primanje i prenošenje informacija svih vrsta, bez obzira na fizičke granice.

Osim toga, izveštaj naglašava vezu poslovnih modela internet posrednika za prikupljanje podataka sa državnim praćenjem. Praćenje ima zastrašujući efekat na onlajn izražavanje pojedinaca, a posebno novinara i civilnog društva, sa neproporcionalnim uticajem na marginalizovane pojedince i grupe. Ovo se isto tako odnosi i na državno, kao i na korporativno praćenje. Izveštaj predočava skup proaktivnih, preventivnih i odgovarajućih preporuka za države članice OEBS-a. Ove preporuke usredsređene na ljudska prava fokusiraju se na očuvanje apsolutne slobode mišljenja, na obezbeđivanje značajne transparentnosti, na obezbeđivanje regulatornih mera onlajn targetiranja i na opšte principe za sprečavanje država da se oslanjaju na poslovne modele zasnovane na praćenju. Iako je potrebno hitno rešiti određene izazove, kao što je nedostatak mogućnosti za dobijanje objašnjenje, transparentnosti i odgovornosti sistema zasnovanih na veštačkoj inteligenciji koji su povezani sa oglašavanjem i upravljanjem sadržajem, izveštaj takođe identifikuje potrebu za rešavanjem pitanja šireg ekosistema zasnovanog na praćenju kako bi se istinski štitila i promovisala sloboda mišljenja i izražavanja u digitalnom dobu.

Ova publikacija je realizovana zahvaljujući finansijskim doprinosima Austrije, Bugarske, Češke, Finske, Francuske, Holandije, Švedske, Švajcarske i Sjedinjenih Američkih Država.

Prevod na srpski jezik publikacije objavljene u januaru 2022. godine kao „Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual“ pomogle su Američka agencija za međunarodni razvoj (USAID) u okviru projekta Podrška reformi medija u Srbiji, i Misija OEBS-a u Srbiji.

OSCE
The Representative on
Freedom of the Media