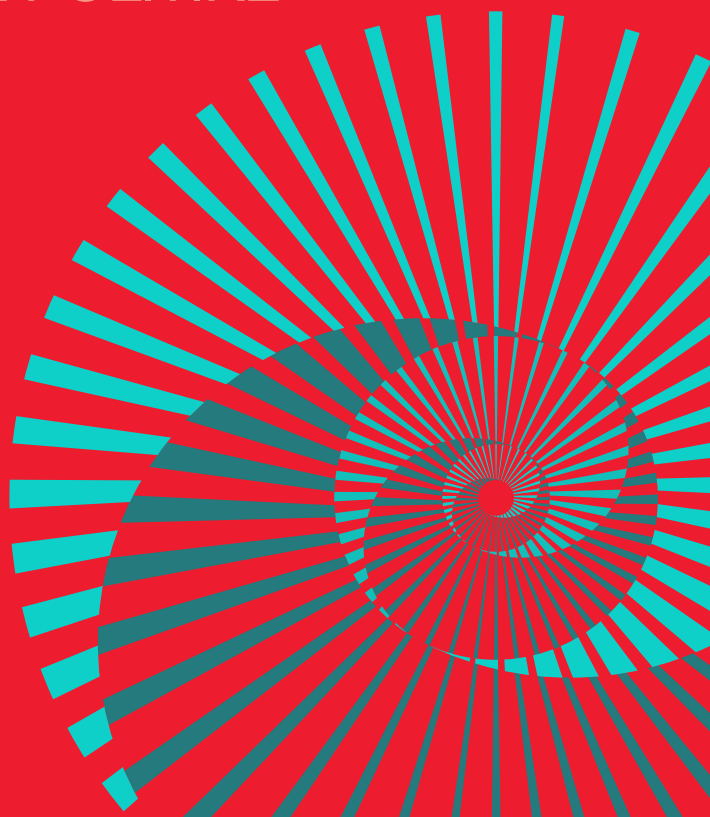


Razmislite ponovo: Sloboda misli u doba umjetne inteligencije

SAŽETAK JAVNE POLITIKE



osce
ODIHR



Razmislite ponovo: Sloboda misli u doba umjetne inteligencije

SAŽETAK JAVNE POLITIKE



Priznanja i zahvale

ODIHR izražava iskrenu zahvalnost onima koji su velikodušno podijeli-li svoje znanje i stručnost u pripremi ovog dokumenta. Njihove pažljivo razmotrene recenzije i povratne informacije u raznim fazama procesa od neprocjenjivog su značaja. ODIHR zahvaljuje članovima ODIHR-ovog panela stručnjaka za slobodu vjeroispovijesti ili uvjerenja, s posebnom zahvalnošću Ahmedu Shaheedu, Davidu Griffithsu i Pasqualeu Annicchinu na njihovom doprinosu s dubokim uvidom. Zahvala ide i sljedećim osobama na recenziji različitih nacрта i važnim doprinosima: Cameranu Ashrafu, Christophu Bublitzu, Hervéu Chneiweissu, Emily Elstub, Alexanderu Kriebitzu i Simonu McCarthy-Jonesu, kao i kolegama iz Ureda predstavnika za slobodu medija. Posebna zahvala ide Benjaminu Greenacreu, čije su rasprave o ovoj temi uveliko obogatile konačni ishod, te Mileni Costas Trascasas na kontinuiranom dijalogu i duhu saradnje.

Razmislite ponovo: Sloboda misli u doba umjetne inteligencije

Objavio Ured OSCE-a za demokratske institucije i ljudska prava (ODIHR)
Miodowa 10
00-251 Warsaw
Poland
www.osce.org/odihr

© OSCE/ODIHR 2026., sva prava pridržana.

Sadržaj ove publikacije može se slobodno koristiti i kopirati u obrazovne i druge nekomercijalne svrhe, pod uvjetom da je u svakoj od tih reprodukcija OSCE/ODIHR naveden kao izvor.

ISBN: 978-92-9271-514-4

Dizajn: Filip Andronik



Sadržaj

Sažetak	5
Uvod	7
1. O slobodi misli	9
2. O umjetnoj inteligenciji	14
Generativna UI	15
Korporativni utjecaj u doba UI	17
3. Ljudskopravaške implikacije razmišljanja unutar algoritma	20
UI, pristranost i misao	20
Ljudski nadzor, transparentnost i odgovornost	23
4. Sloboda misli u rekonfiguriranom građanskom prostoru	26
Nova arhitektura građanskog prostora koja utječe na slobodu misli	26
„Informacijski poremećaj“ mijenja istinski izbor	31
Uticaj UI na ljudsku povezanost, razvoj misli i pravo na zaborav	33
5. Neurotehnologija i sloboda misli	36
6. Regulatorni okviri	42
7. Zaključci i preporuke	48

Sažetak

Ovaj sažetak javne politike pruža jezgrovit uvod u pravo na slobodu misli, napominjući da države imaju kontinuiranu dužnost stvarati poticajno okruženje za njegovo potpuno ostvarenje.¹ Počinje definiranjem slobode misli kao temeljnog prava koje se mora štititi i promovirati u svim kontekstima. Zatim istražuje kako nove tehnologije, posebno umjetna inteligencija (UI) i neurotehnologije, mijenjaju mentalni pejzaž. Iako se takve tehnologije, ovisno o načinu na koji se razvijaju i primjenjuju, mogu koristiti za poboljšanje uživanja ljudskih prava², na primjer, boljim pristupom raznovrsnim informacijama, prilagođenim kako bi pomogle obrazovnim procesima ili podržale medicinski napredak, također nose i neviđene rizike. One zahtijevaju sveobuhvatne procjene i robusne zaštitne mjere prije razvoja, uvođenja ili upotrebe.

U ovom sažetku se ispituju implikacije novog informacijskog ekosistema po slobodu misli, kao i potencijalni utjecaji neurotehnologija zasnovanih na umjetnoj inteligenciji, koji pokreću temeljna pitanja o nepovredivosti individualne mentalne autonomije. U sažetku javne politike se ističu i preporuke koje države trebaju uzeti u obzir pri poštivanju slobode misli u digitalnom dobu. One uključuju:

- Regulatorne i institucionalne zaštitne mjere protiv novih rizika, posebno onih koje stvaraju umjetna inteligencija i neurotehnologije
- Mehanizme etičkog nadzora i obavezne procjene utjecaja na ljudska prava ugrađene u cijeli životni ciklus umjetne inteligencije³
- Inicijative za osiguranje transparentnosti, inkluzivne javne rasprave i javnu odgovornost u procesima donošenja odluka

1 Iako prepoznamo značajne implikacije po slobodu misli kod primjene umjetne inteligencije u vojsci ili u kontekstu nadzora, ta područja nisu obuhvaćena ovim sažetkom.

2 Theresa Addie, MS., *Harnessing Technology to Safeguard Human Rights: AI, Big Data, and Accountability*, Centar za istraživanje ljudskih prava, 8. 4. 2025.; Specijalni izvjestilac UN-a (UNSR) o slobodi vjere ili uvjerenja (FoRB) Ahmed Shaheed, *Report of the Special Rapporteur on freedom of religion or belief*, A/76/380, 5. 10. 2021., stavovi 2.- 4.

3 *UN Global Digital Compact*, kao aneks Rezolucije A/RES/79/1 Generalne skupštine UN-a, o Paktu za budućnost, usvojen 22. 9. 2024., uključuje sljedeće elemente u sklopu životnog ciklusa digitalnih tehnologija i tehnologija u nastajanju: preliminarni dizajn, dizajn, razvoj, evaluaciju, testiranje, isporuku, upotrebu, prodaju, nabavku, rad i povlačenje iz upotrebe.

- Izgradnju kapaciteta kroz obrazovanje i edukacije o etici i ljudskim pravima za razvojne programere i operatere tehnologija umjetne inteligencije
- Ojačanu ulogu neovisnih institucija u zaštiti slobode misli i demokratskih vrijednosti

U sažetku se iznosi zaključak da inovacija kojom od samog početka nisu uzeta u obzir i zaštićena ljudska prava potkopava ciljeve koji bi trebali biti vodilja etičkom, naučnom i tehnološkom napretku u korist čovječanstva.⁴

Temeljem rada međunarodnih organizacija i akademske zajednice na području umjetne inteligencije, neurotehnologija i ljudskih prava, u ovom sažetku javne politike ističu se specifične implikacije po slobodu misli. Svrha mu je da služi kao temelj za daljnju raspravu i materijal za inicijative i aktivnosti ODIHR-a, država članice OSCE-a i drugih aktera.

⁴ Deklaracija o korištenju naučnog i tehnološkog napretka u interesu mira i za dobrobit čovječanstva, Rezolucija 3384 (XXX) Generalne skupštine UN-a, usvojena 10.11.1975.

Uvod

Dok se vlade snalaze u brzo razvijajućem području umjetne inteligencije (UI), uključujući njenu integraciju u područja poput neurotehnologije, ključno a često zanemareno pravo zahtijeva hitnu pažnju, a to je sloboda misli. Utvrđena u Univerzalnoj deklaraciji o ljudskim pravima (UDHR), uz pravo na slobodu savjesti i vjeroispovijesti, sloboda misli štiti sposobnost pojedinca da samostalno razmišlja, formira vlastita uvjerenja, donosi autonomne odluke i doprinosi novim perspektivama. Ovo pravo potiče kreativnost, inovacije i razvoj otpornih, pluralističkih društava.⁵

Sloboda misli obuhvata i slobodu da se ne otkrivaju vlastite misli, kao i slobodu od prisile, kažnjavanja zbog vlastitih misli i nedopuštenog mijenjanja misli.⁶ Prepoznajući temeljnu prirodu ovog prava, države članice OSCE-a obavezale su se štititi slobodu misli, počevši od Završnog akta iz Helsinkija.⁷

5 **Međunarodni pakt o građanskim i političkim pravima**, Rez 2200A (XXI) Generalne skupštine UN-a, usvojen 16.12.1966., čl. 18.:

- “1. Svako ima pravo na slobodu misli, savjesti i vjeroispovijesti. To pravo uključuje slobodu da ima ili prihvati vjeroispovijest ili uvjerenje po svom izboru, te slobodu da, pojedinačno ili u zajednici s drugima, javno ili privatno, iskazuje svoju vjeroispovijest ili uvjerenje bogoslužjenjem, obredima, praksom i podučavanjem.
2. Niko ne smije biti podvrgnut prisili koja bi umanjila njegovu ili njenu slobodu da ima ili prihvati vjeru ili uvjerenje po svom izboru.
3. Sloboda iskazivanja vjere ili uvjerenja može biti podvrgnuta samo ograničenjima koja su propisana zakonom i koja su nužna za zaštitu javne sigurnosti, reda, zdravlja ili morala ili temeljnih prava i sloboda drugih.
4. Države strane ovog Pakta obavezuju se da će poštivati slobodu roditelja i, kada je to primjenjivo, zakonskih skrbnika da osiguraju vjerski i moralni odgoj svoje djece u skladu s vlastitim uvjerenjima.”

6 Ovo se odnosi na svaku modifikaciju misli učinjenu na način koji je suprotan standardima ljudskih prava, uključujući i promjenu i manipuliranje strukturom i funkcijom mozga. Vidi Ahmed Shaheed, **Izveštaj specijalnog izvjestioca**.

7 Završni akt iz Helsinkija s Prvog samita šefova država ili vlada Konferencije za sigurnost i saradnju u Evropi (CSCE), od 1.8.1975., potvrđuje kao jedno od svojih deset vodećih principa „poštivanje ljudskih prava i temeljnih sloboda, uključujući slobodu misli, savjesti, vjeroispovijesti ili uvjerenja”. To pravo je ponovo potvrđeno u više kasnijih dokumenata OSCE-a (npr. **Zaključni dokument s Trećeg sastanka**, Beč, 4.11.1986. do 19.1.1989.; **Dokument sastanka s Konferencije o ljudskoj dimenziji CSCE-a** u Kopenhagu, 29.6.1990.; **Odluka Ministarskog vijeća OSCE-a, br. 4/03**, “Tolerancija i nediskriminacija”, Maastricht, 2.12.2003.; **Odluka Ministarskog vijeća OSCE-a, br. 3/13** “Sloboda misli, savjesti, vjeroispovijesti ili uvjerenja”, Kijev, 9.12.2013.).

Ipak, s obzirom na brzi razvoj novih tehnologija, uključujući umjetnu inteligenciju i neurotehnologije, ostvarenje prava na slobodu misli je pod sve većim pritiskom. Kako ove tehnologije sve više oblikuju način na koji mislimo, komuniciramo i međusobno djelujemo, tako otvaraju složena pravna, etička i društvena pitanja koja zahtijevaju pažnju i djelovanje država članica.

1. O slobodi misli

Član 18. Univerzalne deklaracije o ljudskim pravima (UDHR) štiti pravo na „slobodu misli, savjesti i vjeroispovijesti”.⁸ Naglasak se obično stavlja na religiju, pri čemu se član 18. često sužava na „slobodu vjeroispovijesti ili uvjerenja“ ili „vjersku slobodu“. Donedavno se inherentna zaštita slobode misli smatrala toliko očitom da se definicija činila nepotrebnom, iako je ona temelj ljudskog dostojanstva i autonomije te svih ostalih ljudskih prava. Dok je René Cassin, koautor UDHR-a, opisao slobodu misli kao „izvor svih ostalih prava“, ni on ni drugi autori nisu to dalje razrađivali, a nakon toga je uslijedilo vrlo malo sudske prakse u oblasti slobode misli.⁹

Istovremeno, nakon Drugog svjetskog rata, autori Univerzalne deklaracije o ljudskim pravima bili su itekako svjesni da bi to pravo moglo biti ugroženo. Pokušaji utjecaja na misli ili promjene misli ili kažnjavanja stvarnih ili pretpostavljenih misli i mišljenja imaju dugu historiju.¹⁰ Ljudi koji odstupaju od mišljenja većine i osporavaju postojeće norme postaju mete mjera prisile, koje se obično intenziviraju kako raste nivo autokratije.¹¹

Uprkos nepostojanju definicije u međunarodnom pravu o tome šta tačno predstavlja „misao“,¹² ipak je moguće prepoznati njene obrise i attribute potrebne za rad na njenoj zaštiti u praksi.

Kao dio člana 18. Univerzalne deklaracije o ljudskim pravima i Međunarodnog pakta o građanskim i političkim pravima (ICCPR), ovo se pravo općenito shvata u kontekstu slobode savjesti i slobode vjeroispovijesti ili

⁸ Univerzalna deklaracija o ljudskim pravima, UNGA rezolucija 217 A, 10.12.1948.

⁹ Ahmed Shaheed, *Izveštaj specijalnog izvjestioca*.

¹⁰ Od nacističke propagande koja je dovela do Holokausta do totalitarizma i autoritarizma te drugih situacija u kojima su države poduzele mjere kako bi se omela sloboda misli, poput intervencija za sprječavanje nasilnog ekstremizma i radikalizacije koja vodi do terorizma. Vidi, npr. *'This is the Thought Police': The Prevent Duty and Its Chilling Effect on Human Rights*, (London: Amnesty International, 2023.); Ahmed Shaheed, *Report on the relationship between freedom of religion or belief and national security*, A/73/362, 5.9.2018., stav 43.

¹¹ 71% svjetske populacije živjelo je pod autokratskim režimima 2023. u uporedbi s 50% u 2003.; u 2024. godini 42. zemlje (u kojima živi 35% svjetske populacije) prolazile su kroz „autokratizaciju“, a 18 zemalja (u kojima živi 5% svjetske populacije) prolazilo je kroz demokratizaciju. Vidi Marina Nord, Martin Lundstedt i Staffan I. Lindberg, *Sloboda medija, demokracija i sigurnost, Izveštaj o istraživanju*, OSCE/RFoM, 15.7.2024., str. 6.

¹² Ahmed Shaheed, *Izveštaj specijalnog izvjestioca*, stav 11.

uvjerenja. Međutim, sloboda misli ne odnosi se samo na vjersku misao. Odbor UN-a za ljudska prava, koji je dao autoritativno tumačenje člana 18. ICCPR-a u Općem komentaru br. 22, objasnio je da on „obuhvata slobodu misli o svim pitanjima“.¹³

Važno je napomenuti da se od člana 18. ne smije odstupati, što znači da države ne mogu obustaviti svoje obaveze u vezi s ovim pravom, čak ni u uvjetima „vanredne situacije koja ugrožava život nacije“.¹⁴ Sloboda misli općenito se shvata kao apsolutno i bezuvjetno pravo. U Općem komentaru br. 22 napominje se da član 18. „ne dopušta nikakva ograničenja slobode misli i savjesti ili slobode vjere ili prihvatanja vjere ili uvjerenja po vlastitom izboru“.¹⁵ Ove dvije slobode, zajedno s pravom na vlastito mišljenje bez uplitanja (član 19.1. Međunarodnog pakta o građanskim i političkim pravima), čine ono što je u zakonima o ljudskim pravima poznato kao forum internum ili „unutrašnje svetište“ ljudskog uma.¹⁶

Forum internum se razlikuje od pojma forum externum, koji se odnosi na izražavanje misli, uvjerenja ili vjerovanja. Prava forum externum nisu apsolutna.¹⁷ Ona mogu podlijeagati ograničenjima od strane države „kako je propisano zakonom i kada je to potrebno za zaštitu javne sigurnosti, reda, zdravlja ili morala ili osnovnih prava i sloboda drugih“ (a u slučaju člana 19.3, iz dodatnih razloga zaštite nacionalne sigurnosti).

Ova formulacija pravi razliku između aspekata članova 18. i 19. koji imaju apsolutnu zaštitu i onih koji se mogu ograničiti u određenim okolnostima. Međutim, posmatranje slobode misli isključivo u smislu forum internuma je rizično jer ona

13 Opći komentar br. 22, „Član 18. (Sloboda misli, savjesti ili vjere)“, Odbor UN-a za ljudska prava, 48. sesija, 27.9.1993., stav 1.

14 Od niza drugih prava ne može se odstupiti, uključujući član 6. – pravo na život, član 7. – pravo da se ne bude podvrgnut mučenju ili okrutnom, nečovječnom ili ponižavajućem postupanju ili kažnjavanju, član 8. (stavovi 1. i 2.), pravo da se ne bude držan u ropstvu ili podređenosti te obavezu zabrane ropstva i trgovine robljem u svim njihovim oblicima.

15 Ahmed Shaheed, *Izveštaj specijalnog izvjestioca*, stav 3.

16 *Ibidem*, stav 2.

17 ICCPR, čl. 18.3. „Sloboda iskazivanja vjere ili uvjerenja može biti podložna samo ograničenjima koja su propisana zakonom i koja su nužna za zaštitu javne sigurnosti, reda, zdravlja ili morala ili temeljnih prava i sloboda drugih“. ICCPR, čl. 19.3. „Ostvarivanje prava predviđenih stavom 2. ovog člana nosi sa sobom posebne dužnosti i odgovornosti. Stoga može biti podložno određenim ograničenjima, ali ona će biti samo ona koja su propisana zakonom i koja su nužna radi:

- (a) Poštivanja prava i ugleda drugih;
- (b) Zaštite državne sigurnosti ili javnog reda (*ordre public*) ili javnog zdravlja i morala“.

time može dobiti pasivan karakter. Misao nije samo pokretač ljudske aktivnosti, već sama aktivnost kojoj treba dopustiti da se odvija neometano.

Kao i sva ljudska prava, sloboda misli često se preklapa s drugim pravima i slobodama zbog značajnih međusobnih odnosa i ovisnosti. Možda je jedan od najbitnijih primjera njen odnos s privatnošću, pravom koje se u određenim situacijama može ograničiti, u smislu zaštite (mentalnog) prostora za generiranje misli. Ovo preklapanje također ilustrira potrebu za daljnjim definiranjem opsega slobode misli i boljim pojašnjenjem njene zaštite, posebno s obzirom na apsolutnu prirodu ovog prava. Na primjer, pod kojim okolnostima mentalna privatnost i integritet zaslužuju apsolutnu zaštitu u okviru prava na slobodu misli? I kada bi ograničena zaštita u okviru prava na privatnost bila dovoljna?¹⁸

Slobodu misli također treba shvatiti u odnosu na slobodu savjesti i slobodu vjeroispovijesti ili uvjerenja. Savjest, koja nam kao „unutrašnji sud“¹⁹ omogućava na svom najosnovnijem nivou da slobodno odlučujemo šta je ispravno, a šta pogrešno, funkcionira samo ako proizlazi iz autonomnog mišljenja, u smislu da nema uplitanja u autonomiju osobe da oblikuje svoje misli. Isto tako, postoji rizik da bi prakticiranje bilo koje religije ili sistema vjerovanja moglo biti prisilno ako ne postoji sloboda misli.

Proces formiranja misli ne događa se u vakuumu, već u interakciji s drugim ljudima, prostorima i informacijama, gdje zajednice također igraju ulogu. Iz ove perspektive, postoji važno preklapanje s pravima sadržanim u članu 19. koji govori o slobodi mišljenja i izražavanja.²⁰ Misao i izražavanje postoje

18 Akademska literatura razvila je niz kriterija koji pomažu u razlikovanju situacija koje zahtijevaju apsolutnu i ograničenu zaštitu kada je u pitanju sloboda misli, odnosno pravo na privatnost. Jedan kriterij uzima u obzir psihičke posljedice uplitanja (uglavnom u pogledu „ozbiljnosti“ i „značaja“) uz druge relevantne faktore, kao što je važnost ličnog interesa u neotkrivanju određenih vrsta mentalnog sadržaja (npr. seksualne orijentacije, političkog mišljenja, vjerskih uvjerenja) ili metoda uplitanja kojom se otkriva ili mijenja mentalno stanje osobe (npr. indoktrinacija ili ispiranje mozga). Nivo kontrole nad vlastitim mislima nakon uplitanja također je ovdje relevantan, kao i da li metoda podriva takvu kontrolu ili je zaobilazi. Drugi kriterij uzima u obzir karakteristike žrtve, gdje je ranjivost važna. Djeca, starije osobe, osobe s mentalnim invaliditetom ili osobe koje su lišene slobode ili se nalaze u drugim pritvorskim okruženjima posebno su ranjive. Razlike u moći važne su u tom smislu i mogu uključivati i situacije na radnom mjestu. Drugi kriterij je kontekst u kojem se uplitanje dogodilo, što zahtijeva procjenu od slučaja do slučaja. Vidi Sjors Ligthart i Naomi van de Pol, Sloboda misli: Apsolutna zaštita mentalne privatnosti i mentalnog integriteta? Razmatranje slučaja neurotehnologije u krivičnom pravosuđu, Patrick O’Callaghan i Bethany Shiner (urednici), Priručnik o pravu na slobodu misli, (Cambridge: Univerzitetsko izdanje, 2025.).

19 Sofie Møller, *Kantov sud razuma, Pravna metafora i normativnost u kritici čistog uma*, (Cambridge: Univerzitetsko izdanje, 2020.), Poglavlje 6 – **Moralna savjest kao praktični unutrašnji sud**, str. 96.-112.

20 Ostala relevantna prava uključuju: pravo svih na uživanje najvišeg dostižnog standarda fizičkog i mentalnog zdravlja (čl. 12. Međunarodnog pakta o ekonomskim, socijalnim i kulturnim pravima – **ICESCR**);

u „stalnoj povratnoj petlji“.²¹ Također ostaju različiti i zaštićeni na različite načine u međunarodnom okviru ljudskih prava. Misao se odnosi na aktivnost razmišljanja, od poticaja do rezultata, dok se mišljenje može shvatiti kao jedan od privremenih rezultata toka misli.²²

„(...) ljudi se razvijaju i napreduju u interakcijama s drugim ljudskim bićima i u poticajnom materijalnom i kulturnom okruženju, (...) autonomija nije samo individualna, već i relacijska jer proizlazi iz nečijih interakcija i pripadnosti zajednici i utječe na njih.“

— **Nacrt teksta Preporuke o etici neurotehnologije**, UNESCO, 9.4.2025., stavovi 16. i 44.

U ključnom izvještaju iz 2021. godine (i prvom značajnom komentaru unutar sistema UN-a o pravu na slobodu misli), specijalni izvjestilac UN-a za slobodu vjeroispovijesti ili uvjerenja (UNSR), Ahmed Shaheed, navodi četiri moguća atributa tog prava:

- Sloboda neotkrivanja vlastitih misli
- Sloboda od kažnjavanja za vlastite misli, stvarne ili pretpostavljene
- Zaštita od nedopuštenog mijenjanja misli
- Pozitivna obaveza država da stvore poticajno okruženje za slobodu misli.

Dok su prva dva atributa jednostavnija, zaštita od nedopuštenog mijenjanja misli općenito se odnosi na zabranu miješanja u mentalnu autonomiju. To

pravo na obrazovanje — također u smislu da „obrazovanje treba biti usmjereno na puni razvoj ljudske osobnosti i osjećaja dostojanstva“ (čl. 13. ICESCR); pravo na učešće u kulturnom životu i uživanje u blagodatima naučnog napretka i njegove primjene, pravo na zaštitu moralnih i materijalnih interesa koji proizlaze iz bilo kojeg naučnog, književnog ili umjetničkog djela čiji je osoba autor, ili obavezu država da poštuju slobodu neophodnu za naučno istraživanje i stvaralačku djelatnost (čl. 15. ICESCR) ili zabranu prisiljavanja bilo koga na pripadnost udruženjima (čl. 20.2. ICESCR-a) (**UDHR**).

21 Ahmed Shaheed, **Izvještaj specijalnog izvjestioca**, stav 18.

22 *Ibidem*, stav 21.; Vidi i Christoph Bublitz, *‘The Mind and Conscience are the Person’s Most Sacred Possessions’: The Origins of Freedom of Thought in the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights*, Patrick O’Callaghan and Bethany Shiner (urednici), Priručnik o pravu na slobodu misli, (Univerzitetsko izdanje, 2025.)

pretpostavlja odsutnost prisile, modifikacije misli putem utjecaja na sam rad mozga odnosno umne manipulacije.

Sloboda misli, međutim, ne jamči „imunitet“ od tuđih misli ili od svakodnevnih procesa uvjeravanja. UNSR nadalje predlaže, između ostalih faktora, moguće kriterije za prepoznavanje neprimjerenih oblika manipulacije, koji bi se trebali procjenjivati od slučaja do slučaja:

- Pristanak (osobe koja je sposobna dati pristanak, slobodna i informirana)
- Prikrivanje i zamagljivanje (ako bi „razumna osoba“ bila svjesna namjeravanog utjecaja)
- Asimetrična moć između onoga koji vrši utjecaj i nositelja prava i kako je korištena
- Namjeravana ili počinjena šteta

Četvrti atribut, obaveza stvaranja poticajnog okruženja za slobodu misli, povezan je s drugim pravima, uključujući prava na pristup informacijama, komunikaciji i obrazovanju.²³

„Sloboda misli nam je potrebna kako bismo se borili protiv klimatskih promjena, rasizma i globalnog siromaštva te kako bismo se zaljubili, smijali i sanjali. Pravo na slobodu misli je individualno, ali je ključno za kulturni, naučni, politički i emocionalni život naših društava. Sloboda misli nam daje priliku da pomislimo nešto ružno i odgurnemo od sebe takve misli prije nego što djelujemo na temelju njih ili im dopustimo da se ukorijene. Omogućava nam da biramo kako ćemo se ponašati prema drugima, da svoj govor prilagodimo kontekstu i publici, te da budemo svoji. Sloboda misli nam omogućava da zamislimo nove budućnosti bez potrebe da ih prvo dokažemo, održava nas dinamičnim i avanturistički raspoloženim, štiti nas i povrh svega, čini nas ljudima.“

— Susie Alegre, *Sloboda misli: Zaštita osnovnog ljudskog prava u digitalnom dobu*, (London, Atlantic Books, 2022.), Uvod.

²³ Ibidem, stavovi 25.-47.

2. O umjetnoj inteligenciji

U međunarodnom pravu ne postoji univerzalno prihvaćena definicija „umjetne inteligencije“. Akt Evropske unije (EU) o umjetnoj inteligenciji (UI),²⁴ defini- ra „UI sistem“ kao „mašinski sistem dizajniran za rad s promjenjivim nivoima autonomije i koji nakon uvođenja može pokazati prilagodljivost, te koji, za eksplicitne ili implicitne ciljeve, na osnovu ulaznih vrijednosti koje prima, zaključuje kako generirati izlazne vrijednosti kao što su predviđanja, sadržaj, preporuke ili odluke koje mogu utjecati na fizička ili virtuelna okruženja“.²⁵ Slična definicija postoji i u Konvenciji Vijeća Evrope o umjetnoj inteligenciji i ljudskim pravima, demokratiji i vladavini prava.²⁶

UI se oslanja na algoritme²⁷ napisane u kompjuterskom kodu koji obra- đuju skupove podataka tako što prikupljaju i transformiraju neobrađene podatke kako bi proizveli određene izlaze.²⁸ Za razliku od algoritma, kod kojeg je proces donošenja odluka unaprijed definiran, UI „donosi odluke“ na osnovu analize podataka.²⁹ Identificira obrasce u podacima, „uči“ iz njih i pravi predviđanja, koja zatim može prilagođavati dok nastavlja prikupljati nove podatke.³⁰

Potencijalne primjene UI gotovo su beskrajne, a posljedice po živote ljudi značajno variraju. Umjetna inteligencija funkcioni- ra na osnovu skupova po- dataka koji odražavaju društvene stvarnosti i sadrže lične podatke, uklju- čujući podatke iz oblasti zdravstva, zapošljavanja, bankarstva, kupovine ili provedbe zakona. Na primjer, u oblasti zdravstvene zaštite, može analizirati skupove medicinskih podataka kako bi predvidjela zdravstvene ishode koji pomažu u donošenju ljekarskih odluka. Umjetna inteligencija se također

24 Akt EU o umjetnoj inteligenciji, Uredba (EU) 2024/1689, službeno izdanje od 13.6.2024.

25 *Ibidem*, čl. 3. (1).

26 Okvirna konvencija o UI i ljudskim pravima, demokratiji i vladavini prava, Vijeće Evrope, usvojena 5.9.2024.

27 Algoritmi su set instrukcija usmjerenih na rješavanje određenog problema. Procesi njihovog rada su složeni i raznoliki. Vidi, na primjer, Andrew Williams, *What Is an Algorithm? Defining And Applying Algo- rithms*, Forbes, 12.1.2024.

28 UNSR o promociji i zaštiti prava na slobodu mišljenja i izražavanja, David Kaye, A/73/348, 29.8.2018.

29 Kaya Ismail, *AI vs. Algorithms: What's the Difference?*, CMSWire, 26.10.2016.

30 Office of Communications, College of Education, *Traditional AI vs. Generative AI: What's the Difference?*, University of Illinois Urbana-Campaign, 11.11.2024.

koristi za analizu velikih online skupova podataka o ponašanju potrošača kako bi se utvrdile njihove individualne sklonosti. Te informacije zatim može povezati s ličnim i kontekstualnim karakteristikama (npr. dob, geografska lokacija, spol, ranije kupovine i potrošačke navike) kako bi kreirala detaljne profile kupaca. Kompanije koriste ove profile za segmentaciju tržišta, što im omogućava da pojedincima nude njima prilagođene proizvode ili usluge s ciljem povećanja prodaje ili interakcije s određenim sadržajem. Veća interakcija omogućava prikupljanje novih korisničkih podataka.

Iako su korisnici manje-više svjesni korištenja svojih podataka ili utjecaja na njihove izbore, ti procesi su i dalje problematični u smislu slobode misli i potencijala za manipulaciju. Isto tako, korisnici nemaju dovoljno informacija za učešće u raspravama o tome kako se nove tehnologije mogu koristiti u javnom interesu.

Generativna UI

Generativna umjetna inteligencija (GenAI) nova je generacija razvoja umjetne inteligencije. Odnosi se na umjetnu inteligenciju koja „kreira“ sadržaj, a koji se uglavnom koristi u kontekstu odgovaranja na upite (promptove) putem interfejsa zasnovanih na prirodnom jeziku (NLI).³¹ „Tradicionalna“ umjetna inteligencija proizvodi izlaze na bazi postojećeg sadržaja. GenAI generira „novi“ sadržaj na bazi postojećeg sadržaja.³² To može uključivati tekst, slike, videozapise, muziku, pa čak i softverski kod. GenAI se obučava na ogromnim količinama podataka. Kako bi generirao svoje izlaze, provodi sofisticiranu analizu obrazaca i odnosa unutar podataka, kao što su nizovi riječi ili rasporedi piksela.³³ Zasniva se, i njen razvoj je omogućen na osnovu takozvanih velikih jezičkih modela i dubokog učenja,³⁴ čiji su razvoj i unapređenje trajali godinama prije nego što je ChatGPT postao široko dostupan za eksperimentiranje i svakodnevnu upotrebu.

Nakon što je OpenAI učinio ChatGPT (svoj GenAI model obučen na internet-skim podacima) javno dostupnim, on je postao najbrže rastuća aplikacija,

31 Fengchun Miao, Wayne Holmes, *Guidance for generative AI in education and research*, (Paris: UNESCO, 2023).

32 Ured za komunikacije, Obrazovni koledž, *Traditional AI vs. Generative AI: What's the Difference?*.

33 Miao i Holmes, *Guidance for generative AI in education and research*, str. 8.

34 Cambridge English Dictionary Online, Entry for *Large Language Model*.

što je pokrenulo utrku među velikim tehnološkim kompanijama u razvoju GenAI alata i njihovoj integraciji u različite proizvode i platforme.³⁵ Posebno je važno napomenuti da je, uslijed brzog širenja ove tehnologije, došlo do naglog porasta sadržaja generiranog umjetnom inteligencijom, koji kroz nadogradnju sadržaja generiranog ljudskim djelovanjem, mijenja ukupnu prirodu podataka na internetu, a time i dostupno znanje.³⁶

Jedna od bojazni koja pogađa i ovu industriju, iako bi se to moglo promijeniti s novim tehnološkim napretkom, odnosi se na činjenicu da su istraživanja pokazala kako dolazi do kolapsa GenAI modela kada se rekurzivno obučavaju na umjetno generiranim podacima, za razliku od podataka koje generiraju ljudi.³⁷

„Kolaps modela predstavlja degenerativni proces koji pogađa generacije obučenih generativnih modela, u kojem podaci koje generiraju na kraju zagađuju skup podataka za obuku sljedeće generacije. Budući da su obučeni na zagađenim podacima, oni zatim pogrešno percipiraju stvarnost.“

— Ilya Shumailov, Zakhar Shumaylov, et.al., *AI models collapse when trained on recursively generated data*, *Nature*, 631, 24. 6. 2024., str. 755.-759.

Početkom 2025. godine, lansiranje DeepSeek-a, GenAI modela razvijenog u Kini, ponudilo je visoke tehničke performanse uz znatno niže troškove, uključujući i u pogledu zaštite okoliša. DeepSeek je pokrenut kao otvoreni kod, omogućujući svima da ga preuzmu, kopiraju i nadgrade bez potrebe za značajnim troškovima izgradnje modela od nule,³⁸ čime su nova naučna otkrića postala javno dostupna.³⁹ Međutim, njegovo pojavljivanje je izazvalo

35 Drugi primjeri GenAI modela: Claude (Anthropic), Llama (Meta), Gemini (Google), Le Chat (Mistral) i dr. Velike kompanije ostvaruju veliki razvoj, ali primjena nije ograničena na „Big Tech“ (velike tehnološke kompanije).

36 Vidi Jason Koebler, *Project Analyzing Human Language Usage Shuts Down Because 'Generative AI Has Polluted the Data'*, 404media, 19.9.2024.

37 Vidi Ilya Shumailov, Zakhar Shumaylov, et.al., *AI models collapse when trained on recursively generated data*, *Nature*, 631, str. 755-759, 24.7.2024.

38 Charlotte Edmond, *What is open-source AI and how could DeepSeek change the industry*, *World Economic Forum*, 5.2.2025.

39 Alex He, *DeepSeek and China's AI Innovation in US-China Tech Competition*, *Centre for International Governance Innovation*, 11.4.2025.

i zabrinutost u vezi sa pravom na privatnost i cenzurom,⁴⁰ te su neke zemlje EU-a zabranile njegovu upotrebu pozivajući se na zaštitu podataka.⁴¹

Ipak, utjecaj DeepSeek-a na konkurenciju u tom području i širenje GenAI modela tek treba u potpunosti sagledati.⁴²

Bilo da su zaista blizu toga ili ne,⁴³ velike tehnološke kompanije govore i o konkretnim koracima prema sljedećoj fazi umjetne inteligencije: Općoj umjetnoj inteligenciji (OUI). Očekuje se da će OUI raditi autonomno i pokazivati sposobnost djelovanja sličnu ljudskoj.⁴⁴ To ima implikacije po slobodu misli, jer bi takvi sistemi mogli utjecati na ovo pravo, podržavati ga ili podrivati. Stoga je ključno da vlade, programeri i regulatori osiguraju da dizajn i upravljanje potencijalnom općom umjetnom inteligencijom (OUI) bude zasnovano na standardima ljudskih prava.

Korporativni utjecaj u doba UI

Ekonomsko i regulatorno okruženje u kojem nove tehnologije djeluju pro- tekle dvije decenije već je imalo dubok utjecaj na demokratiju i ljudska prava. Korištenje UI od strane velikih tehnoloških kompanija, često nazivanih „Big Tech“, koje upravljaju glavnim internetskim platformama (uključujući pretraživače i društvene mreže), značajno je transformiralo prostor za javni diskurs. Ova transformacija promijenila je način na koji se vodi demokrat- ska debata, kao i način na koji ljudi formiraju svoja mišljenja, prave izbore i donose odluke, s direktnim implikacijama po slobodu misli.⁴⁵ Kako kor-

40 Robert Booth and Dan Milmo, *Experts urge caution over use of Chinese AI DeepSeek*, *The Guardian*, 28.1.2025.

41 Hakan Ersen and Miranda Murray, *DeepSeek faces ban from Apple, Google app stores in Germany*, *Reuters*, 27.6.2025.

42 Hamilton Mann, *Seek Deeper On DeepSeek For Artificial Integrity Over Intelligence*, *Forbes*, 28.1.2025; Robert Booth and Dan Milmo, *Chinese AI chatbot DeepSeek censors itself in realtime, users report*, *The Guardian*, 28.1.2025; Cade Metz, *How Did DeepSeek Build Its A.I. With Less Money?*, *The New York Times*, 12.2.2025.

43 Kate Brennan, Amba Kak and Dr. Sarah Myers West, *Artificial Power: 2025 Landscape Report*, *AI Now Institute*, 3.6.2025.

44 Tom Allen, *Real-world agentic AI is 'complex' cautions Google's Demis Hassabis*, *Computing*, 7.3.2025; Will Knight, *Google's AI Boss Says Gemini's New Abilities Point the Way to AGI*, *Wired*, 28.5.2025; Robert Booth, *Meta to announce \$15bn investment in bid to achieve computerised 'superintelligence'*, *The Guardian*, 11.6.2025.

45 Vidi detaljnije o argumentu u Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, (New York, Public Affairs, 2019). Vidi i OSCE RFoM o UI i slobodi izražavanja.

poracije i vlade dobivaju sve veći pristup velikim količinama podataka o ponašanju i sklonostima pojedinaca, povećava se i potencijal za utjecaj ili čak manipuliranje najdubljim mislima ljudi. Bez obzira financiraju li se javno ili privatno, tehnologijama utemeljenim na UI, uz nekoliko iznimki, trenutno nedostaju djelotvorni, na ljudskim pravima zasnovani javni regulatorni okviri koji bi osigurali prethodnu procjenu utjecaja ili naknadnu odgovornost.

Neurotehnologije se također ubrzano razvijaju i postaju sve sofisticiranije kroz svoju konvergenciju s umjetnom inteligencijom. Ove tehnologije otvaraju mogućnost čitanja i utjecaja na misli ljudi, budući da djeluju na mozak i neuronski krug. Komercijalizacija uređaja koji prikupljaju neuronske podatke povećava mogućnosti profiliranja i nadzora. To je direktno povezano sa slobodom misli. Iako ovakvi uređaji potencijalno nude revolucionarna rješenja, posebno na polju medicine, mogu omogućiti i pretpostavlja se da mogu ponuditi neviđene uvide u nešto što se dugo smatralo nepovredivom unutarnjom sferom, a to je svetost uma. Kada se koriste za profit, ove tehnologije imaju potencijal da doprinesu i pojačaju moć i sofisticiranost utjecaja na individualne izbore koji nameću velike online platforme, što predstavlja nove etičke izazove i izazove na polju ljudskih prava.⁴⁶

Tehnologije umjetne inteligencije prvenstveno razvijaju i primjenjuju akteri iz privatnog sektora, koji djeluju unutar komercijalnih okvira gdje je generiranje profita ključno za održivost. Kao rezultat toga, te se tehnologije uglavnom oblikuju van demokratskog nadzora i tradicionalnog „društvenog ugovora“ između vlada i građana, uprkos njihovom značajnom utjecaju na pitanja od javnog interesa. Ova nepovezanost znači da postojeće strukture upravljanja trenutno ne mogu adekvatno regulirati širinu utjecaja koji te tehnologije imaju na društvo. To također znači da komercijalni prioriteti tih privatnih aktera mogu, ali i ne moraju biti usklađeni sa širim ciljevima javne dobrobiti, što stvara probleme kada se želi osigurati da razvoj umjetne inteligencije uvijek služi općem dobru.

S obzirom na dubok utjecaj umjetne inteligencije na demokratske institucije, ljudska prava i interakciju između ljudi, važno je odgovoriti na pitanje o tome ko bi trebao razvijati i primjenjivati takve tehnologije, pod kojim uvjetima, u skladu s kojim propisima te s kojim ciljem. Osim dominantnih velikih

46 Vidi, na primjer, Nita A. Farahany, *The Battle for Your Brain. Defending the Right to Think Freely in the Age of Neurotechnology*, (New York: St Martin's Griffin, 2023).

tehnoloških kompanija, postoje brojne novoosnovane (start-up) kompanije i drugi konkurenti na polju umjetne inteligencije.⁴⁷ Industrijski, a posebno etički standardi, još se uvijek razvijaju, a stvarne prakse često ovise o ekonomskoj moći ili poslovnoj kulturi pojedinih kompanija ili okruženju u kojem posluju.

Države moraju preuzeti odgovornost i osigurati da „revolucija umjetne inteligencije“ poštuje ljudska prava i osnovne slobode svih. Pritom je važno da države promoviraju jednake uvjete koji potiču poštenu konkurenciju među svim akterima i sprječavaju „utrku do dna“. Države imaju stalnu dužnost stvarati i štititi poticajno okruženje za slobodu misli i sva druga ljudska prava, istovremeno pažljivo prateći utjecaj velikih tehnoloških kompanija kako bi osigurale odgovornost i usklađenost s javnim interesom.⁴⁸

47 Rashi Shrivastara (ed.), *Forbes 2025 AI 50 List - Top Artificial Intelligence Companies Ranked*, Forbes, 10.4.2025.; Kate Brennan et al, *Artificial Power: 2025 Landscape Report*.

48 Ostala relevantna područja sa specifičnim implikacijama po slobodu misli, koja ovdje nisu obrađena, uključuju korištenje umjetne inteligencije u različite svrhe nadzora, posebno u svrhu nacionalne sigurnosti ili odbrane, na granicama i u kontroli migracija, tokom pritvora, od strane vojske ili kao dio autoritarnih režima. Vidi i *Digital technologies at borders: A threat to people on the move*, UN OHCHR, 9.10.2023.

3. Ljudskopravaške implikacije razmišljanja unutar algoritma

UI, pristranost i misao

Učinci umjetne inteligencije ovise o svrsi i dizajnu sistema umjetne inteligencije, kvaliteti podataka i načinu na koji se umjetna inteligencija primjenjuje. Određeni sistem umjetne inteligencije odražavat će i pristranosti i nesvjesne predrasude (slijepe tačke) onih koji su učestvovali u njegovom stvaranju, kao i one ugrađene u skupove podataka korištene za njegov razvoj i obuku. Nadalje, može se primijeniti u svrhe drugačije od onih za koje je stvoren i testiran, dok se u algoritmima za mašinsko učenje kao takvima može pojaviti niz inherentnih problema.⁴⁹ U područjima koja tradicionalno uključuju društvenu interakciju (npr. bankarstvo, socijalna davanja, stanovanje, zapošljavanje), ako skupovi podataka već sadrže pristranosti, to će stvoriti posebno problematične posljedice, jer sistemi umjetne inteligencije imaju moć da šire i potencijalno produbljuju takve pristranosti.⁵⁰ Sistemi umjetne inteligencije također često odražavaju vrijednosti i stvarnost globalnog sjevera, što povećava vjerojatnost da će umjetna inteligencija pogrešno predstavljati druge regije. Sva ta pitanja nose rizike po ljudska prava, uključujući pravo na slobodu misli. Međutim, mora se i dalje raditi na rješavanju ovog problema i smanjenju stepena pristranosti, između ostalog, promocijom raznolikosti tokom cijelog životnog ciklusa umjetne inteligencije, radom na rješavanju tehničkih izazova i, u nekim slučajevima, nekorištenjem određenih vrsta umjetne inteligencije ako je rizik previsok, ako se želi unaprijediti moć razmišljanja, zaštititi raznolikost i osigurati različite perspektive.

49 *Bias in Algorithms: Artificial Intelligence*, EU Agency for Fundamental Rights (FRA), (Vienna, FRA, 2022).

50 Na primjer, na prediktivne algoritme policijskog djelovanja može utjecati niz faktora, uključujući i to je li skup podataka utemeljen na prethodno zabilježenim stopama kriminala. Na te stope, opet, mogu utjecati predrasude prema određenim manjinama koje žive u određenim područjima, a koje se zatim češće pojavljuju u skupovima podataka kao područja s visokim stopama kriminala nego što bi to bio slučaj u stvarnim (evidentiranim i neevidentiranim) stopama kriminala. Algoritamska predviđanja tako mogu postati pristrana prema određenim manjinama, čija naselja mogu postati predmet prekomjernog policijskog nadzora. Za sveobuhvatno objašnjenje pristranosti u prediktivnim sistemima policijskog djelovanja s umjetnom inteligencijom, vidi FRA, *Bias in Algorithms*. Akt EU-a o UI zabranjuje individualne prediktivne policijske sisteme umjetne inteligencije isključivo zasnovane na profiliranju ili osobinama ličnosti, osim kada se koriste za nadopunu procjena pojedinaca zasnovanih na objektivnim, provjerljivim činjenicama direktno povezanim s kriminalnim aktivnostima (vidi i poglavlje 6 u nastavku).

„Povratna petlja nastaje kada predviđanja koja generira sistem utječu na podatke koji se zatim koriste za ažuriranje tog istog sistema. To znači da algoritmi utječu na algoritme, jer njihove preporuke i predviđanja oblikuju stvarnost u praksi. Ta stvarnost tada postaje osnova za prikupljanje podataka za ažuriranje algoritama. Time, izlaz sistema postaje budući ulaz u isti taj sistem. Svaka pristranost u algoritmima na taj način se potencijalno može vremenom pojačati i pogoršati.“

— EU Agency for Fundamental Rights, *Bias in Algorithms: Artificial Intelligence and Discrimination*, (Vienna, FRA, 2022), str. 8.

Kada se ljudske interakcije i odluke zamijene sistemima umjetne inteligencije ili kada se previše oslanjamo na takve sisteme, okruženje za informirano donošenje odluka iskrivljuju postojeće ili sistemski generirane nejednakosti i netačnosti, koje sistem zatim dodatno pojačava. Tamo gdje pristran sadržaj na društvenim mrežama ili profiliranje pomoću umjetne inteligencije utječe na političke i druge društvene aktivnosti, osobe sa određenim zaštićenim karakteristikama ili obrascima ponašanja često su najdirektnije pogođene.⁵¹ U konačnici, ovo utječe na društvo u cjelini.

S obzirom na njihovu složenost i obim, uređivanje skupova podataka s ciljem izbjegavanja pristranosti, umjesto pokušaja njenog smanjivanja, izazovan je zadatak. Čak i kada se u najboljoj namjeri uloži trud u uklanjanje pristranosti iz alata umjetne inteligencije, greške u procesu mogu same po sebi dovesti do (daljnjih) kršenja ljudskih prava. Na primjer, platforme društvenih mreža koje koriste umjetnu inteligenciju za moderiranje (filtriranje, označavanje, uklanjanje ili ograničavanje širenja nezakonitog i uvredljivog sadržaja) već djeluju u širokim razmjerama. U nastojanju da se nosi sa složenošću ljudske komunikacije na tim platformama, umjetna inteligencija često ne uspijeva razumjeti ni kontekst ni nijanse (npr. satiru), što ograničava kvalitet moderiranja ljudske komunikacije općenito, a da ne spominjemo složenost pristranosti.⁵² Osim toga, značenje varira ovisno o okolnostima, zajedničkom razumijevanju i jeziku te se s vremenom mijenja. Prilikom pokušaja korištenja umjetne inteligencije za moderiranje pristranosti, platforme društvenih mreža mogu nenamjerno cenzurirati marginalizirane zajednice

51 Mehnaz Rafi, *When AI plays favourites: How algorithmic bias shapes the hiring process*, *The Conversation*, 14.10.2024. Vidi i *ICO considers uses of neurotechnology in employment in the UK*, *Withers*, 31.7.2023.

52 *Spotlight on AI and Freedom of Expression – A Policy Manual*, OSCE/RFoM, 20.1.2022.

zbog korištenja i interakcije na određenom jeziku, što se odnosi i na zaštićene protunarative koji se suprotstavljaju štetnom govoru i govoru mržnju ili ga osporavaju.⁵³

Istovremeno, sadržaj koji koristi „kodirani“ jezik može izbjeći detekciju. Ti su izazovi izraženiji kod sadržaja koji nije na engleskom jeziku, budući da kompanije ulažu manje resursa u moderiranje i provjeru činjenica.⁵⁴ Napori usmjereni na uređivanje podataka radi smanjenja pristranosti, koji se uglavnom oslanjaju na umjetnu inteligenciju, trebali bi od samog početka životnog ciklusa sistema umjetne inteligencije podrazumijevati adekvatno učešće različitih i dobro informiranih ljudi.

Zabrinjavajuće je da kompanije koje upravljaju društvenim mrežama trenutno mijenjaju svoje politike o moderiranju sadržaja i zaštitnim mehanizmima koji se odnose na povjerenje i sigurnost korisnika, uključujući mjere za ublažavanje govora mržnje, uznemiravanja i širenja netačnog sadržaja. Značajno su smanjile broj osoblja potrebnog za adekvatno moderiranje⁵⁵ i zato postaju sve manje efikasne u suzbijanju štetnog sadržaja.⁵⁶ S obzirom na štetu koju takav sadržaj nanosi marginaliziranim zajednicama (uključujući opasnost koju govor mržnje na internetu može izazvati u stvarnom životu), napore usmjerene na ublažavanje pristranosti trebalo bi u najmanju ruku pojačati.⁵⁷

Jedna posebno važna karakteristika digitalnog okruženja je dominacija engleskog jezika, što ne odražava raznolikost jezika koji se govore širom svijeta. To je produbilo digitalni jaz u smislu pristupa internetu za one koji ne govore engleski ili druge dominantne jezike na internetu.⁵⁸ GenAI dodatno produbljuje ovaj problem jer se obučava na ograničenom broju jezika

53 *FRA, Bias in Algorithms: Artificial Intelligence and Discrimination.*

54 *Content Moderation in a New Era for AI and Automation, Meta Oversight Board*, septembar, 2024.; vidi i *OSCE/RFoM, Spotlight on AI and Freedom of Expression – A Policy Manual.*

55 *David Evan Harris and Aaron Shul, Generative AI, Democracy and Human Rights*, Centre for International Governance and Policy, Policy Brief No. 12., februar, 2025.

56 *Joel Kaplan, More Speech and Fewer Mistakes, Meta*, 7.1.2025.; *Clare Duffy, Calling women 'household objects' now permitted on Facebook after Meta updates its guidelines*, *cnn.com*, 8. 1. 2025.; *Dia Kayyali, Meta's Content Moderation Changes are Going to Have a Real World Impact. It's Not Going to be Good*, *TechPolicy.press*, 9. 1. 2025.; *Adrian Kopps, Four key policy changes of X under Musk*, *Digital Society Blog*, *Alexander von Humboldt Institut für Internet und Gesellschaft*, 28.10.2024.

57 *Hate Speech*, *United Nations-dedicated webpage*; *X's design and policy choices created fertile ground for inflammatory, racist narratives targeting Muslims and migrants following Southport attack*, *Amnesty International*, 6.8.2025.

58 *Internet Access: UNESCO and ICANN join forces to improve linguistic diversity online*, *UNESCO*, 27.2.2025.

i stoga popunjava informacijsko okruženje prvenstveno podacima na tim jezicima. Ovaj trend ugrožava jezičku raznolikost i malo je vjerojatno da će se promijeniti bez značajnih ulaganja kompanija.⁵⁹

Maternji jezici nisu samo bitni za kulture i identitete, nego su i sastavni dio načina na koji ljudi organizuju svoje misli u forum internumu i načina na koji se oblikuju svjetonazori ljudi. S pozitivne strane, pored ostalih dobrobiti za javnost,⁶⁰ umjetna inteligencija uspješno se koristi za dokumentiranje, očuvanje i zaštitu ugroženih jezika, pa čak i za oživljavanje jezika koji se više ne koriste.⁶¹ U ovom smislu, od suštinske važnosti je etos ljudskih prava i razumijevanje koje oblikuje razvoj, primjenu i korištenje umjetne inteligencije.

Bez odgovarajuće brige za osiguranje raznolikosti i zastupljenosti, uključujući i uklanjanje digitalnog jaza, budućnost umjetne inteligencije, u kojoj se (ponovo) stvara ogroman disbalans moći korištenjem novih tehnologija koje djeluju u širokim razmjerama, nosi rizik da od nestanka svjetonazora koji nisu zastupljeni u sistemima umjetne inteligencije.

Ljudski nadzor, transparentnost i odgovornost

Iz perspektive osiguranja poštivanja ljudskih prava, nivo ljudske intervencije u procesu donošenja odluka dobija ključni značaj. Potpuno automatizirano donošenje odluka znači da nema ljudske intervencije (uključujući ljudsku intuiciju ili sposobnost rasuđivanja) u načinu na koji umjetna inteligencija isporučuje rezultat. Nadalje, složenost GenAI-a otežava ili čak onemogućava razumijevanje ili objašnjenje na koji način je postignut određeni rezultat. Dodatnu komplikaciju predstavlja interakcija između čovjeka i mašine, odnosno fenomen poznat kao „pristranost automatizacije“, gdje su studije pokazale da ljudi općenito više vjeruju mašinama nego ljudima i zato lakše prihvataju rezultate generirane umjetnom inteligencijom.⁶² Sloboda misli i stvarna ljudska autonomija u odlučivanju dovode se u pitanje kada postaje

59 Viorica Marian, *AI could extinguish languages and ways of thinking*, *The Washington Post*, 19.4.2023.

60 Vidi UN-ov *AI for Good platform* and Stéphanie Bascou, « *IA d'intérêt public* » : *cette ONG veut créer un label éthique, similaire au « Commerce équitable », et facilement identifiable pour les consommateurs*, (UI u javnom interesu. Ova nevladina organizacija želi stvoriti etičku oznaku, sličnu „Poštenoj trgovini“ (Fair Trade) i lako prepoznatljivu za potrošače), 01.net, 5.10.2024. (na francuskom).

61 *AI: The Unexpected Hero in the Battle to Save Dying Languages*, *apolitical website*, 20.10.2024.

62 Chris Baraniuk, *Why we place too much trust in machines*, *BBC*, 20. 10. 2021.

teže provjeriti i preispitati rezultate, posebno u posljedičnim situacijama, čak i ako su ljudi uključeni u odlučivanje o tome kako djelovati na osnovu tih rezultata.⁶³ Nivo rizika varira ovisno o području u kojem se koristi umjetna inteligencija. Na primjer, ljekari bi trebali u potpunosti razumjeti umjetnu inteligenciju koju koriste i biti obučeni za prevladavanje pristranosti automatizacije, sistematski procjenjujući rezultate koju generira umjetna inteligencija za svakog pojedinačnog pacijenta, budući da su rizici u pogledu negativnih zdravstvenih ishoda vrlo visoki.⁶⁴

Mogućnost provjere i osporavanja rezultata je često otežana pozivanjem na prava intelektualnog vlasništva nad kodom, parametrima i/ili drugim internim operacijama umjetne inteligencije. Ti se argumenti koriste kako bi se izbjeglo otkrivanje načina na koji je određen rezultat postignut. Kada se uklone transparentnost i mogućnost odlučivanja o informacijskom okruženju (npr. ono što algoritmi društvenih mreža prikazuju u feed-ovima), to dodatno otežava informirano donošenje odluka ili čak pravilno razumijevanje vlastitog okruženja.

„Razumljivost se odnosi na stepen do kojeg osoba može razumjeti određeni sistem. Dok se predvidljivost sistema odnosi na pitanje - *šta će sistem učiniti*, razumljivost se odnosi na pitanje - *zašto on to radi*?“

Arthur Holland Michel, *The Black Box, Unlocked*, UNIDIR, 22.9.2020., str. 9.

Nemogućnost razumijevanja načina na koji umjetna inteligencija donosi odluke koje utječu na ljudska prava, bilo zato što je sistem previše složen ili zato što njegov rad i sredstva za njegovu kontrolu ne predstavljaju javno dostupne informacije, izaziva ozbiljnu zabrinutost u pogledu odgovornosti i transparentnosti - dva ključna principa za pravilno funkcioniranje pravosudnih sistema i demokratskih država.

63 Radi sveobuhvatnijeg objašnjenja predvidljivosti i razumljivosti, vidi Arthur Holland Michel, *The Black Box, Unlocked*, UNIDIR, 22.9.2020.

64 *The application of artificial intelligence in healthcare and its impact on the "patient-doctor" relationship*, Council of Europe, Steering Committee for Human Rights in the Fields of Biomedicine and Health (CD-BIO), septembar 2024., str. 20.

„1. Digitalne tehnologije dramatično mijenjaju naš svijet. One nude ogroman potencijal za unapređenje dobrobiti i napretka ljudi, društava i naše planete (...) 3. Svjesni smo da tempo i snaga novih tehnologija stvaraju nove mogućnosti, ali i nove rizike po čovječanstvo, od kojih neki još nisu u potpunosti poznati. Svjesni smo potrebe za identificiranjem i ublažavanjem rizika te osiguranjem ljudskog nadzora nad tehnologijom na način koji doprinosi održivom razvoju i punom ostvarivanju ljudskih prava.“

— *UN Global Digital Compact*, annexed to the United Nations General Assembly resolution A/RES/79/1 on the Pact for the Future, usvojen 22.9.2024.

Ovi rizici se mogu ublažiti drugačijim razvojem umjetne inteligencije,⁶⁵ na odgovoran način usmjeren na ljudska prava, uz osiguravanje regulatornog nadzora i odgovarajućeg nivoa razumljivosti, transparentnosti i odgovornosti. Nadalje, smislen ljudski nadzor ključan je kako bi se osiguralo da sistemi umjetne inteligencije doprinose postizanju najboljih ishoda za ljude i ne ugrožavaju njihova prava, uključujući pravo na slobodu misli.

⁶⁵ *Recommendation on the Ethics of Artificial Intelligence*, UNESCO, 23.11.2024.; vidi takode Rumman Chowdhury i Dhanya Lakshmi, “Your opinion doesn’t matter, anyway”: exposing technology-facilitated gender-based violence in an era of generative AI, (Pariz, UNESCO 2023.); ili *Resources* stranicu organizacije Humane Intelligence.

4. Sloboda misli u rekonfiguriranom građanskom prostoru

Nova arhitektura građanskog prostora koja utječe na slobodu misli

Istinska sloboda mišljenja i izražavanja ovisi o istinskoj slobodi misli.⁶⁶ Kao što je objašnjeno, i misao i mišljenje dijelom su *forum internuma*, privatnog unutrašnjeg domena svakog pojedinca, područje koje uživa apsolutnu zaštitu od uplitanja, prema međunarodnom pravu ljudskih prava. Konceptualno, u demokratskoj teoriji, slobodna misao oblikuje se i razvija kroz razgovor dok osobe slobodno razmjenjuju mišljenja i informacije.⁶⁷ U ovakvom idealnom građanskom prostoru mora se omogućiti izražavanje, osporavanje i testiranje svih ideja i mišljenja, kako bi se osiguralo da najbolje ideje i istine u konačnici prevladaju. Mogućnost ove dinamične razmjene, gdje se ideje mogu susresti i sukobiti, razlog je zašto sloboda izražavanja (tj. izražavanje misli i mišljenja) uživa snažnu zaštitu i podršku. Ključna je za zdravu javnu raspravu koja je bitna za demokratiju. Novinarstvo u javnom interesu igra ključnu ulogu u omogućavanju informirane demokratske rasprave, kao i mogućnost slobodnog udruživanja i organiziranja, između ostalog. Međutim, postoje i ograničenja. Na primjer, države trebaju ograničiti „govor mržnje“ koji dostiže određeni prag ozbiljnosti, u skladu s međunarodnim pravom ljudskih prava.⁶⁸

S porastom Big Tech kompanija, veliki dio komunikacije premjestio se na platforme društvenih mreža koje sada posreduju (moderiraju) u ljudskim interakcijama i djeluju kao čuvari pristupa vijestima i informacijama. Ove platforme koriste snažne algoritme mašinskog učenja za reorganizaciju sadržaja na načine koji imaju implikacije po slobodu misli,

66 Sjors Ligthard, Christoph Bublitz, Thomas Douglas, Lisa Forsberg and Gerben Meynem, *Rethinking the Right to Freedom of Thought: A Multidisciplinary Analysis*, in *Human Rights Law Review*, 2022, 22, str. 1.-14.

67 Vidi i Simon McCarthy-Jones, *Freethinking: Protecting Freedom of Thought Amidst the New Battle for the Mind*, (London: Oneworld, 2023).

68 ICCPR, čl. 20: „1. Svaka ratna propaganda zabranjena je zakonom.

2. Svako zagovaranje nacionalne, rasne ili vjerske mržnje koje predstavlja poticanje na diskriminaciju, neprijateljstvo ili nasilje, zabranjeno je zakonom.“

jer temeljno mijenjaju način na koji informacije, mišljenja i ideje cirkuliraju, u usporedbi s informacijskim okruženjem kojeg su uglavnom organizirali tradicionalni mediji (u javnom interesu).⁶⁹

Online sistemi preporuka zasnovani na algoritmima su oblikovani i vođeni prvenstveno komercijalnim interesima. Ovi sistemi su optimizirani kako bi se postigao maksimalan angažman korisnika i poticaj na kupovinu umjesto utvrđivanja prioriteta u smislu tačnosti, raznolikosti ili ciljeva od javnog interesa, uključujući medijski pluralizam i novinarstvo u javnom interesu.⁷⁰ Maksimiziranje angažmana i poticanje kupovine podrazumijeva korištenje informacija o prethodnom ponašanju i individualno profiliranje kako bi se predvidjelo s kojim će sadržajem korisnici najvjerojatnije stupiti u interakciju. Kao rezultat toga, korisnici imaju ograničenu kontrolu nad onim što vide, čak i ako i dalje zadržavaju autonomiju u odlučivanju da li će s tim sadržajem stupiti u interakciju ili ne. Također, mogu biti nesvjesni činjenice da se sadržaj koji vide razlikuje od onog koji vide drugi korisnici, uključujući osobe iz njihovog kruga interakcije, zbog algoritamske personalizacije.

Ovaj model stavlja značajan utjecaj nad informacijskim ekosistemom svakog pojedinaca u ruke malog broja platformi. To je doprinijelo fragmentaciji medijskog informacijskog prostora, slabljenju neovisnog novinarstva i potencijalnom narušavanju zajedničkog osjećaja stvarnosti, koji je dugo oblikovao interakcije, odabire i demokratski angažman pojedinaca.⁷¹

69 „Za razliku od tradicionalnih algoritama, koje inženjeri „hardkodiraju“ (fiksno ugrađuju u kod), algoritmi mašinskog učenja „obučavaju se“ na ulaznim podacima kako bi prikupili informacije o korelacijama unutar njih. Obučeni algoritam, poznat kao model mašinskog učenja, zatim može automatizirati buduće odluke. Algoritam obučen na podacima o klikovima na oglase, na primjer, mogao bi naučiti da će žene češće kliknuti na oglase za helanke za jogu nego muškarci. Rezultirajući model tada će ženama prikazivati više tih oglasa.“ Vidi Karen Hao, *The Facebook whistleblower says its algorithms are dangerous. Here's why*, MIT Technology Review, 5.10.2021.

70 OSCE/RFoM, *Spotlight on AI and Freedom of Expression – A Policy Manual*.

71 OSCE/RFoM, *Spotlight on AI and Freedom of Expression – A Policy Manual*; i šire u radu OSCE RFoM on *Media and Big Tech*; vidi takođe *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes*, Odbor ministara Vijeća Evrope, usvojeno 13.2.2019.: “8. Savremeni alati za mašinsko učenje imaju sve veću sposobnost ne samo predviđanja izbora već i utjecaja na emocije i misli te promjene očekivanog toka djelovanja, ponekad i podsvjesno. (...) 9. Fino granulirani, podsvjesni i personalizirani nivoi algoritamskog uvjeravanja mogu imati značajan utjecaj na kognitivnu autonomiju pojedinaca i njihovo pravo na formiranje mišljenja i donošenje neovisnih odluka. (...) Centralni stubovi [Vijeća Evrope], ljudska prava, demokratija i vladavina prava, zasnivaju se na temeljnom vjerovanju u ravnopravnost i dostojanstvo svih ljudi kao neovisnih moralnih subjekata. (...) Odbor ministara - skreće pažnju na rastuću prijetnju po pravo ljudi da formiraju mišljenja i donose odluke neovisno o automatiziranim sistemima, koja proizlazi iz naprednih digitalnih tehnologija. Pažnju treba posebno posvetiti njihovoj sposobnosti korištenja ličnih i neličnih podataka za sortiranje i mikro ciljano oglašavanje, identificiranje individualnih ranjivosti i korištenje tačnog prediktivnog znanja i rekonfiguraciju društvenih okruženja kako bi se ostvarili specifični ciljevi i interesi”;

„Metode kojima online platforme uređuju sadržaj putem sistema preporuka nisu transparentne i vrlo rijetko podliježu javnoj i/ili državnoj kontroli. (...). Sistemi za preporuke sadržaja mogu imati i neželjene posljedice iz perspektive širih društvenih ciljeva te mogu negativno oblikovati i ometati apsolutno pravo na slobodu misli i slobodu mišljenja. Osim toga, procesi sistema za preporuke internetskih posrednika obično isključuju izbor, kontrolu i djelovanje pojedinačnih korisnika, što su preduvjeti za osiguranje individualne autonomije u traženju i prenošenju raznih informacija i ideja. (...) Postoje brojni dokazi da moć online platformi da oblikuju mišljenja ima sposobnost usmjeravanja i pojačavanja utjecaja određenih javnih narativa i vrsta diskursa u odnosu na druge. Za zemlje s krhkim ili opresivnim političkim sistemima, ova moć oblikovanja mišljenja, zajedno s algoritamskim pojačavanjem, može imati katastrofalne posljedice po individualno uživanje ljudskih prava“.

— *Spotlight on AI and Freedom of Expression – A Policy Manual*, OSCE/RFoM, 20.1.2022. str. 64. - 78.

Predstavnik OSCE-a za slobodu medija (RFoM) obavio je opsežno istraživanje u pokušaju da objasni utjecaj umjetne inteligencije na digitalno tržište ideja, medije od javnog interesa, slobodu mišljenja i izražavanja te demokratske procese, nudeći sveobuhvatne preporuke državama članicama.

Vidi više o tome u *OSCE RFoM Spotlight on Artificial Intelligence and Freedom of Expression* — *SAIFE Resource Hub*.

Kako se razvijaju mogućnosti umjetne inteligencije, oglašavanje koje se oslanja na poslovni model prikupljanja podataka zasnovan na umjetnoj inteligenciji, koji analizira individualno i kolektivno ponašanje i karakteristike, postaje sve djelotvornije. Ovaj sistem prati korisnike na različitim web lokacijama i uređajima, objavljujući ciljani sadržaj, uključujući vijesti, kako bi se maksimalno privukla pažnja i potaknule kupovine.

Pojava društvenih mreža nesumnjivo je omogućila nove kontakte i komunikaciju, olakšala mobilizaciju javnosti po pitanjima od zajedničkog interesa i dala glas onima koji su pod cenzurom. Međutim, u usporedbi s tradicionalnim javnim prostorima, platforme društvenih mreža mijenjaju način na koji se odvija međusobna interakcija ljudi i njihovo povezivanje unutar različitih

grupa. Ovaj komunikacijski okvir postavljen je tako da koristi ljudske psihološke pristranosti, posebno kada je riječ o dinamici vlastite grupe u odnosu na druge grupe i ljudskoj potrebi za pripadanjem, kako bi korisnicima pružao njima prilagođeni sadržaj prema kriterijima koji uglavnom nisu javno dostupni (tj. način na koji su algoritmi koji upravljaju sadržajem postavljeni i svrha kojoj služe). Platforme su transformirale velik dio javnog prostora i način na koji se možemo angažirati u njemu, a istovremeno održavaju jaku iluziju da imamo izbora.⁷² Kroz suptilne, ali sistematske efekte, ljudi postaju podložni utjecaju i manipulaciji mislima i ponašanjem, što podriva ideju o samostalnom razmišljanju i umanjuje našu autonomiju u djelovanju.⁷³

Moglo bi se reći da dominantan položaj velikih tehnoloških kompanija pojedincima ne ostavlja puno izbora nego da prihvate model nadzora i koriste platformu. To je postao uvjet za mnoge svakodnevne društvene i profesionalne interakcije, posebno tamo gdje se grupne aktivnosti odvijaju ili druge važne informacije prenose samo putem jedne platforme. To ima posljedice po slobodu misli. Izlažući se algoritmima kompanija koje filtriraju značajan dio njihovog doživljaja svijeta, ljudi riskiraju da izgube kontrolu nad svojim identitetom. Samo saznanje da su nadzirani može utjecati i na ono što ljudi misle i kako se ponašaju, potencijalno u strahu od posljedica.⁷⁴ Važno je da je

72 Novo informacijsko okruženje, posebno u kontekstu političke polarizacije, dovodi do stvaranja „eho komora“, tj., rezultata koji proizlazi iz sklonosti ljudi, uz istovremenu stvarnu mogućnost izbora na društvenim mrežama gdje mogu odabrati interakciju s ljudima poput sebe s čijim se idejama slažu. Iako je literatura podijeljena o postojanju eho komora, ona je vrlo ograničena i geografski i jezički te se općenito proizvodi bez odgovarajućeg pristupa podacima potrebnim za pravilnu analizu. Na osnovu jedne velike, kolaborativne meta-istraživačke studije na temu političkog angažmana zaključeno je da postoje dokazi da ljudi uranjaju u homogene, istomišljeničke mreže kada koriste društvene medije. No, slika je mnogo složenija. Autori također ističu da, iako društvene mreže mogu ponuditi ljudima više vijesti i ideja nego što bi inače saznali u svom stvarnom životu, kada su online, različita mišljenja i informacije se percipiraju iz polarizirane pozicije jer ljudi ne reagiraju na suprotstavljene stavove ulaskom u demokratsku raspravu, nego traženjem odobrenja od online zajednice, odnosno „svoje grupe“ nasuprot „druvoj grupi“, što u konačnici dovodi do još dublje ukorijenjenih pozicija i polarizacije. Vidi slučaj *Jonathan Haidt and Chris Bail (u toku)*, *Social Media and Political Dysfunction, A collaborative review*, neobjavljeni rukopis, Univerzitet u Njujorku, poglavlje 2.4).

73 Termin „nadzorni kapitalizam“ popularizirala je Shoshana Zuboff, koja ga opisuje kao novu ekonomsku logiku usmjerenu na izdvajanje ličnih podataka u svrhu oblikovanja ponašanja korisnika. „(...) Nadzorni kapitalisti otkrili su da najprediktivniji podaci o ponašanju dolaze iz intervencije u trenutno stanje kako bi se ponašanje potaknulo, pridobilo, podesilo i usmjerilo ka profitabilnim ishodima. Konkurentski pritisci proizveli su ovaj pomak, u kojem automatizirani mašinski procesi ne samo da poznaju naše ponašanje, već ga umnogome i oblikuju. S ovim preusmjeravanjem sa znanja na moć, više nije dovoljno automatizirati tokove informacija o nama, cilj je automatizirati nas.“ *Shoshana Zuboff, The Age of Surveillance Capitalism: the fight for a human future at the new frontier of power*, (New York: Profile Books, 2019), str. 8.

74 Nakon Snowdenovih otkrića o vladinom nadzoru, ljudi su promijenili svoje online ponašanje kako bi isključili pretraživanja o temama zbog kojih bi mogli biti prijavljeni vlastima. Vidi *Jon Penney, Chilling Effects: Online Surveillance and Wikipedia Use*, *Berkley Technology Law Journal*, svezak 31, br. 1, str. 117.-183., 27. 4. 2016. Vidi i *Ahmed Shaheed, Report of the Special Rapporteur*, stav 54.

EU počela poduzimati korake na rješavanju problema ciljanog oglašavanja u okviru različitih aspekata svog specifičnog zakonodavnog okvira. U trenutku pisanja ovog teksta, neki su slučajevi već bili riješeni, a drugi su još u toku.⁷⁵ Neki od ovih slučajeva će imati implikacije po zaštitu autonomije i djelovanja, a time i slobode misli u novom građanskom prostoru.

„...činjenica da operater online društvene mreže ima dominantan položaj na tržištu online društvenih mreža sama po sebi ne isključuje mogućnost da korisnici takve mreže daju valjanu saglasnost (...) tom operateru da obrađuje njihove lične podatke. To je ipak važan faktor u utvrđivanju je li saglasnost zaista valjana i slobodno data, što operater mora dokazati.“

— *Press release no. 113/23, Court of Justice of the European Union, Judgment of the Court in Case C-252/2001 | Meta Platforms and Others (General terms of use of a social network), Luxembourg, 4.7.2023., stav 154.*

Akademski literatura o slobodi misli sugerira da ovo novo okruženje ugrožava mentalnu autonomiju kroz: (1) nadzor i korištenje ličnih podataka koji daju vrlo detaljne uvide o pojedincima ili grupama; (2) netransparentnost tehnologija, kako u načinu na koji koriste umjetnu inteligenciju za generiranje i širenje sadržaja, tako i u smislu unutrašnjeg funkcioniranja same tehnologije; (3) iskorištavanje kognitivnih pristranosti; i (4) njen do sada nezabilježeni doseg, oblikujući društvene sklonosti i izbore na način koji zaobilazi demokratske procese.⁷⁶

⁷⁵ *Commission finds Apple and Meta in breach of the Digital Markets Act, European Commission Press release, 23.4.2025.; Vidi također Press release no. 116/24, Judgment of the Court in Case C-446/21 | Schrems (Communication of data to the general public), Court of Justice of the European Union, Luxembourg, 4.10.2024.; i Press release no. 113/23, Judgement of the Court in Case C-252/2001 | Meta Platforms and Others (General terms of use of a social network), Court of Justice of the European Union, Luxembourg, 4.7.2023. Sličan slučaj riješen je i u Velikoj Britaniji početkom 2025. Vidi Dan Milmo, Meta to stop targeting UK citizen with personalised ads after settling privacy case, The Guardian, 22.3.2025.*

⁷⁶ *Nina Keese and Mark R. Leiser, Online Manipulation as a Potential Interference with the Right to Freedom of Thought in Patrick O'Callaghan and Bethany Shiner (eds.), The Cambridge Handbook of the Right to Freedom of Thought, (Cambridge: Cambridge University Press, 2025.). Vidi i Kate Brennan et al, Artificial Power: 2025 Landscape Report.*

„Informacijski poremećaj“ mijenja istinski izbor

Reorganizacija našeg informacijskog ekosistema također je uticala na način na koji funkcioniraju demokratska rasprava i odlučivanje, pri čemu su mediji od javnog interesa posebno teško pogođeni. Širenje digitalnih tehnologija unaprijeđenih umjetnom inteligencijom, koje pokreću cikluse dezinformacija i pogrešnih informacija radi ostvarivanja određene političke, ideološke ili komercijalne koristi, a koje se putem digitalnih platformi šire mnogo brže nego prije,⁷⁷ stvorilo je stanje „informacijskog poremećaja“.⁷⁸ Ovaj fenomen narušio je povjerenje u neovisne medije i integritet informacija te doprinio zamagljivanju granica između istine i neistine. Kako se neizvjesnost povećava, opada i javno povjerenje u institucije i širi informacijski ekosistem, što utječe na kvalitet informiranog donošenja odluka potrebnog za učešće u demokratskom i svakodnevnom životu.⁷⁹

Kako je objasnio OSCE RFoM,⁸⁰ GenAI izaziva dodatne probleme novinarstvu. Zamjenjuje originalni sadržaj koji su stvorili novinari, a istovremeno stvara probleme zbog plagiranja bez isplaćivanja naknade, što bi moglo dodatno potisnuti novinarstvo i sadržaj od javnog interesa iz javne sfere.⁸¹

Paralelno s promjenom uloge medija i njihove sposobnosti djelovanja u javnom interesu i održavanja zdravog informacijskog okruženja,⁸² GenAI i sposobnost stvaranja *deepfake*-ova koje je sve teže otkriti stvaraju daljnje rizike za informacijsko okruženje, a posebno za izborne procese.

77 Report of the UNSR on FoOE Irene Khan, *Disinformation and Freedom of Opinion and Expression*, 13. 4. 2021., *A/HRC/47/25*, stav 2.

78 Vidi Claire Wardle PhD and Hossein Derakhshan, *Information Disorder: Towards an interdisciplinary framework for research and policy making*, (Strasbourg: Vijeće Evrope, 2018.). Wardle i Derakhshan koriste konceptualni okvir za opis informacijskog poremećaja u novom komunikacijskom okruženju, koji se bavi vrstama informacijskog poremećaja (dezinformacije, pogrešne informacije i maliciozne informacije) u tri faze informacijskog poremećaja (stvaranje, produkcija i distribucija), a zatim elementima informacijskog poremećaja (subjekt, poruka i interpretator). Fokusiraju se na štetu i lažnost kako bi definirali vrste informacija koje čine ekosistem informacijskog poremećaja. Pogrešne informacije se stoga javljaju kada se lažne informacije dijele bez štetnih namjera. Dezinformacija je lažna informacija koja se svjesno dijeli s namjerom da se nanese šteta. Maliciozna informacija je istinita informacija koja se dijeli s namjerom da se nanese šteta (npr. objavljivanje privatnih informacija).

79 UNSR on FoOE Irene Khan, *Disinformation and Freedom of Opinion and Expression*.

80 Vidi *Workshop on Big Tech and Media Freedom - Outcome Report*, OSCE/RFoM, 15.10.2024.

81 Pored novinarstva, GenAI izaziva i velike probleme kada je u pitanju poštivanje člana 15. ICESCR-a u vezi s pravom svake osobe da ima koristi od zaštite moralnih i materijalnih interesa koji proizlaze iz bilo kojeg naučnog, književnog ili umjetničkog djela čiji je autor.

82 Vidi rad OSCE RFoM-a na temu: *Media and Big Tech Initiative | OSCE*. Pored toga, u 2025. godini, OSCE RFoM priprema smjernice za države o očuvanju medijske slobode u doba Big Tech kompanija i UI, podržavajući na taj način razvoj zdravih online informacijskih prostora.

Ove se tehnologije mogu lako iskoristiti za manipulaciju, širenje dezinformacija i lažnih informacija s razornim efektom. *Deepfake*-ovi se često koriste u svrhu rodno zasnovanog uznemiravanja, posebno kroz proizvodnju i širenje seksualiziranih slika. Prvenstveno su usmjereni na žene i djevojčice, a među čestim metama su i LGBTI osobe. Žene u politici su posebno izložene riziku, kao i one koje žive u konzervativnim, patrijarhalnim društvima, gdje *deepfake* koji prikazuje osobu u „neprimjerenjoj“ situaciji može imati ozbiljne posljedice.⁸³

Pokazalo se, na primjer, da zlonamjerni akteri mogu koristiti GenAI modele otvorenog koda za počinjenje ili dodatno širenje rodno zasnovanog nasilja (GBV) u digitalnom prostoru.⁸⁴ Iako GenAI nije stvorio rodno zasnovano nasilje, njegova moć generiranja, umnožavanja i masovnog širenja rodno zasnovanog nasilja i drugih vrsta štetnog sadržaja uz male troškove⁸⁵ čini ga destruktivnijim ako ostane nereguliran i dostupan takvim akterima, dok se oni koji su njime pogođeni istiskuju iz online sfere ili su previše zastrašeni da bi reagirali.

GenAI modeli nude i specifičan tehnički izazov nazvan „halucinacije“. To je sadržaj koji proizvodi GenAI, a koji se udaljava od činjenične stvarnosti ili je jednostavno izmišljen. Međutim, obično je vrlo uvjerljiv i stoga ga je teško ili nemoguće otkriti. Halucinacije proizlaze prvenstveno iz načina na koji GenAI funkcionira. Problem nije samo u bazi podataka na osnovu koje GenAI radi (uključujući netačnosti ili lažne informacije npr. u slučaju podataka dostupnih na internetu), već i u načinu na koji je GenAI dizajniran. Naime, on je osmišljen tako da nastavi tok interakcije i ponudi odgovor.⁸⁶ Kompanija OpenAI je prepoznala da halucinacije, zajedno s drugim elementima poput namjernih dezinformacija ili društvenih predrasuda, mogu izazvati sumnju u cijelo informacijsko okruženje, ugrožavajući našu sposobnost razlikovanja činjenica od fikcije.⁸⁷ Ovaj „gubitak osjećaja za stvarnost“ može imati stvaran utjecaj na ljudsku autonomiju i sposobnost samostalnog djelovanja, kao i na slobodu misli. Iako kompanije i istraživači rade na smanjenju halucinacija,⁸⁸ između ostalog i kroz tehnička rješenja koja čine izvore informacija vidljivima, što je važno iz mnogo

83 Meta Oversight Board, *Content Moderation in a New Era for AI and Automation* web stranica. Vidi i OSCE/ODIHR *CHANGE: Capitalizing on the Human Dimension Mandate to Advance Gender Equality*, projekat.

84 Chowdhury and Lakshmi, *“Your opinion doesn’t matter, anyway”*.

85 Laura Bates, *Online brothels, sex robots, simulated rape: AI is ushering in a new age of violence against women*, *The Guardian*, 3.6.2025.

86 Kate Brennan et al, *Artificial Power: 2025 Landscape Report*, str. 49.

87 Colleen M Shannon, P. Eng., LL.M., *Do AI Hallucinations Disguise Gender Bias?*, *Medium*, 27.9.2023.

88 Reece Rogers, *Reduce AI Hallucinations with This Neat Software Trick*, *Wired*, 14 June 2024.

razloga, malo je vjerojatno da će se problem riješiti unutar postojećih modela i bez određenih kompromisa u performansama GenAI tehnologije.⁸⁹

Akt EU o umjetnoj inteligenciji predstavlja prvi pokušaj reguliranja korištenja GenAI-a, uključujući vrlo sofisticirane *deepfake* sadržaje, kako bi se pojedinci zaštitili od zloupotreba omogućenih umjetnom inteligencijom. Iako su neke grupe očiglednije pogođene ovim problemima, „informacijski poremećaj“ ima globalni utjecaj i predstavlja rizik ne samo po slobodu misli, već i po ljudska prava i sigurnost svih ljudi.⁹⁰

Utjecaj UI na ljudsku povezanost, razvoj misli i pravo na zaborav

Način na koji ljudi obrađuju i razumiju svijet, uključujući i kritičko razmišljanje, oblikovan je njihovim obrazovanjem, odgojem i uvjetima za razvoj mozga tokom formativnih godina. Umjetna inteligencija, a posebno GenAI, već je duboko utjecala na obrazovanje, a to će se i nastaviti, vjerojatno transformirajući podučavanje i učenje, i to tempom koji često ostavlja malo prostora za rasprave o rizicima i koristima ili donošenje pravilnih odluka nakon toga. Digitalna i medijska pismenost ključna je za sve, ali nastavni planovi i programi moraju se ažurirati kako bi učenici, mladi i stari, stekli znanja i vještine potrebne za razumijevanje, snalaženje i djelovanje u okruženju u kojem posreduje sve prisutnija umjetna inteligencija.

Postoje potencijalne koristi od upotrebe umjetne inteligencije u oblasti obrazovanja. Međutim, njeno korištenje, uključujući GenAI, u obrazovanju nosi i specifične rizike kojima se treba pažljivo pozabaviti. UNESCO je istaknuo potencijalne etičke izazove povezane s primjenom GenAI-a u obrazovnim okruženjima, pozivajući na sveobuhvatne procjene dugoročnog utjecaja.⁹¹ Glavni problem je potencijalno smanjenje interakcije među ljudima. Ta interakcija i ljudska povezanost ključne su za socio-emocionalne aspekte učenja, koji igraju važnu ulogu u kognitivnom razvoju i emocionalnom zdravlju.

Postoji rizik da iskustva učenja i stjecanje vještina razmišljanja budu zamijenjeni rješenjima koja nudi GenAI. Akademski istraživanja o korištenju GenAI-a

89 Kate Brennan et al, *Artificial Power: 2025 Landscape Report*, str. 49.

90 Vidi i Marina Nord et.al., *Media Freedom, Democracy, and Security*.

91 Miao and Holmes, *Guidance for generative AI in education and research*.

pokazuju zabrinjavajući pad sposobnosti kritičkog razmišljanja.⁹² Nedavna istraživanja stručnjaka iz ove oblasti slično su pokazala smanjenje vještina višeg nivoa razmišljanja među profesionalcima koji koriste GenAI modele za obavljanje svog posla.⁹³ Postoji i zabrinutost da bi GenAI mogao suziti izloženost mišljenjima različitih grupacija, jer obično odražava „standardne“ odgovore, dodatno marginalizirajući nedovoljno zastupljene grupe.⁹⁴

Da bi umjetna inteligencija imala pozitivan učinak u obrazovanju, integracija GenAI-a u obrazovne okvire zahtijeva jasno razumijevanje njenog utjecaja na učenje i stjecanje „vještina razmišljanja“, nakon čega slijede jasne smjernice i obuka za obrazovne radnike. Da bi bila korisna, mora biti i u službi obrazovnih ciljeva vezanih za razvoj sposobnosti razmišljanja i ljudski napredak. Na taj način, umjetna inteligencija npr. može poboljšati personalizirano učenje kroz prilagodljive platforme za učenje, omogućiti pristup obrazovnim resursima u nedovoljno opsluženim područjima i osloboditi nastavno osoblje suvišnih zadataka poput automatskog ocjenjivanja testova s višestrukim izborom odgovora, čime bi se dao prostor i prilika za jačanje slobode misli.

Sve više zabrinjava primjena chatbot-ova osmišljenih za simuliranje ljudskih odnosa radi druženja, uključujući životne partnere, u novoj generaciji umjetne inteligencije (GenAI). Mladi ljudi ili osobe s mentalnim zdravstvenim problemima posebno su izloženi riziku. Njihovo stanje se može pogoršati, a strašne posljedice nastaju kada chatbot-ovi potiču nezdrava ponašanja ili čak zagaravaju samoubistvo.⁹⁵ Posebno je alarmantno kada se umjetna inteligencija koristi za rješavanje problema mentalnog zdravlja. Razvoj odgovarajućih usluga mentalnog zdravlja temeljenih na umjetnoj inteligenciji je složen, a ogromni su rizici kod upotrebe chatbot-ova, uključujući i one koji su namijenjeni za nešto drugo (npr. zabavu ili marketing). Stručnjaci, uključujući udruženja psihologa, upozoravaju na ovakvu opasnost.⁹⁶ Potrebno je hitno regulirati ovo

92 Andrew R Chow, *ChatGPT's Impact on Our Brains According to an MIT Study*, Time, 23.6.2025.

93 *Despite plans to invest \$80 bn, Microsoft admits that AI is making us dumb*, Business Standard, 17.2.2025., izvor Parmy Olson, Bloomberg; *Amanda Silberling, Is AI making us dumb?*, TechCrunch, 10.2.2025.

94 Miao and Holmes, *Guidance for generative AI in education and research*.

95 Eileen Guo, *An AI chatbot told a user how to kill himself — but the company doesn't want to "censor" it*, MIT Technology Review, 6.2.2025.; Kate Payne, *In lawsuit over teen's death, judge rejects arguments that AI chatbots have free speech rights*, Associated Press, 21.5.2025.

96 Zara Abrams, *Using generic AI chatbots for mental health support: A dangerous trend*; American Psychological Association, 12. 3. 2025.; *APA calls for guardrails, education, to protect adolescent AI users*, American Psychological Association, Press Release, 3.6.2025.; Nathalie Koubayová, *Meet ChatPal, the European bot against loneliness*, Algorithm Watch, 22.5.2023.

pitanje, što bi obuhvatilo zabranu određenih aplikacija ili upotreba, kako bi se spriječila šteta i omogućio siguran razvoj tehnologija u ovoj oblasti.

Stepen do kojeg je osoba svjesno prisutna i aktivna u javnom prostoru značajno se mijenja u doba online prostora u kojem vlada UI. Urođeni dio naše slobode misli i svijesti jeste naš potencijal, kao ljudskih bića, da se stalno razvijamo i mijenjamo.⁹⁷ Međutim, sada je javno prisustvo na internetu trajno (npr. kroz objave na internetu), uključujući javno dostupne činjenice i digitalizaciju novinskih arhiva. Na primjer, misli koje je osoba nekad iznijela na internetu danas se mogu koristiti protiv nje. Može se tvrditi da to ograničava mogućnost i slobodu razvoja u vlastitim, kao i u očima drugih. Kao takvo, „pravo na postajanje“ koje je Charles Malik, autor UDHR-a, opisao kao bit slobode, pod rizikom je od zaustavljanja na samom izvoru.

„Pravo na zaborav“ počinje se prepoznavati u sudskoj praksi u odnosu na pravo na privatnost. Odnosi se na pravo na brisanje informacija o sebi koje bi inače ostale trajno dostupne u javnoj sferi, osim u određenim situacijama, kada javni interes prevladava u smislu pristupa određenim informacijama.⁹⁸

Međutim, digitalna prisutnost ljudi razvila se u ono što se naziva digitalnim otiskom. To je zapis svih online aktivnosti osobe, a ne samo ono što namjerno objavljuje da drugi vide. Obuhvata i web stranice koje posjećuje (zabilježene putem kolačića) i sve što one mogu reći o njoj. Ovi podaci omogućavaju sve sofisticiranije profiliranje, s opsegom i rezultatima neviđenih razmjera, što utječe na individualnu autonomiju. Osim osnovnog ciljanja korisnika prilagođenim sadržajem, ovakvo profiliranje moglo bi postati dio širih sistema nadzora te predstavljati prijetnju kibernetičkoj sigurnosti ili se koristiti za progon, npr. branitelja ljudskih prava, disidenata ili slobodnih mislilaca u autoritarnim režimima.⁹⁹

97 Vidi, na primjer, *Hannah Arendt, The Human Condition*, (University of Chicago Press: 1953).

98 „Izbor mjere koja će se provesti u specifičnim okolnostima bilo kog slučaja može varirati ovisno o faktorima kao što su tačnost ili netačnost informacije, stepen u kojem informacija doprinosi raspravi od javnog interesa, ima li bilo kakav istorijski, istraživački ili statistički interes, negativne posljedice po lični prostor pojedinca zbog kontinuirane dostupnosti informacije na internetu, kao i količina vremena koja je protekla od događaja na koje se odnosi članak ili vremena koje je proteklo od objave informacije“. *Joint Factsheet: The Right to be Forgotten. ECtHR and CJEU Case-Law*, European Court of Human Rights and the EU Agency for Fundamental Rights, 30.10.2024.

99 Vidi *Front Line Defenders webpage on Digital Protection*.

5. Neurotehnologija i sloboda misli

Neurotehnologija je još jedno područje koje je izuzetno bitno za slobodu misli, a posebno kada se prepliće s UI. Ovo poglavlje nudi pregled glavnih izazova koje ovo područje nameće slobodi misli.

U ključnom izvještaju iz 2024. godine,¹⁰⁰ Savjetodavni odbor Vijeća UN-a za ljudska prava opisuje „neurotehnologiju“ kao „niz uređaja i sistema koji su u međusobnoj interakciji sa centralnim nervnim sistemom putem električnih, magnetskih, optogenetskih i drugih sredstava“. Neki od ovih uređaja i sistema prvenstveno su namijenjeni razumijevanju načina rada mozga, dok drugi direktno interveniraju u mentalne procese kako bi obnovili izgubljene funkcije i poboljšali kognitivne sposobnosti.

Najnovija dostignuća u neurotehnologiji povećala su naše razumijevanje mozga i omogućila nove terapije za različita neurološka ili mentalna oboljenja. Velike istraživačke inicijative, koje finansiraju državni nivoi, pridonijele su takvom napretku.¹⁰¹ Ipak, brzi razvoj neurotehnologija u kombinaciji s umjetnom inteligencijom izaziva ozbiljnu zabrinutost za mentalnu privatnost, integritet i autonomiju.

U preporuci iz 2019. godine, Vijeće OECD-a prepoznalo je da „postoje etička, pravna i društvena pitanja koja se otvaraju kroz određene primjene neurotehnologija s obzirom na percipiranu centralnu ulogu mozga i kognitivne funkcije za pojmove ljudskog identiteta, slobode misli, autonomije, privatnosti i ljudskog prosperiteta“ te da je „potrebna široka javna rasprava o najboljem načinu primjene neurotehnologije u društvu“.¹⁰² U nacrtu teksta

¹⁰⁰ UN Human Rights Council Advisory Committee, *Impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights*, *A/HRC.57/61*, 8.8.2024., stav 4. Druga definicija neurotehnologija glasi: „... uređaji i postupci koji se koriste za pristup, praćenje, istraživanje, procjenu, manipuliranje i oponašanje strukture i funkcije nervnih sistema životinja ili ljudi“, *Preliminary study on the technical and legal aspects relating to the desirability of a standard-setting instrument on the ethics of neurotechnology*, UNESCO, 6.4.2023.

¹⁰¹ Na primjer, *the US Brain Research through Advancing Innovative Technologies (BRAIN) Initiative*, *the EU Human Brain Project*, ili *China Brain project*, između ostalog. Vidi i Hermann Garden, David E Winickoff, Nina Maria Frahm, Sebastian Pfotenhauer, *Responsible innovation in neurotechnology enterprises*, OECD Science, Technology and Industry Working Papers 2019/05, str. 15; ili *International Brain Initiative*.

¹⁰² *Recommendation of the Council on Responsible Innovation in Neurotechnology*, Organisation for Economic Co-operation and Development, OECD/LEGAL/0457, usvojene 11.12.2019.

preporuke o etici neurotehnologije,¹⁰³ UNESCO također razmatra etička, pravna i društvena pitanja i probleme koje sa sobom nosi primjena neurotehnologija u vezi s ljudskim pravima i ljudskim dostojanstvom, uključujući „autonomiju, privatnost, mentalni i fizički integritet, lični identitet, slobodu misli, rizik od diskriminacije, neravnopravnost i izazove po demokratiju (...).“ Nadalje, prepoznaje da osjetljivost intervencija u odnosu na vrlo složeni nervni sistem proizlazi iz njegove uloge u koordinaciji ponašanja i mentalnih procesa: „Omogućava ostvarivanje individualne autonomije, sposobnost djelovanja u svojstvu moralnih subjekata, odgovornost za postupke, saradnju s drugima, promišljanje o kolektivnim odlukama i razvoj ličnosti.“

Iako se neurotehnologija već neko vrijeme koristi u medicini, njeno spajanje s inženjerskim i kompjuterskim naukama, uz rast medicinskih i komercijalnih primjena, dramatično je promijenilo njen utjecaj, unaprijedivši medicinsku nauku, ali je donijelo i rizike po slobodu misli i druga ljudska prava.¹⁰⁴

„... neviđena sposobnost koju [neurotehnologije] nude vanjskim akterima da utječu na uživanje prava pojedinca otvara krucijalna etička pitanja i dovodi u pitanje samo razumijevanje temeljnih principa ljudskih prava“. Neurotehnologije su društveno razorne jer općenito: „(a) omogućavaju izloženost kognitivnih procesa, (b) omogućavaju direktnu promjenu mentalnih procesa i misli osobe, (c) zaobilaze svjesnu kontrolu ili svijest pojedinca, (d) omogućavaju neodobren vanjski pristup mislima, emocijama i mentalnim stanjima, (e) hrane se „neuropodacima“, koji su potrebni za njihovo vlastito funkcioniranje, kalibraciju i optimizaciju i (f) prikupljaju, analiziraju i obrađuju velike skupove ličnih podataka vrlo osjetljive prirode.“

— *UN Human Rights Council Advisory Committee, Impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights, A/HRC.57/61*, 8.8.2024.

Pored upotrebe u medicini, korištenje proizvoda temeljenih na neurotehnologiji putem aplikacija i uređaja, igara, praćenja zdravlja, pomagala za meditaciju ili čak ugrađenih u slušalice, omogućilo je tehnološkim kompanijama

¹⁰³ *Draft text of the Recommendation on the Ethics of Neurotechnology, UNESCO*, 9.4.2025.

¹⁰⁴ Savjetodavni odbor Vijeća za ljudska prava, *A/HRC.57/61*.

da dobiju pristup podacima o mozgu, što je imalo posljedice po slobodu misli.¹⁰⁵ UNESCO tvrdi da: „Kompanije mogu koristiti neuronske podatke dobivene iz neinvazivnih neurotehnoloških uređaja u marketinške svrhe. Otkrivanjem signala povezanih s našim sklonostima i izborima, te kompanije mogu utjecati na ponašanje kupaca radi maksimalnog povećanja profita. Time se otvaraju alarmantna pitanja o efektima nadzora, marketinških takтика i političkog utjecaja na naše najintimnije misli i emocije, što u konačnici ugrožava demokratije naših država i temelje društva u cjelini.“¹⁰⁶ Druge senzorske tehnologije mogu indirektno prikupljati podatke o našoj neuronskoj aktivnosti (npr. praćenje očiju, prepoznavanje i analiza glasa, prepoznavanje emocija lica, itd.) i postaju problem kada se koriste za donošenje zaključaka o mentalnim stanjima.¹⁰⁷

„Ovdje nije u pitanju samo zdravlje, nego naš pogled na ljudsku osobu, naše dostojanstvo i punu sposobnost ostvarivanja naših prava u kontekstu tenzije između zdravstvenih potreba i tržišnih ciljeva. S jedne strane, imamo velike zdravstvene potrebe, budući da na bolesti nervnog sistema, neurološke i mentalne bolesti otpada trećina naših zdravstvenih troškova (...). S druge strane, na potrošačkom tržištu „neuronski podaci“ (ili „podaci o mozgu“) postaju tražena vrsta podataka i robe izvan medicinskog sektora, uključujući digitalno fenotipiranje, afektivno računarstvo, neuroigre i neuromarketing. Među problemima koje neurotehnologije izazivaju treba spomenuti i javno povjerenje, poštovanje mentalne privatnosti, brzi tehnološki i ekonomski razvoj te činjenicu da se takav razvoj malo ili slabo nadzire“.

— Hervé Chneiweiss, *Ethics issues and global governance of neurotechnologies*, in *The risks and challenges of neurotechnologies for human rights*, UNESCO, University of Milan-Biocca, and State University of New York (SUNY) Downstate, 2023, str. 48.-49.

U spomenutom izvještaju, Savjetodavni odbor Vijeća UN-a za ljudska prava zaključuje kako neurotehnologije utječu na ljudska prava na jedinstven način. U izvještaju se razmatraju i opisuju glavni rizici po ljudska prava,

¹⁰⁵ Farahany, *The Battle for Your Brain*.

¹⁰⁶ *Ethics of neurotechnology*, UNESCO web stranica.

¹⁰⁷ UNESCO, *Draft text of the Recommendation on the Ethics of Neurotechnology*.

uključujući pravo na slobodu misli, koje donosi novo okruženje neurotehnologija i njihova (moguća) upotreba:

- Tehnologije koje su prvobitno razvijene kako bi pomogle osobama s neurološkim problemima sada se razvijaju i prodaju u komercijalne svrhe kao aplikacije za „kognitivno poboljšanje“ i druge, nemedicinske svrhe. Ove tehnologije bi korisnicima omogućile da mislima kontroliraju određene elemente vanjskog okruženja ili da komuniciraju s drugima koji imaju ugrađene slične tehnologije. Iako medicinska rješenja koja one pružaju znaju biti revolucionarna, nedostatak odgovarajućih etičkih okvira i okvira ljudskih prava kada su u pitanju nemedicinski uređaji izaziva značajnu zabrinutost, posebno u pogledu slobode misli i nediskriminacije, u što spada i osiguranje jednakog pristupa medicinskim uslugama. Zabrinutost zbog moguće ugroženosti prava djece postoji npr. ako se sučelja mozak-računalo (koja nisu testirana u smislu kratkoročnih ili dugoročnih posljedica po fizičko ili mentalno zdravlje) koriste u igrama ili tamo gdje neurotehnološki uređaji mogu poboljšati intelektualne sposobnosti.
- Čak i ako potpuno „čitanje misli“ još nije moguće, neurotehnologije unaprijeđene umjetnom inteligencijom sve su sposobnije za donošenje suptilnih zaključaka o mislima i mentalnim stanjima te omogućavaju profiliranje osoba na posebno invazivne načine. Otvaraju se specifična pitanja o privatnosti jer neurotehnologije mogu generirati detaljne zaključke o identitetima ljudi (uključujući osobine ličnosti, kognitivne performanse ili seksualnu orijentaciju). Dodatni rizici pojavljuju se ako se ovi podaci koriste u pravosudnim ili nacionalnim sigurnosnim sistemima, prilikom ispitivanja osumnjičenih osoba ili svjedoka, uključujući detektore laži koji su unaprijeđeni uz pomoć neurotehnologije. To bi moglo dovesti do kršenja prava osobe da ne svjedoči protiv sebe i rizika od kažnjavanja pojedinaca zbog njihovih misli. To uključuje situacije u kojima je tehnologija još uvijek previše neprecizna da bi mogla dati tačne podatke o određenim mislima. Otkrivanje neizrečenih misli bez pristanka „vlasnika“ u suprotnosti je sa suštinom slobode misli.¹⁰⁸
- Uređaji za praćenje mentalnog stanja već su u upotrebi u radnim sredinama, posebno u ekstremnim uvjetima, kako bi se osigurala budnost

108 Savjetodavni odbor Vijeća za ljudska prava, [A/HRC.57/61](#).

i izbjegle nesreće zbog umora.¹⁰⁹ Iako su ulozi možda visoki, njihovo korištenje otvara ozbiljna pitanja o ljudskim pravima, uključujući prava radnika i slobodu misli. To bi moglo otvoriti vrata neuronadzoru u svrhu produktivnosti.¹¹⁰ EU je kroz svoj Akt o umjetnoj inteligenciji poduzela korake ka rješavanju ovog problema u specifičnim okruženjima.¹¹¹ Općenito, s obzirom na njihove nepoznate dugoročne posljedice, posebno ako se primjenjuju bez jasnog etičkog nadzora i praćenja stepena poštivanja ljudskih prava, primjena ovih tehnologija od strane korisnika predstavlja rizik kako za lični integritet tako i za (mentalno) zdravlje.

- Ranjive kategorije, poput starijih osoba, djece, osoba s invaliditetom i osoba lišenih slobode ili u pritvorskim okruženjima posebno su izložene riziku sve dok se ne osigura odgovarajuća regulacija razvoja i korištenja neurotehnologija. Postoje posebni problemi u vezi s pristankom, koji uvijek treba biti prethodno dat, slobodan, na osnovu odgovarajućih informacija, stvaran, transparentan, djelotvoran i nikad zasnovan na pretpostavkama. S obzirom na medicinske svrhe za koje se mnoge od ovih tehnologija razvijaju, osobe s invaliditetom i njihovi predstavnici bi npr. trebali biti uključeni u proces razvoja, a njihove potrebe, prava i perspektive kao krajnjih korisnika trebaju biti prioritet. Poštovanje prava osoba s invaliditetom uključuje i osiguranje njihovog potpunog pristupa neurotehnologijama nakon njihovog razvoja, te da one budu sigurne, djelotvorne i da poštuju ljudska prava u svom dizajnu, razvoju i korištenju.¹¹² Ovaj pristup usklađen je i sa socijalnim modelom Konvencije o pravima osoba s invaliditetom, putem kojeg se invaliditet posmatra kao dio ljudske raznolikosti i društva, za razliku od ableističkog medicinskog modela s tradicionalnim fokusom na prevenciju i liječenje.¹¹³

109 Ovo obuhvata područja poput rudarstva, izgradnje, kamionskog prevoza, avioprevoza, željeznica. Farahany, *The Battle for Your Brain*. See also José M. Muñoz, Laura Isaza, and Tarini Mehta, *Tech is coming for your brain data: how a Chilean politician turbocharged the "neurorights" movement*, *The Boston Globe*, 10.9.2024.

110 Savjetodavni odbor Vijeća za ljudska prava, *A/HRC.57/61*.

111 *High-level summary of the AI Act*, EU Artificial Intelligence Act, 27.2.2024., ažurirano 30.5.2024. Prema ovom Aktu, zaključivanje o emocijama na radnom mjestu ili u obrazovnim institucijama, osim u medicinske ili sigurnosne svrhe, spada u zabranjene UI sisteme. Vidi i Nora Santalu, *Neurotechnologies under the EU AI Act: Where law meets science*, IAPP, 12.5.2025.

112 Savjetodavni odbor Vijeća za ljudska prava, *A/HRC.57/61*.

113 *International Disability Alliance, Submission to the Human Rights Committee Advisory Committee call for inputs on neurotechnology and human rights*, 2.7.2023., vidi i OHCHR *Neurotechnology and human rights* web stranica.

Propisima bi se trebalo ciljati na zaštitu ljudskih prava definiranjem jasnih okvira za komercijalnu upotrebu koji su u skladu s ljudskim pravima, gdje bi, istovremeno, postizanje medicinskih otkrića trebalo biti prioritet.

6. Regulatorni okviri

Regulacija umjetne inteligencije tokom cijelog njenog životnog ciklusa zaostaje za jednim od najbrže rastućih, najsloženijih i najutjecajnijih područja u historiji čovječanstva. Iako postoje razni neobavezujući dokumenti na regionalnom ili globalnom nivou, ne postoji jedinstveni, univerzalni, pravno obavezujući standard o umjetnoj inteligenciji koji bi štitiio ljudska prava.¹¹⁴

Usvajanjem Preporuke o etici UI 2021. godine, na Općoj konferenciji UNESCO-a prepoznati su „duboki i dinamični pozitivni i negativni utjecaji umjetne inteligencije (UI) na društva, okoliš, ekosisteme i ljudske živote, uključujući ljudski um, dijelom zbog novih načina na koje njena upotreba utječe na ljudsko razmišljanje, interakciju i donošenje odluka te utječe na obrazovanje, humanističke, društvene i prirodne nauke, kulturu, komunikaciju i informacije.“ Preporuka prvenstveno ima za cilj zaštititi ljudska prava i dostojanstvo, na osnovu niza principa kao što su transparentnost, pravednost i ljudski nadzor. U preambuli ove Preporuke upozorava se i na različite asimetrije moći oko UI, koje stvaraju različite i ozbiljne rizike i potencijalne posljedice, te je izražena potreba za jačom globalnom saradnjom i solidarnošću, uključujući multilateralni pristup.

— *Preporuka o etici umjetne inteligencije*, UNESCO, 23.11.2024.

Reguliranjem ovog područja putem Akta o UI, EU je ostvarila značajan napredak, između ostalog i u zaštiti ljudskih prava. Budući da je tek nedavno usvojen i potrebne su daljnje smjernice i razvoj instrumenata, kao i okviri za provedbu na državnom nivou, njegova sposobnost da ublaži utjecaje na ljudska prava još uvijek nije potvrđena u praksi. Vijeće Evrope je 2024. godine usvojilo i Okvirnu konvenciju o umjetnoj inteligenciji, ljudskim pravima, demokratiji i vladavini prava, koja svojim okvirom obuhvata veći broj država.¹¹⁵

¹¹⁴ *Artificial Intelligence*, UNESCO web stranica; UNESCO, *Draft text of the Recommendation on the Ethics of Neurotechnology*; *Artificial intelligence*, OECD web stranica; *Governing AI for Humanity: Final Report*, Savjetodavno tijelo UN-a o UI, septembar 2024.; *UN Global Digital Compact*.

¹¹⁵ Mogu je potpisati i države koje nisu članice Vijeća Evrope koje su pomogle u njenoj izradi, a postoje posebne, odvojene procedure koje omogućavaju takvim državama da se pridruže. Od januara 2025. godine, potpisale su je Sjedinjene Države i Izrael, kao države nečlanice VE. Kako bi postala obavezujuća, moraju je i ratificirati.

Iako su izražene i zabrinutosti u pogledu opsega i djelotvornosti zaštite oba ova instrumenta,¹¹⁶ važno je da se dokumenti poput Okvirne konvencije ratificiraju, u potpunosti provedu i ojačaju gdje je to potrebno. To zahtijeva resurse i volju pojedinačnih država i, u slučaju Akta EU o umjetnoj inteligenciji, institucija EU. Iako nisu pravno obavezujući i ne mogu zamijeniti regulatorne okvire, industrijski standardi također doprinose uspostavljanju etičkih okvira.¹¹⁷ ODIHR ukazuje i na rastući broj akademskih istraživanja i inicijativa koje se posebno bave potrebom za globalnim regulatornim okvirima i doprinose raspravama o ovoj temi.¹¹⁸

Akt EU o umjetnoj inteligenciji predstavlja do sada najsveobuhvatniji pokušaj uspostavljanja regulatornog okvira za razvoj i korištenje umjetne inteligencije, znatno napredniji u odnosu na druge regulatorne okvire. Definira pristup zasnovan na procjeni rizika, pri čemu sisteme umjetne inteligencije razvrstava u kategorije: zabranjeni, visokorizični, ograničenog rizika i minimalnog rizika. Regulira umjetnu inteligenciju opće namjene, odnosno sisteme zasnovane na temeljnim modelima, poput GenAI-a, te nameće niz obaveza, uključujući transparentnost i odgovornost. Definira i umjetnu inteligenciju opće namjene koja predstavlja sistemske rizike (tj. moćne modele), za koje propisuje dodatne zahtjeve te pojašnjava da se takvi sistemi mogu koristiti kao visokorizični sistemi umjetne inteligencije ili biti integrirani u njih, što sa sobom nosi posebne obaveze saradnje radi osiguravanja njihove usklađenosti s Aktom. Važno je napomenuti da modeli umjetne inteligencije opće namjene dostupni pod slobodnim i otvorenim licencama podliježu manjem broju obaveza prema Aktu EU o umjetnoj inteligenciji.¹¹⁹ Okviri zasnovani na procjeni rizika poput Akta EU o umjetnoj inteligenciji mogu poslužiti kao model i za druge regije.

116 Vidi *EU: Artificial Intelligence rulebook fails to stop proliferation of abusive technologies*, Amnesty International, 13.3.2024. i *ENNHRI Calls on Council of Europe member States to ensure strong human rights protection in the draft Convention on AI, Human Rights, Democracy and Rule of Law*, ENNHRI, 17.5.2024. Radi sveobuhvatne analize Akta i drugih sličnih zakona, vidi Sandra Wachter, *Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This means for the European Union, the United States, and Beyond*, *Yale Journal of Law and Technology*, tom 26., izdanje 3., str. 671.-718. Radi diskusije o specifičnim neurotehnologijama i nedostacima Akta EU, vidi Christoph Bublitz, *Banning biometric mind reading: the case for criminalizing mind probing*, *Law, Innovation and Technology*, tom 16., izdanje 2., str. 432.-462.

117 Vidi, *IEEE Frameworks*, Institute of Electrical and Electronics Engineers (IEEE) Tech Ethics.

118 Vidi, Alexander Kriebitz, Caitlin C Corrigan (eds.), *Promoting and Advancing Human Rights in Global AI Ecosystems: The Need for A Comprehensive Framework under International Law*, februar, 2025.

119 *High-level summary of the AI Act*, EU Artificial Intelligence Act, 27.2.2024., ažuriran 30.5.2024.

Zabranjeni sistemi umjetne inteligencije uključuju one koji: koriste subliminalne, manipulativne ili obmanjujuće tehnike s ciljem promjene ponašanja i narušavanja informiranog donošenja odluka, pri čemu uzrokuju značajnu štetu; iskorištavaju ranjivosti pojedinaca (npr. dob, invaliditet, socioekonomske okolnosti) kako bi utjecali na njihovo ponašanje na način koji može uzrokovati značajnu štetu; biometrijske sisteme kategorizacije (uz izuzetak organa za provedbu zakona), provode društveno vrednovanje zasnovano na društvenom ponašanju ili osobinama ličnosti; procjenjuju rizik da će pojedinac počiniti kazneno djelo isključivo na osnovu profiliranja ili osobina ličnosti, osim kad se koriste kao podrška ljudskoj procjeni zasnovanoj na objektivnim i provjerljivim činjenicama direktno povezanim s kriminalnom aktivnošću; kreiraju baze podataka za prepoznavanje lica putem neselektivnog prikupljanja slika lica s interneta ili sistema videonadzora; zaključuju o emocijama osoba na radnim mjestima ili u obrazovnim ustanovama, osim u slučajevima zdravstvenih ili sigurnosnih razloga; vrše daljinsku biometrijsku identifikaciju u stvarnom vremenu i na javnim mjestima za potrebe provedbe zakona, uz određene izuzetke koji zahtijevaju, između ostalog, procjenu utjecaja na osnovna prava.

Pružaoци visokorizičnih sistema UI podliježu specifičnim zahtjevima koji se, između ostalog, odnose na upravljanje rizicima, kvalitet podataka, transparentnost, odgovornost i ljudski nadzor. Prije stavljanja u upotrebu, uz određene izuzetke i pod propisanim uvjetima, takvi sistemi moraju proći i procjenu utjecaja na osnovna prava. Sistemi umjetne inteligencije koji profiliraju pojedince uvijek se smatraju visokorizičnim.

— Isječak iz *High-level summary of the AI Act*, EU Artificial Intelligence Act web stranica, 27.2.2024., ažuriran 30.5.2024.

Prije Akta o umjetnoj inteligenciji, EU je usvojila Akt o digitalnim uslugama (DSA), koji regulira online platforme i posredničke usluge, sa specifičnim pravilima za vrlo velike platforme i pretraživače. Fokus DSA je i na rješavanju sistemskih rizika, uključujući rizike povezane sa nezakonitim sadržajem, slobodom izražavanja, slobodom i pluralizmom medija, diskriminacijom, zaštitom potrošača i pravima djece, javnom sigurnošću i izbornim procesima, rodno zasnovanim nasiljem te mentalnim i fizičkim blagostanjem. Također zahtijeva određeni nivo transparentnosti i nadzora od strane nadležnih

tijela, kao i mogućnost korištenja sistema preporuka koji se ne temelje na profiliranju korisnika.¹²⁰

Ovaj regulatorni okvir EU-a djeluje u izuzetno složenoj oblasti, a njegovi efekti tek postaju vidljivi i razumljivi. Potrebni instrumenti i institucionalni okviri, posebno na državnom nivou, još se uvijek razvijaju i uspostavljaju u vrijeme pisanja ovog teksta.¹²¹ Nadalje, napredovanje ka osiguranju zaštite ljudskih prava u vezi s umjetnom inteligencijom kroz propise nije završeno.¹²² Iako važan, okvir EU-a ne odnosi se na sve države članice OSCE-a, a kamoli na čitav svijet, zbog čega postoji potreba za globalnim regulatornim pristupom.¹²³ I Akt EU o umjetnoj inteligenciji i Okvirna konvencija Vijeća Evrope sadrže izuzetke kad je u pitanju nacionalna sigurnost. S obzirom na to da je EU najnapredniji regulator umjetne inteligencije na svijetu, ovi izuzeci izazivaju značajne zabrinutosti u vezi s ljudskim pravima, te potencijalno mogu doprinijeti novoj globalnoj utrci u naoružanju, gdje se nadmoć u dometima umjetne inteligencije smatra ključem dominacije.¹²⁴ Kako bi se takvi rizici ublažili, međunarodna saradnja je ključna za razvoj usklađenih standarda koji bi zatvorili postojeće praznine i spriječili zloupotrebe, te istovremeno podstakli odgovorne inovacije.

Iako razvoj proizvoda u oblasti medicine podliježe određenim propisima, neurotehnologije koje se razvijaju u komercijalne svrhe nose nove rizike po ljudska prava i osnovne slobode. Zato su potrebni jači okviri zaštite. Uzimajući u obzir da postoje određeni, ali općenito ograničeni, elementi regulacije na nacionalnom ili regionalnom¹²⁵ nivou ili unutar Akta EU o umjetnoj inteligenciji, ne postoje univerzalni, međunarodni, pravno obavezujući standardi o neurotehnologijama. U okviru UNESCO-a, države članice

120 Evropska komisija, *DSA: Very large online platforms and search engines* (web stranica ažurirana 12.2.2025.) i evropska komisija, *The Digital Services Act package* (web stranica ažurirana 12.2.2025.).

121 Vidi *Towards meaningful fundamental rights impact assessments under the DSA*, European Center for Not-for-Profit Law and Access Now, 15.9.2023. i Aninda Chakraborty, *LatticeFlow AI unveils EU compliance framework for Generative AI*, Tech Monitor, 17.10.2024. *Overview of all AI Act National Implementation Plans*, EU Artificial Intelligence Act web stranica, 8.11.2024. (ažurirana 19.5.2025.).

122 *Withdrawal of the AI Liability Directive Proposal Raises Concerns Over Justice for AI Victims*, Center for Democracy and Technology, 12.2.2025.

123 Neke države nečlanice OSCE-a usvojile su razne regulatorne okvire, ali nisu predmetom ove analize. EU je i dalje najnapredniji regulator na ovom polju.

124 Ilaria Carroza, Nicholas Marsh and Gregory M. Reichberg, *Dual-Use AI Technology in China, the US and the EU: Strategic Implications for the Balance of Power*. PRIO Paper, (Oslo: PRIO, 2022.).

125 Vidi, na primjer, OECD, *Recommendation of the Council on Responsible Innovation in Neurotechnology*. OECD/Legal/0457, usvojene 11.12.2019.

trenutno pregovaraju o prvom, neobavezujućem, globalnom instrumentu za postavljanje standarda o etici neurotehnologije, koji proizlazi iz ranijeg rada o etici umjetne inteligencije.¹²⁶ Osim toga, potrebna je sveobuhvatna regulacija koja će obuhvatiti punu složenost novih izazova koje predstavljaju neurotehnologije, kako kroz postojeće okvire zaštite koji bi se mogli proširiti i obuhvatiti specifične nove rizike po ljudska prava, tako i kroz nove instrumente o razvoju i korištenju neurotehnologija.

U izvještaju iz 2024. godine,¹²⁷ Specijalni izvjestilac UN-a za pravo na privatnost predložio je tekst rezolucije kojom se ažurira Rezolucija Generalne skupštine iz 1990. godine, br. 45/95, pod nazivom „Smjernice za regulaciju kompjuterizovanih baza podataka“. Predloženi tekst navodi da se obrada neuropodataka „ne smije koristiti za manipuliranje ili mijenjanje slobode misli i svijesti pojedinca, čineći ga ovisnim o trećoj strani ili mijenjanje njegove ideje, sigurnosti ili neovisnosti ili njegovog prirodnog moždanog identiteta ili neurokognitivnog integriteta“. Nadalje, takvi podaci ne smiju se „obrađivati u svrhe koje nisu promocija zdravlja, dijagnostika, rehabilitacija i ublažavanje bolesti u kontekstu prava na zdravlje i naučnih istraživanja na polju biologije, psihologije i medicine čiji je cilj ublažavanje patnje i poboljšanje zdravstvenog stanja“. U aprilu 2025. godine, putem Rezolucije 58/6, Vijeće UN-a za ljudska prava prepoznalo je da „kontinuirani razvoj nekih od njenih aplikacija može pokrenuti izvjestan broj etičkih, pravnih i društvenih pitanja s implikacijama po ljudsko dostojanstvo i autonomiju, zbog čega je neophodno da se ljudska prava efektivno poštuju, štite i ispunjavaju u ovom kontekstu“. Vijeće je zatražilo od Savjetodavnog odbora da izradi niz smjernica za primjenu postojećeg okvira za ljudska prava na koncept, dizajn, razvoj, testiranje i korištenje neurotehnologija.¹²⁸

I pored složenosti i implikacija koje su još uvijek nejasne, razumijevanje ljudi o tome kako društvene mreže koriste njihove podatke brzo se poboljšalo posljednjih godina. Nasuprot tome, njihovo razumijevanje implikacija po ljudska prava u smislu otkrivanja njihovih neuropodataka je još na izrazito

126 *The Ethics of Neurotechnology: UNESCO appoints international expert group to prepare a new global standard*, Izdanje za štampu UNESCO-a, 22.4.2024.

127 *A/79/173*: Izvještaj Specijalnog izvjestioca za pravo na privatnost, Ane Brian Nougères – Prijedlog ažuriranja Rezolucije Generalne skupštine br. 45/95 od 14.12.1990., pod nazivom „Smjernice za regulaciju kompjuterskih datoteka ličnih podataka“, OHCHR, 17.7.2024.

128 *Human rights guidelines on neurotechnology*, web stranica Vijeća UN-a za ljudska prava.

niskom nivou, posebno s obzirom na nepostojanje odgovarajućeg regulatornog okvira.¹²⁹

„Revolucija umjetne inteligencije“ dobila je ime po načinu na koji i dalje transformira ljudski život. Ali, ova se revolucija uglavnom dogodila bez ikakve javne rasprave ili doprinosa u donošenju odluka. Imati potrošače na digitalnom tržištu, koji su samo djelomično svjesni svojih izbora i nad njima imaju ograničenu kontrolu nije isto što i voditi demokratsku raspravu. Isto tako, brzi razvoj neurotehnologija, posebno komodifikacija uređaja koji komuniciraju s ljudskim mozgom ili neuronskim krugovima, utječu na ljudska prava na neviđen i snažan način. Države članice OSCE-a dužne su ponovo uspostaviti uvjete potrebne za javnu raspravu o pitanjima od ključne društvene važnosti. Pitanja poput „kognitivnog poboljšanja“ zdravih osoba, na primjer, ili možda čak „povećanja ljudskih sposobnosti“, imaju tako duboke implikacije po čovječanstvo da zaslužuju istinsku i temeljitu javnu raspravu kako bi se mogle donijeti ispravne odluke o tome kako regulirati takve tehnologije. Rasprave i donošenje odluka moraju uključivati širi krug ljudi, a ne samo malobrojne koji učestvuju u razvoju tehnologije ili nekolicinu uključenih u istraživanje koji mogu tvrditi da imaju odgovarajuće mišljenje o toj temi.¹³⁰ Čak i ako su teme složene, njihove implikacije po društvo treba pažljivo sažeti u jasno razumljive informacije i podijeliti s javnošću radi istinske demokratske rasprave o tome treba li koristiti umjetnu inteligenciju ili ne.

„Čitava pompa oko umjetne inteligencije (...) isisala je zrak iz ionako zagušljive prostorije, da se čini uzaludnim, ponekad nemogućim, zamisliti išta drugo osim stalnog marširanja ka neizbježnoj nadmoći umjetne inteligencije. Ali bez obzira koliko se to činilo istinito, to je samo osjećaj. Nije stvarnost, barem ne još uvijek. Zapravo, postoje mnoge alternative ovoj verziji umjetne inteligencije, mnogi načini oblikovanja novih svjetova. Međutim, kao i umjetna inteligencija, ni one nisu neizbježne. Uvođenje tih alternativa počinje postavljanjem i odgovaranjem na jedno pitanje: Je li ovo svijet koji želimo?“

— Kate Brennan, Amba Kak and Dr. Sarah Myers West, *Artificial Power: 2025 Landscape Report*, AI Now Institute, 3. 6. 2025., str. 9. i 12.

¹²⁹ Farahany, *The Battle for Your Brain*.

¹³⁰ Vidi i OECD, *Recommendation of the Council on Responsible Innovation in Neurotechnology*, Recommendation No. 5: Enable societal deliberation on neurotechnology.

7. Zaključci i preporuke

Potreba za zaštitom prava na slobodu misli postaje sve hitnija, s obzirom na mjeru u kojoj tehnologije utemeljene na UI oblikuju i društva i svijest pojedinaca o sebi. Trenutni okvir za razvoj i primjenu tih tehnologija produbljuje neravnotežu moći i predstavlja rizik po demokratiju. Misli i donošenje odluka moraju biti zaštićeni od novih oblika suptilne, ali djelotvorne manipulacije, dok neizrečene misli moraju ostati privatne. U nedostatku odgovarajućih regulatornih okvira za zaštitu ljudskih prava, širenje tehnologija temeljenih na UI čini se da zamagljuje, ako ne i briše, konture demokratske i javne odgovornosti. Države su dužne poštivati i štititi ljudska prava, što zahtijeva posebnu regulaciju na državnom i međunarodnom nivou. Države možda kasne s izradom propisa, ali još uvijek ima vremena za to, posebno za potrebe razvoja i korištenja neurotehnologija za ublažavanje rizika od štete širokih razmjera koju mogu prouzročiti.

Nadovezujući se na niz neiscrpnih kriterija koje je predložio Specijalni izvjestilac UN-a za slobodu vjeroispovijesti ili uvjerenja kako bi se, od slučaja do slučaja, procijenilo šta može predstavljati nedopuštenu manipulaciju mislima, Susie Alegre i Aaron Schull opisuju skup faktora potrebnih za pravni test procjene tehnološki omogućene, nezakonite manipulacije. Napominju da se pravo na slobodu misli, kao dio općeg okvira UN-a za ljudska prava, „razvilo kao odgovor na zločine koje je počinila nacistička Njemačka uz prepoznavanje neviđenih rizika koje nosi ispiranje mozga stanovništva po ljudska prava u širem smislu“. Granice koje okružuju zaštitu prava na slobodu misli ključne su za budućnost demokratije i ljudskih prava širom svijeta.” Dalje predlažu da se „na osnovu Općeg komentara Vijeća za ljudska prava utvrdi novi pravni test kojim se uzima u obzir niz dodatnih faktora:

- Obim i razmjere primjene prakse;
- Da li su građani općenito ili relevantni pojedinci svjesni taktike utjecanja;
- Da li šira javnost ili relevantni pojedinci mogu razumjeti utjecaj određene prakse;

- Da li je praksa osmišljena kako bi zaobišla moć rasuđivanja;
- Intenzitet i vremenski period izloženosti;
- Ciljanje na posebne kognitivne predrasude ili ranjivosti (posebno na osnovu „insajderskog znanja“ dobivenog iz podataka);
- Bilo kakva neravnoteža moći; i
- Praktična mogućnost da se kaže – „ne“.

— *Susie Alegre and Aaron Schull, **Freedom of Thought: Reviving and Pro-protecting a Forgotten Human Rights**, Center for International Governance Innovation, 2024., str. 12.*

U svjetlu argumenata iznesenih u ovom dokumentu i radi stvaranja poticajnog okruženja za slobodu misli, državama članicama OSCE-a preporučuje se da:

- **Usvoje politike, zakone i propise zasnovane na ljudskim pravima na državnom nivou i uspostave institucionalnu infrastrukturu** kako bi se osiguralo poštivanje, zaštita i ispunjenje prava na slobodu misli i svih ljudskih prava i osnovnih sloboda, u kontekstu rastuće uloge tehnologija utemeljenih na UI i onih koje koriste UI, uključujući neuro-tehnologije. Osim toga, kako bi se osiguralo poštivanje ljudskog dostojanstva i demokratskih principa, ključno je regulirati neurotehnologije isključivo na temelju etičkih i ljudskih prava, uz izbjegavanje poslovnog modela društvenih mreža i zauzimanje pristupa zasnovanih na provjeri rizika, te uz osiguranje zaštite ljudskih prava tokom cijelog životnog ciklusa UI kojeg razvija industrija.
- **Usvoje metodologije procjene utjecaja na ljudska prava** kao standardnu praksu, gradeći pravni i institucionalni okvir potreban za osiguranje široke primjene takvih metodologija. Ovi alati moraju dati prioritet poštovanju, zaštiti i ostvarivanju ljudskih prava tokom cijelog životnog ciklusa UI te se, kroz redovno preispitivanje, prilagođavati brzom tempu razvoja UI i novim izazovima u oblasti ljudskih prava koje ona donosi.
- **Razviju alate za ugradnju slobode misli u procjene utjecaja na ljudska prava** za sve tehnologije utemeljene na UI, kako u smislu zasebnog prava tako i u međusobnoj ovisnosti s drugim pravima, posebno slobodom izražavanja i pravom na privatni život.

- **Prikupljaju podatke o UI tokom cijelog njenog životnog ciklusa, prateći usklađenost sa standardima ljudskih prava (uključujući slobodu misli), moguće dobre prakse i načine ostvarivanja koristi od umjetne inteligencije, uz procjenu pristupačnosti i ravnopravnosti pristupa.** S obzirom na brzo usvajanje novih tehnologija utemeljenih na UI u različitim područjima, svijet zaostaje ne samo u regulaciji, već i u djelotvornom praćenju i mapiranju potrebnom za razumijevanje posljedica po ljudska prava, predviđanje implikacija i prepoznavanje novih prilika. Države su dužne pratiti i osigurati poštivanje ljudskih prava. Posebna se pažnja posvećuje najranjivijima, uključujući osobe s invaliditetom, djecu, starije ili osobe lišene slobode.
- **Potiču i podržavaju nova istraživanja i konceptualna pojašnjenja o standardima** promocije i zaštite slobode misli, posebno u kontekstu novih tehnologija utemeljenih na UI. To podrazumijeva međunarodnu saradnju u okviru UN-a (npr. donošenjem rezolucija Vijeća za ljudska prava), kao i kroz aktiviranje mehanizama UN-a za ljudska prava (npr. traženjem općih komentara ili izvještaja) ili putem drugih sličnih sredstava.
- **Sarađuju s obrazovnim sistemima i pripremaju ih** da uzmu u obzir nove načine društvenog života i pomognu učenicima da razumiju i djeluju kao kritički mislioci, autonomno i djelotvorno, kako bi mogli sačuvati svoju slobodu misli u svijetu koji sve više pokreće umjetna inteligencija. To će vjerojatno zahtijevati više od onoga što se trenutno radi na digitalnom opismenjavanju. Novi državni obrazovni planovi i programi trebali bi obuhvatiti sadržaj, vještine i ponašanja potrebna za pripremu sadašnjih i budućih generacija za ove promjene, koje utječu na sva područja života. Trebalo bi provoditi i kampanje za obuke odraslih osoba i širenje svijesti.
- **Sarađuju sa svim obrazovnim i edukacijskim ustanovama kako bi se osigurala edukacija o etici i ljudskim pravima za one koji rade ili odlučuju o korištenju umjetne inteligencije tokom cijelog njenog životnog ciklusa, kako bi se osiguralo da bude razvijena na način da poštuje i, u idealnom slučaju, jača zaštitu i poštivanje ljudskih prava.** Obuka tokom studiranja (npr. na univerzitetu) i kontinuirana obuka za one koji već rade trebale bi osigurati da oni koji razvijaju, uvode ili koriste umjetnu inteligenciju i nove tehnologije (od stručnjaka za podatke do inženjera i drugih) budu dovoljno obučeni o etičkim implikacijama tehnologije po ljudska prava, uključujući i one koje se odnose

na slobodu misli. Stručnjaci koji rade s umjetnom inteligencijom tokom cijelog njenog životnog ciklusa (uključujući one koji donose odluke o nabavci i primjeni) trebali bi znati kako provoditi procjene utjecaja na ljudska prava i ugraditi zaštitu ljudskih prava u tehnologije umjetne inteligencije već pri samom dizajnu.

- **Štite neovisne medije i jačaju njihovu ulogu u javnom interesu na polju zaštite slobode misli, slobode mišljenja i slobode izražavanja, uključujući i putem provedbe** preporuka Predstavnik OSCE-a za slobodu medija o ovoj vrlo važnoj temi.¹³¹
- **Potiču istinsku javnu raspravu o etičkim implikacijama UI i neurotehnologija i njihovim implikacijama po ljudska prava, uključujući i po slobodu misli.** Te bi rasprave trebale biti strukturirane i multidisciplinarne i uključivati različite aktere, posebno one čija bi prava mogla biti pogođena, ili krajnje korisnike tehnologija. Rasprave bi trebale uključivati medije od javnog interesa kako bi se širila svijest javnosti o tim implikacijama i osiguralo da se informacije sažimaju i omogućavaju adekvatne javne rasprave. Trebale bi biti transparentne i javno odgovorne, a rezultati bi se trebali odražavati u političkim odlukama.
- **Sarađuju s neovisnim institucijama OSCE-a, kao što su ODIHR i Ured predstavnika za slobodu medija, kako bi se osigurala regulacija i primjena životnog ciklusa UI koji poštuje ljudska prava.**

¹³¹ Za više informacija, vidi *OSCE RFoM Spotlight on Artificial Intelligence and Freedom of Expression — SAIFE Resource Hub* and the *Policy manual* in particular as well as upcoming work on *Media and Big Tech*.

