

ALGORITAMSKA DISKRIMINACIJA****

Apstrakt

Algoritamska diskriminacija može nastati radom sistema veštačke inteligencije i u javnom i u privatnom sektoru. U javnom sektoru to su, primera radi, preporuke za odmeravanje kazni za krivična dela, donošenje odluka o penzijskim i socijalnim davanjima ili naknadama za nezaposlene itd. U privatnom sektoru to su odluke koje se donose u vezi sa zapošljavanjem, stanovanjem, odobravanjem kredita, reklamiranjem proizvoda itd. Postojeći pravni propisi o nediskriminaciji i zaštiti podataka, zaštiti potrošača i tome slično pružaju određene mogućnosti za borbu protiv algoritamske diskriminacije. Međutim, neophodno je da se postojeći propisi unaprede i da se donesu novi opšti i sektorski propisi, kojima će se regulisati upotreba sistema veštačke inteligencije zasnovanih na algoritamskom odlučivanju, na bezbedan način i uz visok nivo zaštite pojedinačnih prava i sloboda. Jedno od dobrih rešenja za otklanjanje algoritamske diskriminacije jeste da se pravnim propisima zahteva ispunjavanje određenih uslova i davanje sertifikata za sisteme veštačke inteligencije koji su visokorizični, odnosno koji mogu ugroziti ljudska prava i slobode, kao što je to predviđeno Predlogom uredbe EU o veštačkoj inteligenciji 2021. godine.

Ključne reči: pravo, pravna informatika, ljudska prava, veštačka inteligencija, algoritamska diskriminacija.

1. Uvod

Svakodnevni život sve više zavisi od upotrebe veštačke inteligencije. Veštačka inteligencija upravlja saobraćajem, snabdevanjem energijom, prepoznaje govor, filtrira neželjenu elektronsku poštu, analizira rendgenske snimke i na mnoge druge načine utiče na naš svakodnevni život i na ekonomski razvoj celokupnog društva. U svim tim procesima veštačka inteligencija funkcioniše tako što softver na bazi algoritma donosi odluke, koje najčešće obuhvataju veliki broj pojedinaca.

* Naučni saradnik, Institut za uporedno pravo, Beograd, d.prlja@iup.rs

** Naučni saradnik, Institut za uporedno pravo, Beograd, g.gasmi@iup.rs

*** Naučni saradnik, Matematički institut SANU, Beograd, vanja@mi.sanu.ac.rs

**** Rad je nastao kao rezultat naučnoistraživačkog rada Instituta za uporedno pravo koji finansira Ministarstvo nauke, tehnološkog razvoja i inovacija Republike Srbije prema Ugovoru o realizaciji i finansiranju naučnoistraživačkog rada NIO u 2023. godini (evidencioni broj: 451-03-47/2023-01/200049 od 3. 2. 2023).

Te odluke bi trebalo da budu racionalne, neutralne, nepristrasne i jednake za sve one na koje se odnose. Praksa pokazuje da odluke donete u sistemima veštačke inteligencije na bazi algoritma često izazivaju neki oblik diskriminacije, koja se naziva algoritamska diskriminacija.

Veštačka inteligencija je pojam koji danas obuhvata široko područje primene najnovijih tehnoloških dostignuća za inteligentno rešavanje različitih vrsta problema. U jednom periodu razvoja veštačke inteligencije, osamdesetih i devedesetih godina 20. veka, dominirala je ideja razvoja ekspertnih sistema. To su bili kompjuterski programi pisani sa namerom da zamene ljudski intelekt i znanje.¹ *Ekspertni sistemi* u oblasti prava, odnosno inteligentni sistemi zasnovani na znanju pravnih eksperata, rešavali su pravne probleme na način kako su to radili pravni eksperti specijalizovani za određenu oblast prava. Oni su funkcionisali kao inteligentni asistenti koji pružaju rešenje pravnog problema na osnovu podataka i prethodno prikupljenog i formalizovanog znanja pravnih eksperata.² Zahvaljujući napretku mašinskog učenja u poslednjoj deceniji, nije više bilo potrebno znanje eksperata već probleme rešavaju sistemi veštačke inteligencije na bazi korišćenja velike količine podataka, koji služe kao primeri na osnovu kojih se formiraju obrasci za rešavanje problema. *Mašinsko učenje* je automatizovan proces otkrivanja korelacija između varijabli u velikom skupu podataka kako bi se predvideo ili procenio ishod, odnosno donosile odluke.³ U osnovi tog automatizovanog sistema donošenja odluka je algoritam.

2. Pojam algoritamske diskriminacije

Algoritam je izraz koji je nastao u matematici i označavao je postupak računanja decimalnim brojevima, ali se danas prevashodno povezuje sa područjem informatike i računarstva i označava niz precizno definisanih pravila i procedura, na osnovu kojih se ulazne vrednosti transformišu u izlazne ili se opisuje izvršavanje nekog postupka. Preciznije, to je uputstvo kako rešiti neki zadatak ili problem. Algoritam je ključni pojam u računarskoj obradi podataka. On je *računarski program koji računaru objašnjava koje naredbe i kojim redosledom treba da obavi*. Algoritam u stvari sadrži niz instrukcija koje računar određenim redosledom treba da izvrši. On mora biti jasno definisan, što podrazumeva da treba da budu predviđene sve moguće situacije koje se mogu pojaviti. Faze rešavanja zadataka uz pomoć algoritma obuhvataju analizu i preciznu identifikaciju problema, opšti opis i detaljni opis algoritma za rešavanje problema, pisanje programa za računar na osnovu detaljnog algoritma, testiranje programa, primenu programa i analizu njegovog funkcionisanja.

¹ D. Prlja, „Pravni ekspertni sistemi”, *Kompjuteri i pravo* 1–2/1993, 44.

² R. Susskind, *Expert Systems in Law*, Oxford University Press, Oxford 1987, 12.

³ F. Zuiderveen Borgtesius, *Diskrimination, Artificial Intelligence And Algorithmic Decision-Making*, Council of Europe, Strasbourg 2018, 13–14.

Algoritamski sistem donošenja odluka može se definisati kao računarski proces, uključujući onaj koji je izveden iz mašinskog učenja, statistike ili druge obrade podataka ili druge tehnike veštačke inteligencije, a koji sam donosi odluke ili podržava ljudsko donošenje odluka.⁴

Algoritam se može opisati kao apstraktan, formalizovan opis računске procedure. Algoritamska odluka je rezultat, nalaz ili ishod te procedure.

Ponekad algoritam odlučuje potpuno automatski. Na primer, filter neželjene pošte za uslugu e-pošte može potpuno automatski filtrirati neželjene poruke iz *inboxa* korisnika. Ponekad ljudi donose odluke pomoću algoritama i takve odluke su delimično automatske. Na primer, na osnovu procene klijentovog kredita koju je uradio sistem veštačke inteligencije, službenik banke može odlučiti da li klijent može pozajmiti novac od banke.⁵ Međutim, treba razlikovati odluke donete na bazi algoritma koje su *potpuno automatizovane* i one koje su samo *delimično automatizovane*. To je važno da bi mogla da se utvrdi odgovornost u slučaju kršenja ljudskih prava i zbog načina na koji se mogu izvršiti izmene sistema automatskog donošenja odluka, kako kršenja ljudskih prava ne bi bilo u budućnosti. Uočena je tendencija da sistem algoritamskog donošenja odluka uz delimično učešće ljudi smanjuje svoju odgovornost tako što se jednostavno slede preporuke kompjutera. Ta pojava se naziva pristrasnost na bazi automatizacije (*automation bias*).⁶

Prava koja poseduju ljudska bića proizilaze iz činjenice da jesu ljudska bića zbog čega prihvatanje nejednakosti ljudi ruši suštinski celokupan koncept ljudskih prava.⁷ *Načelo jednakosti* je osnovno načelo ljudskih prava te iz njega proizilazi *načelo nediskriminacije*, odnosno zabrana diskriminacije među ljudima.

Sama reč diskriminacija dolazi iz latinskog jezika i znači razlikovanje, ali je, bar u pravu, izgubila neutralnost i dobila negativno značenje nedozvoljenog razlikovanja. U oblasti ljudskih prava, diskriminacija je razlikovanje u posedovanju i obimu prava, koje nije dozvoljeno zbog osnova i načina razlikovanja.⁸

Zabrana diskriminacije u savremenom pravu odnosi se na razlikovanje na osnovu rase, boje, pola, jezika, veroispovesti, političkog ili drugog uverenja, nacionalnog i socijalnog položaja, porekla, imovine, rođenja i drugog statusa. Nabranjenje nije konačno, odnosno moraju se imati u vidu i drugi slični osnovi diskriminacije.⁹ U novijim pravnim dokumentima zabranjena je diskriminacija na osnovu roda, seksualne orijentacije, rodnog identiteta, uzrasta, zdravstvenog stanja, invaliditeta, bračnog stanja, migrantskog statusa, odnosno drugog statusa, što znači da ni ovaj spisak nije konačan.¹⁰

⁴ European Law Institute, *Model Rules on Impact Assessment of Algorithmic Decision Making Systems Used by Public Administration*, European Law Institute, Vienna 2022, 16.

⁵ F. Zuiderveen Borgtesius, 11.

⁶ *Ibid.*, 12.

⁷ V. Dimitrijević, M. Paunović, *Ljudska prava*, Beogradski centar za ljudska prava, Beograd 1997, 181.

⁸ *Ibid.*, 185.

⁹ *Ibid.*, 186.

¹⁰ G. Gasmi, „Principi nediskriminacije i savesnog ispunjavanja obaveza država – Teorijski pogled na Istanbul Konvenciju Saveta Evrope i praksu Evropskog suda za ljudska prava”, *Kultura Polisa*1/2016, 149–167.

Jedan od najčešće prijavljivanih negativnih uticaja veštačke inteligencije na ljudska prava jeste uticaj na zabranu diskriminacije, odnosno na pravo na jednak tretman.¹¹

3. Primeri algoritamske diskriminacije

Sistemi veštačke inteligencije zasnovani na pristrasnim informacijama mogu prouzrokovati algoritamsku diskriminaciju, odnosno diskriminatorne algoritamske odluke ili ponašanja. Ukoliko sistem veštačke inteligencije uči na bazi prethodnih podataka zasnovanih na diskriminatornim odlukama, onda i on sam može na bazi „povratnih petlji” da donosi diskriminatorne odluke, odnosno može da ugrozi ljudska prava. Primera radi, ako su u prošlosti na određenim radnim mestima zapošljavani mnogo više muškarci nego žene, onda će sistem veštačke inteligencije na osnovu učenja na prethodnim primerima i u budućnosti diskriminirati žene prilikom zapošljavanja na ta radna mesta. Sistemi veštačke inteligencije namenjeni praćenju ponašanja zaposlenih i donošenju odluka od algoritma mogu negativno uticati na ostvarivanje socijalnih i ekonomskih prava zaposlenih. Zaposleni se mogu suočiti sa greškama sistema veštačke inteligencije, koje za posledicu imaju neopravdano manju platu, neisplaćen regres, neadekvatno raspoređivanje na drugo radno mesto itd. Posledica upravljanja algoritma radnim procesima mogu biti dehumanizacija i ugrožavanje prava zaposlenih.

Postoji više načina koji mogu prouzrokovati algoritamsku diskriminaciju: izbor ciljne varijable i oznaka klase, izbor podataka za obuku sistema, prikupljanje podataka za obuku sistema, izbor karakteristika, posredne pretpostavke i zlonamerno korišćenje sistema veštačke inteligencije sa ciljem diskriminacije.¹²

Ciljne varijable su korelacije koje algoritam pronalazi u skupovima podataka. Ako je svrha algoritma da ukloni neželjenu elektronsku poštu, onda softver treba da pronađe karakteristike na osnovu kojih će neku elektronsku poštu označiti kao neželjenu. Na primer, elektronske poruke koje predstavljaju neželjenu poštu često sadrže istu formulaciju, kao što je „dobili ste milion dolara” ili dolaze sa iste elektronske adrese sa koje su identifikovani prethodni slučajevi slanja neželjene pošte ili slučajevi prevare. Na bazi mašinskog učenja, algoritam prikuplja te ciljne varijable. Oznaka klase deli te ciljne varijable na međusobno isključive kategorije. Nekada je potpuno jasno kako treba napraviti oznaku klase, na primer za neželjenu elektronsku poštu, ali nekada to i nije tako jednostavno. U slučaju kada treba odrediti oznaku klase za algoritam za zapošljavanje novih radnika, postavlja se pitanje da li je dobar radnik onaj koji dolazi tačno na posao ili onaj koji prodaje više proizvoda. Ako se kao ciljna varijabla i oznaka klase izabere tačnost dolaska na posao, to u nekim slučajevima može biti osnov za diskriminaciju. Logično je pretpostaviti da ljudi imigrantskog porekla i oni siromašniji ne žive u centru grada već na periferiji,

¹¹ I. Ben-Israel *et al.*, *Towards Regulation of Artificial Intelligence Systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law*, Council of Europe, 2020, 28, <https://www.coe.int/en/web/artificial-intelligence/cahai>, 28. mart 2023.

¹² F. Zuiderveen Borgtesius, 15.

te će oni zbog saobraćajnih gužvi i problema sa prevozom češće kasniti na posao. Posledica takvog izbora ciljne varijable i oznake klase može biti algoritamska diskriminacija, zbog načina na koji su definisane ciljne varijable i oznake klase.

Podaci koji se koriste za obuku sistema veštačke inteligencije takođe mogu biti osnov algoritamske diskriminacije. Moguće je da su sami podaci bili uzrok diskriminacije i ranije, pa ako sistem veštačke inteligencije uči na takvim podacima, tada on samo ponavlja diskriminaciju koja je i ranije postojala. Na taj način ranije ustanovljena pristrasnost, kao što je to prilikom zapošljavanja u korist muškaraca u odnosu na žene, biće samo ponovljena i u sistemu veštačke inteligencije. Moguće je da je u prošlosti policija usmerila pažnju na određene etničke grupe i određena naselja, da je češće zaustavljala imigrante te je, naravno, u tim slučajevima evidentirano i više počinilaca krivičnih dela. Stoga će algoritam koji uči na tim primerima izazvati povratnu petlju i reprodukovati pristrasnost, odnosno diskriminaciju.

Uzrok algoritamske diskriminacije može biti izbor karakteristika, odnosno kategorija podataka koje će koristiti sistem veštačke inteligencije. Ukoliko sistem veštačke inteligencije ima namenu da pomogne poslodavcu u izboru kandidata za zapošljavanje, on može biti pristrasan prema određenim grupama kandidata. Ako se kao karakteristika traži studiranje na poznatim i skupim fakultetima, onda će to biti osnov za diskriminaciju onih koji su siromašnijeg imovnog stanja te nisu mogli da studiraju na skupim fakultetima.

Posredne pretpostavke takođe mogu biti osnov algoritamske diskriminacije, ukoliko banka koristi sistem veštačke inteligencije da predvidi koji će korisnici imati probleme sa otplatom kredita. Na osnovu podataka iz prethodnih dvadeset godina, sistem veštačke inteligencije je zaključio da korisnici sa određene poštanske adrese u većoj meri nisu otplaćivali kredite. Poštanski broj je neutralan kriterijum, ali ako na tom poštanskom broju stanuju pripadnici određene etničke manjine i ako banka koristi taj kriterijum da uskrati kredite, rezultat je algoritamska diskriminacija pripadnika te etničke grupe. Skup podataka koji ne sadrži eksplicitne podatke o seksualnoj orijentaciji ipak može na posredan način da ukaže na njih. Recimo, procenat prijatelja na Fejsbuku može da otkrije seksualnu orijentaciju pojedinca. Problem posrednih pretpostavki je izuzetno ozbiljan i teško ga je rešiti.

Osim već pomenutih slučajeva algoritamske diskriminacije koji nisu namerom izazvani, postoje i slučajevi da institucija koja je kreirala sistem veštačke inteligencije pomoću sistema veštačke inteligencije vrši namernu diskriminaciju. To je slučaj kada institucija kreira sistem veštačke inteligencije sa namerom da prednost u zapošljavanju preda muškarcima na štetu žena i tada namerno bira karakteristike uslova za zapošljavanje koje će u većoj meri odgovarati muškarcima, a ne ženama.

Primeri sistema veštačke inteligencije koji vrše algoritamsku diskriminaciju zaista su mnogobrojni.

U javnom sektoru u SAD se koristi sistem veštačke inteligencije *COMPAS*, namenjen profilisanju prestupnika, kako bi se što adekvatnije odredile sankcije. Taj sistem treba da predvidi da li će prestupnik ponovo počiniti krivično delo, kako bi sudije mogle da se odluče u kojim slučajevima će odrediti uslovnu kaznu, a u kojima

neće. Istraživanja su pokazala da taj sistem po rasnoj osnovi diskriminiše crnce, u korist belaca. Sistem veštačke inteligencije koji je koristila privatna kompanija *Amazon* u SAD za odabir programera koje će zaposliti nije bio rodno neutralan već je sam sebe naučio da su muški kandidati poželjniji od ženskih kandidata.¹³

Algoritamska diskriminacija se javlja i u sistemima veštačke inteligencije namenjenim oglašavanju. Na primer, analiza oglasa za posao koje *Google* sistem veštačke inteligencije nudi muškarcima i ženama pokazala je da taj sistem više nudi muškarcima nego ženama oglase za visokoplaćene i bolje poslove. Na taj način neko može biti diskriminisan, a da zapravo i ne zna da je diskriminisan. Holandsko telo za zaštitu podataka utvrdilo je da Fejsbuk omogućava oglašivačima da ciljaju ljude određenih seksualnih preferencija, određene rase ili određenog uzrasta. Tako se, recimo, oglasi za posao nude samo osobama mlađim od određenog broja godina.¹⁴

Algoritamska diskriminacija može da postoji i prilikom nudenja identičnih proizvoda po različitim cenama. Iako su troškovi proizvodnje ili davanja usluga identični, cene proizvoda mogu da variraju zavisno od područja, dela države, dela grada, pa se kupcima za isti proizvod ili uslugu u drugim područjima nudi različita cena. To na posredan način može da izazove diskriminaciju, ako u određenom delu grada živi pretežno populacija određenih rasnih ili drugih karakteristika. Iako cilj kompanije koja nudi proizvod ili uslugu nije bila diskriminacija po rasnoj osnovi već uvećavanje profita na bazi tržišnih analiza, konačan efekat takvog ponašanja može da bude rasna diskriminacija.

Sistemi za pronalaženje i prepoznavanje na slikama takođe mogu imati diskriminacioni efekat. Razlog može biti učenje sistema za prepoznavanje na primerima pripadnika bele rase, a kasnije taj sistem u velikom broju slučajeva greši u prepoznavanju pripadnika drugih rasa.

Takođe, uočena je tendencija da alati za prevođenje na bazi veštačke inteligencije, kakav je na primer *Google Translate*, odražavaju postojeću rodnu neravnopravnost. To znači da bi oni mogli još više da pogoršaju rodnu nejednakost.

Svi navedeni primeri algoritamske diskriminacije ukazuju na to da sistemi veštačke inteligencije često prave razne oblike diskriminacije zahvaljujući tome što uče na bazi prethodnih loših primera. Da bi se eliminisali ti oblici diskriminacije, neophodno je da se sistemi veštačke inteligencije koriguju na osnovu primera idealnog ponašanja bez diskriminacije. Sistemi veštačke inteligencije ne bi trebalo da ponavljaju stereotipna ponašanja koja vode diskriminaciji već se moraju korigovati sa ciljem ukidanja nepravdi i poštovanja propisa o jednakosti svih.

4. Pravo kao sredstvo zaštite od algoritamske diskriminacije

Obavezujuće i neobavezujuće norme glavne su prepreke algoritamskoj diskriminaciji. U prvom redu to su obavezujući propisi na međunarodnom i

¹³ F. Zuiderveen Borgesius, 26.

¹⁴ *Ibid.*, 27.

nacionalnom nivou o nediskriminaciji i o zaštiti podataka, ali i mnogi drugi propisi, standardi i pravila ponašanja.

Univerzalna deklaracija UN o pravima čoveka iz 1948. godine u svom prvom članu garantuje slobodu i jednaka prava svim ljudima, a u drugom članu garantuje nediskriminaciju: „Svakom pripadaju sva prava i slobode proglašene u ovoj Deklaraciji, bez ikakvih razlika u pogledu rase, boje, pola, jezika, veroispovesti, političkog ili drugog mišljenja, nacionalnog ili društvenog porekla, imovine, rođenja ili drugih okolnosti.”¹⁵

Evropska konvencija o zaštiti ljudskih prava i osnovnih sloboda (1950), u skladu sa Univerzalnom deklaracijom UN o pravima čoveka, u članu 14. zabranjuje diskriminaciju sledećom odredbom: „Uživanje prava i sloboda navedenih u ovoj Konvenciji osiguraće se bez diskriminacije po bilo kom osnovu kao što su: pol, rasa, boja kože, jezik, vera, političko ili drugo mišljenje, nacionalno ili socijalno poreklo, povezanost s nacionalnom manjinom, imovina, rođenje ili drugi status.”¹⁶

Direktiva EU iz 2000. godine o implementaciji principa jednakog tretmana osoba bez obzira na rasno ili etničko poreklo poznaje dva oblika diskriminacije: direktnu i indirektnu diskriminaciju. U članu 2. te direktive *direktna diskriminacija* je definisana na sledeći način: „Direktna diskriminacija postoji kada se jedna osoba tretira nepovoljnije od druge osobe, bila je tretirana ili bi bila tretirana nepovoljnije u sličnoj situaciji na osnovu rasnog ili etničkog porekla.” U istom članu *indirektna diskriminacija* je definisana ovako: „Smatra se da se indirektna diskriminacija javlja kada bi naizgled neutralna odredba, kriterijum ili praksa dovela osobe rasnog ili etničkog porekla u posebno nepovoljan položaj u odnosu na druge osobe, osim ako je ta odredba, kriterijum ili praksa objektivno opravdana legitimnim ciljem, a sredstva za postizanje tog cilja su odgovarajuća i neophodna.”¹⁷

Algoritamska diskriminacija je upravo često indirektna diskriminacija, odnosno naizgled neutralna odredba dovodi u posebno nepovoljan položaj određenu grupu ljudi. Dakle, nije relevantno da li je diskriminator imao nameru da diskriminiše već je relevantan efekat koji je u praksi imalo algoritamsko odlučivanje. Indirektna diskriminacija se dešava mnogo češće nego direktna. Primera radi, algoritamskim odlukama koje teraju ljude određenog rasnog porekla da plaćaju proizvođači ili usluge po višim cenama krše se odredbe o zabrani indirektno diskriminacije.¹⁸

Teškoće u dokazivanju indirektno diskriminacije nastupaju kada standarde treba primeniti u praksi. Indirektna diskriminacija ne postoji ako postoji objektivno i razumno opravdan cilj. Prema jednoj od odluka Evropskog suda za ljudska prava, ako teži legitimnom cilju i postoji proporcionalnost između upotrebljenog

¹⁵ O. Račić, B. Milinković, M. Paunović, *Ljudska prava – Pet decenija od usvajanja Opšte deklaracije o pravima čoveka*, Međunarodna politika, Službeni list SRJ, Pravni fakultet, Fakultet političkih nauka, Beograd 1998, 126.

¹⁶ Evropska konvencija za zaštitu ljudskih prava i osnovnih sloboda, Rim 1950, *Službeni list SCG–Međunarodni ugovori*, br. 9/2003.

¹⁷ Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, *Official Journal of the European Union* L180/2000, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000L0043>, 7. april 2022.

¹⁸ F. Zuiderveen Borgtesius, 34.

sredstva i cilja koji se želi postići, a sredstva za postizanje tog cilja su prikladna i neophodna, nema indirektna diskriminacije.¹⁹

Indirektnu diskriminaciju je teško otkriti i zbog razloga *netransparentnosti* sistema veštačke inteligencije. Najčešće korisnici usluga jedne banke nemaju informaciju o tome da odluke o njihovim zahtevima za kredit donose sistemi veštačke inteligencije, a ne ljudi. Zatim korisnici usluga jedne banke, čak i kada znaju da o njihovim zahtevima odlučuje sistem veštačke inteligencije, nemaju adekvatno objašnjenje zbog čega je njihov zahtev odbijen te je logično da klijenti teško mogu da utvrde da li je algoritamska odluka diskriminatorna ili nije.

Može se oceniti da se zabrana direktne i indirektna diskriminacije odnosi i na direktnu i indirektnu algoritamsku diskriminaciju, ali da se oni koji žele da otkriju i dokažu takvu vrstu diskriminacije suočavaju sa mnogobrojnim problemima. Kao i u mnogim drugim slučajevima, pravna pravila postoje, ali njihovo sprovođenje zahteva mnogobrojna razjašnjenja, sudsku praksu i uspostavljanje novih standarda u oblasti kreiranja i primene sistema veštačke inteligencije.

Sistemi veštačke inteligencije baziraju algoritamsko donošenje odluka na velikom broju podataka koje prikupljaju o pojedincima. *Propisi o zaštiti podataka* imaju cilj da obezbede poštovanje svih osnovnih prava i sloboda, da obezbede jednaka prava, a time i nediskriminaciju za sve one čiji se podaci obrađuju.

Opšta uredba o zaštiti podataka EU²⁰ koja je doneta 2016, a počela je da se primenjuje 2018. godine, u velikoj meri je uticala na promenu sistema zaštite podataka u EU i u čitavom svetu. Ona zahteva podizanje nivoa transparentnosti u svim slučajevima obrade podataka, a posebno kada sistemi veštačke inteligencije automatizovano donose odluke, što je i navedeno u članu 13. tačka f. U toj odredbi se navodi da će se pojedincima dati informacije o tome po kojoj logici su odluke donošene, kakav je značaj i kakve su posledice takve obrade podataka i donošenja takvih odluka za pojedince.

Transparentnost se zahteva kada se prikupljaju lični podaci, kada se oni upotrebljavaju, kada se daju na uvid ili na drugi način obrađuju.

Na osnovu načela transparentnosti, u tački 39. Preambule Opšte uredbе o zaštiti podataka EU zahteva se da svaka informacija i komunikacija u vezi sa obradom ličnih podataka bude lako dostupna i razumljiva i da se upotrebljava jasan i jednostavan jezik. Pojedinci imaju pravo da dobiju informaciju o ličnim podacima koji se obrađuju, a treba da budu upoznati sa rizicima, pravilima, zaštitnim merama i pravima u vezi sa obradom ličnih podataka i načinom ostvarivanja svojih prava u vezi sa obradom tih podataka.²¹

¹⁹ European Court of Human Rights, Case of Biao v. Denmark, 2016, 38590/10, [https://hudoc.echr.coe.int/fre#%22itemid%22:\[%22001-163115%22\]](https://hudoc.echr.coe.int/fre#%22itemid%22:[%22001-163115%22]), 7. april 2022.

²⁰ Regulation 2016/679 of the European Parliament and of the Council of 7 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), *Official Journal of the European Union* L 119/2016.

²¹ *Ibid.*

Tačka 71. Preambule i član 22. Opšte uredbe o zaštiti podataka EU direktno regulišu algoritamsko donošenje odluka. Tim odredbama se utvrđuje pravo pojedinca da zahteva da se na njega ne odnosi odluka doneta isključivo na osnovu automatizovane obrade podataka koja proizvodi pravne posledice za njega ili na njega značajno utiče. Navedeni su primeri odbijanja kredita putem interneta ili prakse za pošljavanja putem interneta. Takva automatizovana obrada podataka uključuje izradu profila, odnosno procenjivanje ličnih karakteristika pojedinca, posebno onih koji se odnose na rezultate na poslu, ekonomsko stanje, zdravlje, lične preferencije ili interese, pouzdanost ili ponašanje, lokaciju ili kretanje, kada ona proizvodi pravne posledice koje se odnose na pojedinca ili na njega utiču. Međutim, donošenje odluka koje se temelje na takvoj obradi, što uključuje i izradu profila, može biti dopušteno ako to dozvoljava pravo EU ili pravo države članice kojem podleže rukovalac podataka, između ostalog, u svrhe praćenja i sprečavanja prevare i poreske utaje, u skladu s propisima, standardima i preporukama institucija EU ili nacionalnih organa. Na taj način se osiguravaju bezbednost i pouzdanost usluge koju pruža rukovalac podataka ili ako je to neophodno za sklapanje ili izvršavanje ugovora između pojedinca i rukovaoca ili kada je pojedinac izričito dao svoju saglasnost.

U svakom slučaju na takve bi obrade trebalo primenjivati odgovarajuće zaštitne mere, koje bi trebalo da uključe davanje određenih informacija pojedincu i pravo na učešće čoveka u procesu odlučivanja, pravo na izražavanje sopstvenog mišljenja, na dobijanje objašnjenja odluke donete nakon takve procene i pravo na osporavanje odluke. Da bi se obezbedila poštena i transparentna obrada podataka u odnosu na pojedinca, uzimajući u obzir posebne okolnosti i kontekst u kojem se lični podaci obrađuju, rukovalac podataka treba da preduzme odgovarajuće matematičke i statističke postupke u vezi sa izradom profila. To su tehničke i organizacione mere, kako bi se posebno obezbedilo da budu ispravljani uzroci netačnosti u ličnim podacima i da se rizici od pojave grešaka svedu na minimum. Te mere treba da uzmu u obzir potencijalne rizike za interese i prava pojedinaca i njima se, između ostalog, sprečavaju diskriminacioni efekti na pojedince na osnovu rasnog ili etničkog porekla, političkog mišljenja, vere ili uverenja, članstva u sindikatu, genetskog ili zdravstvenog stanja ili polne orijentacije ili po drugom osnovu sa sličnim posledicama. Automatizovane odluke i izrada profila na temelju posebnih kategorija ličnih podataka mogu se dozvoliti samo pod posebnim uslovima.²²

Primeri automatizovanih odluka na bazi algoritma sa pravnim posledicama jesu sudske odluke ili odluke državnih organa o socijalnim davanjima predviđenim zakonom, kao što su penzije i razna druga socijalna davanja. Automatizovane odluke sa sličnim posledicama jesu odluke koje donose banke kada odobravaju kredit ili odluke koje donose privatne kompanije o cenama proizvoda ili usluga koje dovode pojedince i grupe pojedinaca u nejednak i neravnopravan položaj u odnosu na druge. Zabrana takvih diskriminatornih automatizovanih odluka baziranih na algoritmu ne primenjuje se u slučajevima kada je pojedinac dao saglasnost, kada je to predviđeno ugovorom između rukovaoca podataka i pojedinca i kada je to

²² *Ibid.*

predviđeno zakonom. U tim slučajevima, kako je to predviđeno članom 22. Opšte uredbe o zaštiti podataka EU, neophodno je da rukovalac podataka sprovede odgovarajuće mere zaštite prava i sloboda i legitimnih interesa pojedinaca, a posebno prava na učešće čoveka u donošenju odluka, prava izražavanja sopstvenog mišljenja i prava osporavanja odluke.

Na algoritamsku diskriminaciju odnosi se i tačka 75. Preambule Opšte uredbe o zaštiti podataka EU. U toj odredbi se naglašava da obrada ličnih podataka može prouzrokovati fizičku, materijalnu ili nematerijalnu štetu, posebno ako ta obrada može dovesti do *diskriminacije*, krađe identiteta ili prevare, finansijskog gubitka, štete po ugled, gubitka poverljivosti ličnih podataka zaštićenih poslovnim tajnom, neovlašćenog obrnutog postupka pseudonimizacije ili bilo koje druge znatne ekonomske ili društvene štete; ili ako pojedinci mogu biti uskraćeni za svoja prava i slobode ili sprečeni u obavljanju kontrole nad svojim ličnim podacima; ako se obrađuju lični podaci koji odaju rasno ili etničko poreklo, politička mišljenja, verska ili filozofska uverenja, članstvo u sindikatu i ako genetički podaci, podaci koji se odnose na zdravlje ili polni život ili krivične presude i kažnjiva dela ili povezane bezbednosne mere; ako se procenjuju lične karakteristike, posebno ako se vrši analiza ili predviđanje karakteristika u vezi s rezultatima na poslu, ekonomskim stanjem, zdravljem, ličnim preferencijama ili interesima, pouzdanošću ili ponašanjem, lokacijom ili kretanjem, kako bi se izradili ili upotrebljavali lični profili; ako se obrađuju lični podaci osetljivih pojedinaca, posebno dece; ili ako obrada obuhvata veliku količinu ličnih podataka i utiče na veliki broj pojedinaca.²³

Na osnovu člana 24. Opšte uredbe o zaštiti podataka EU, rukovaoci su dužni da uzmu u obzir *rizike* različitih nivoa verovatnoće i težine koji mogu ugroziti prava i slobode pojedinaca. Tu spada i rizik od diskriminacije pojedinaca tokom obrade podataka. Rukovaoci su dužni da preduzmu odgovarajuće tehničke i organizacione mere kako bi se ti rizici otklonili. To je potrebno posebno kada se primenjuju nove tehnologije, kao što je veštačka inteligencija, i algoritamsko donošenje odluka ukoliko rukovalac nije sproveo *procenu uticaja* na zaštitu podataka, kao što je navedeno u tački 89. Preambule Opšte uredbe o zaštiti podataka EU.

Članom 35. Opšte uredbe o zaštiti podataka EU ustanovljena je obaveza rukovalaca da procenjuju uticaj u slučajevima postojanja visokog rizika po prava i slobode pojedinaca. Procena uticaja u vezi sa zaštitom podataka predstavlja postupak procene rizika po prava i slobode pojedinaca i ispitivanje zaštitnih mera, koje se odnose na obradu podataka o ličnosti. Osim toga, taj postupak obuhvata i procenu neophodnosti i proporcionalnosti obrade podataka o ličnosti. Postupak procene uticaja u vezi sa zaštitom podataka omogućava rukovaocima da dokažu usaglašenost sa Opštom uredbom o zaštiti podataka EU, odnosno da dokažu da su preduzete adekvatne mere zaštite.²⁴

U stavu 3. člana 35. Opšte uredbe o zaštiti podataka EU posebno je definisano da je neophodnost sprovođenja procene uticaja u vezi sa zaštitom podataka

²³ *Ibid.*

²⁴ A. Diligenski, D. Prlja, D. Cerović, *Pravo zaštite podataka GDPR*, Institut za uporedno pravo, Beograd 2018, 163.

obavezna kada se sprovodi sistematska i obimna procena ličnih aspekata u vezi sa pojedincima, koja se temelji na automatskoj obradi podataka, uključujući izradu profila, i na osnovu koje se donose odluke koje proizvode pravna dejstva za pojedince ili na sličan način značajno utiču na pojedince.

Izradom profila i automatskim donošenjem odluka stvara se velika verovatnoća za nastupanje rizika za prava i slobode pojedinaca. Usled profilisanja ili automatskog donošenja odluka, pojedinci mogu biti razvrstani u određene kategorije, što im može ograničiti izbor određenih proizvoda ili usluga, može uticati na pogrešne prognoze i predviđanja, a u krajnjoj liniji, može dovesti do uskraćivanja usluga ili proizvoda, odnosno algoritamske diskriminacije.²⁵

Opšta uredba o zaštiti podataka EU ima veliki značaj za sprečavanje diskriminacije prilikom obrade ličnih podataka, ali i pored toga jedan deo algoritamskog donošenja odluka koji se ne odnosi na obradu ličnih podataka ostaje van domašaja tog pravnog propisa. To su *prediktivni modeli* koji su rezultat rada sistema za veštačku inteligenciju baziranih na algoritmima na osnovu kojih se pojedinci ne mogu identifikovati. Za prikupljanje posebno osetljivih kategorija podataka, prema Opštoj uredbi o zaštiti podataka EU, kao što su podaci o rasnom poreklu, zdravstvenom statusu i tome slično, predviđen je vrlo strog režim obrade podataka. Organizacije koje žele da obrađuju te podatke moraju da ispune mnogobrojne tehničke i organizacione mere. Mnoge organizacije koje žele da procene da li postoji diskriminacija i da se bore protiv diskriminacije ne mogu da ispune te stroge mere za obradu posebno osetljivih kategorija podataka. Time im je otežana mogućnost da efikasno ustanove da li postoji algoritamska diskriminacija i da se bore protiv takvog oblika diskriminacije.²⁶

Dosadašnja primena Opšte uredbi o zaštiti podataka EU, kao i samo njeno donošenje, pokazali su da taj pravni propis značajno utiče na poboljšanje nacionalnih zakonodavstava u oblasti zaštite podataka i nudi pravna sredstva za borbu protiv algoritamske diskriminacije.

Osim propisa o zabrani diskriminacije i propisa o zaštiti podataka u pravnim sistemima, postoji i veliki broj drugih propisa koji mogu biti od pomoći u borbi protiv algoritamske diskriminacije. To su propisi o zaštiti potrošača, o konkurenciji, o slobodnom pristupu informacijama od javnog značaja, o slobodi informisanja itd. Sve je više i pravila neobavezujućeg karaktera o etičkim principima upotrebe veštačke inteligencije, koje formulišu međunarodne organizacije i profesionalna udruženja, koja mogu poslužiti kao dobra osnova za formulisanje obavezujuće pravne regulative.

5. Vanpravna sredstva zaštite od algoritamske diskriminacije

Osim strogo pravnih sredstava za borbu protiv algoritamske diskriminacije, postoje i mnoge druge mere koje mogu da pomognu u toj borbi. Te mere su

²⁵ *Ibid.*, 49.

²⁶ F. Zuiderveen Borgtesius, 44–45.

obrazovanje, procena i ublažavanje rizika, transparentnost javnog sektora, jačanje kapaciteta nadležnih organa koji se bave ravnopravnošću i unapređivanje organizacija za kontrolu poštovanja ljudskih prava itd.²⁷

Veliki broj ljudi, inženjera, pravnika i ekonomista nisu svesni rizika koje donosi upotreba veštačke inteligencije. Upravo zato je neophodno da se na svim nivoima obrazovanja usmeri pažnja na sticanje novih znanja koja mogu pomoći u borbi protiv algoritamske diskriminacije.

Procena i ublažavanje rizika su neophodni u svim projektima koji su povezani sa upotrebom veštačke inteligencije. Svaka institucija koja ima nameru da stvori proizvod ili uslugu baziranu na veštačkoj inteligenciji, odnosno algoritmima, treba da formira tim sačinjen od inženjera, pravnika i ekonomista, koji će proceniti moguće rizike i uticaj na prava i slobode pojedinaca. To nije neophodno samo u fazi kreiranja sistema veštačke inteligencije već i kasnije, u fazi testiranja i implementacije sistema veštačke inteligencije.

Posebna odgovornost za upotrebu sistema veštačke inteligencije postoji kada takve sisteme koriste javne službe i državni organi. Njihove odluke često pogađaju veliki broj pojedinaca, pa je neophodno da iz takvog algoritamskog odlučivanja budu isključeni svi mogući oblici diskriminacije. Da bi se postigle pravičnost i jednakost prilikom upotrebe veštačke inteligencije i algoritamskog odlučivanja u javnom sektoru, te sisteme veštačke inteligencije je neophodno stalno nadzirati i kontrolisati njihovo funkcionisanje.

Institucije koje se bave ravnopravnošću i ljudskim pravima treba da se pozabave pitanjima algoritamske diskriminacije, ali da bi u tome bile uspešne, treba da steknu nova znanja i da koriste posebne ekspertske analize. Zadatak tih institucija je i da rade na podizanju svesti javnosti o rizicima koje donosi algoritamsko odlučivanje. Saradnja institucija koje se bave ravnopravnošću i organizacija koje se bave ljudskim pravima sa institucijama koje sa bave zaštitom podataka i zaštitom potrošača takođe doprinosi uspešnjoj borbi protiv algoritamske diskriminacije.

6. Zaključna razmatranja

Izričito su zabranjeni svi oblici diskriminacije koji narušavaju jednakost pojedinaca u skladu sa važećim pravnim sistemima na nivou međunarodne zajednice i na nacionalnom nivou. Sledi da je i algoritamska diskriminacija izričito zabranjena. Problem nastaje kada je treba identifikovati i otkloniti u praksi. Problem otkrivanja algoritamske diskriminacije direktno je povezan sa problemom transparentnosti sistema veštačke inteligencije, odnosno algoritamskog odlučivanja.

Logika po kojoj algoritam odlučuje mora biti jasna, poštena, pravedna, unapred testirana i transparentno objašnjena pojedincima na koje se algoritamsko odlučivanje odnosi. Algoritamsku diskriminaciju bi bilo lakše otkriti ako bi bili

²⁷ *Ibid.*, 51–60.

transparentno poznati standardi na osnovu kojih se kontroliše primena sistema veštačke inteligencije i meri kvalitet odluka koje algoritmi donose. Nažalost, u praksi je često teško saznati kako algoritam funkcioniše jer je softver na bazi koga je nastao sistem veštačke inteligencije i na osnovu koga algoritam donosi odluke zaštićen poslovnim tajnama, pravima intelektualne svojine i internim aktima kompanija koje ga kreiraju. U slučaju algoritamske diskriminacije, kao i u mnogim drugim slučajevima zaštite ljudskih prava, javlja se sukob javnog interesa da se ostvari transparentnost i zaštite osnovna prava i slobode sa privatnim interesom da se ostvari profit i zaštite neka druga prava.

Algoritamska diskriminacija može nastati radom sistema veštačke inteligencije i u javnom i u privatnom sektoru. U javnom sektoru su to, primera radi, preporuke za odmeravanje kazni za krivična dela, donošenje odluka o penzijskim i socijalnim davanjima ili naknadama za nezaposlene i slično. U privatnom sektoru to su odluke koje se donose u vezi sa zapošljavanjem, stanovanjem, odobravanjem kredita, reklamiranjem proizvoda itd.

Postojeći pravni propisi o nediskriminaciji i zaštiti podataka, zaštiti potrošača itd. pružaju određene mogućnosti za borbu protiv algoritamske diskriminacije. Međutim, te propise treba unaprediti i doneti nove opšte i sektorske propise, kojima će se regulisati upotreba sistema veštačke inteligencije zasnovanih na algoritamskom odlučivanju, na bezbedan način i uz visoki nivo zaštite pojedinačnih prava i sloboda.

Jedno od dobrih rešenja za otklanjanje algoritamske diskriminacije jeste da se pravnim propisima zahteva ispunjavanje određenih uslova i davanje sertifikata za sisteme veštačke inteligencije koji su visokorizični, odnosno koji mogu ugroziti ljudska prava i slobode, kao što je to predviđeno Predlogom uredbe EU o veštačkoj inteligenciji 2021. godine.²⁸

Zbog tehnološke složenosti sistema veštačke inteligencije nije jednostavno pravnim sredstvima urediti upotrebu algoritma i zaštititi osnovna prava i slobode, ali je to moguće i zaista neophodno.²⁹

Literatura

Ben-Israel I. et al., *Towards Regulation of Artificial Intelligence Systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law*, Council of Europe, 2020, <https://www.coe.int/en/web/artificial-intelligence/cahai>.

²⁸ European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial intelligence Act) and Amending Certain Union Legislative Acts, Brussels, 21.april 2021. COM(2021) 206 final, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF, 20. septembar 2021.

²⁹ F. Zuiderveen Borgtesius, 63.

- Diligenski A., Prlja D., Cerović D., *Pravo zaštite podataka GDPR*, Institut za uporedno pravo, Beograd 2018.
- Dimitrijević V., Paunović M., *Ljudska prava*, Beogradski centar za ljudska prava, Beograd 1997.
- European Law Institute, *Model Rules on Impact Assessment of Algorithmic Decision Making Systems Used by Public Administration*, European Law Institute, Vienna 2022.
- Gasmi G., Principi nediskriminacije i savesnog ispunjavanja obaveza država – Teorijski pogled na Istanbul Konvenciju Saveta Evrope i praksu Evropskog suda za ljudska prava, *Kultura Polisa*1/2016.
- Prlja D., Pravni ekspertni sistemi, *Kompjuteri i pravo* 1–2/1993.
- Račić O., Milinković B., Paunović M., *Ljudska prava – Pet decenija od usvajanja Opšte deklaracije o pravima čoveka*, Međunarodna politika, Službeni list SRJ, Pravni fakultet, Fakultet političkih nauka, Beograd 1998.
- Susskind R., *Expert Systems in Law*, Oxford University Press, Oxford 1987.
- Zuiderveen Borgtesius F., *Diskrimination, Artificial Intelligence And Algorithmic Decision-Making*, Council of Europe, Strasbourg 2018.
- Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, *Official Journal of the European Union* L180/2000, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000L0043>.
- European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial intelligence Act) and Amending Certain Union Legislative Acts, Brussels 21. april 2021. COM(2021) 206 final, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.
- European Court of Human Rights, Case of Biao v. Denmark, 2016, 38590/10, <https://hudoc.echr.coe.int/fre#%7B%22itemid%22:%5B%22001-163115%22%5D%7D>.
- Evropska konvencija za zaštitu ljudskih prava i osnovnih sloboda, Rim 1950, *Službeni list SCG – Međunarodni ugovori* 9/2003.
- Regulation 2016/679 of the European Parliament and of the Council of 7 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), *Official Journal of the European Union* L 119/2016.

Dragan Prlja, PhD

Institute of Comparative Law, Belgrade

Gordana Gasmi, PhD

Institute of Comparative Law, Belgrade

Vanja Korać, PhD

Mathematical Institute SANU, Belgrade

THE ALGORITHMIC DISCRIMINATION

Summary

Algorithmic discrimination can result from the operation of artificial intelligence systems in both the public and private sectors. In the public sector, these are, for example, recommendations for sentencing for criminal offenses, making decisions on pension and social benefits or unemployment benefits, etc. In the private sector, these are decisions that are made in relation to employment, housing, loan approval, product advertising, etc. Existing legal regulations on non-discrimination and data protection, consumer protection, etc. provide certain opportunities to combat algorithmic discrimination. However, it is certainly necessary to improve these regulations and to adopt new general and sectoral regulations, which will regulate the use of artificial intelligence systems based on algorithmic decision-making, in a safe manner and with a high level of protection of individual rights and freedoms. One of the good solutions for eliminating algorithmic discrimination is that legal regulations require meeting certain conditions and issuing certificates for artificial intelligence systems that are high-risk or that may endanger human rights and freedoms, as provided by the EU Regulation on Artificial Intelligence from 2021.

Key words: law, legal informatics, human rights, artificial intelligence, algorithmic discrimination.