

5 ALGORITEMSKA DISKRIMINACIJA

OSKAR PEČE

Univerza v Mariboru, Pravna fakulteta, Maribor, Slovenija
oskar.pece@student.um.si

O algoritemski diskriminaciji govorimo, kadar avtomatiziran sistem, ki temelji na umetni inteligenci, posameznika obravnava manj ugodno kakor drugega posameznika na primerljivem položaju. Kljub temu, da avtomatizirani sistemi domnevno odločajo bolj objektivno in nepristransko kakor človek ter omogočajo formalno izključitev obravnave določenih spornih lastnosti, lahko odločajo diskriminatorno. Za namen boja proti algoritemski diskriminaciji je možna regulacija lastnosti, ki so upoštevane v postopku odločanja, vendar je ta pogosto neučinkovita in v nasprotju s samim smislom uporabe tovrstnih sistemov. Morda je najbolj učinkovito orodje za boj zoper algoritemsko diskriminacijo pravica, da za posameznika ne velja odločitev, ki temelji zgolj na podlagi avtomatizirane obdelave podatkov. Pri regulaciji avtomatiziranih sistemov za odločanje moramo vselej biti pozorni, da so regulatorni ukrepi tehnično izvedljivi in sorazmerni.

Ključne besede:
prepoved diskriminacije,
avtomatizirano odločanje,
umetna inteligenca,
zaščita osebnih podatkov,
podatkovno rudarjenje

DOI
[https://doi.org/
10.18690/um.pf.4.2023.4](https://doi.org/10.18690/um.pf.4.2023.4)

ISBN
978-961-286-774-4

Keywords:
non-discrimination,
automated decision making,
artificial intelligence,
personal data protection,
data mining

5 THE ALGORITHM OF DISCRIMINATION

OSKAR PEČE

University of Maribor, Faculty of Law, Maribor, Slovenia
oskar.pece@student.um.si

Discrimination is prohibited in the European Convention on Human Rights, the Charter of Fundamental Rights of the European Union, multiple directives of the European Union, the Constitution of the Republic of Slovenia and the Protection against discrimination Act. Algorithmic discrimination occurs when an automated system based on artificial intelligence treats an individual less favorably than another individual in a comparable position. Despite the fact that automated systems make decisions more objectively and impartially than humans, and allow for the formal exclusion of the treatment of certain controversial features, they they are still able of making discriminatory decisions. This is often due to human influence in data mining processes. Demonstrating algorithmic discrimination is very difficult due to limited insight into the operation of automated systems and understanding of the decision-making process itself. In proving unfavorable algorithmic discrimination, we move from the paradigm of proving causality to proving a meaningful correlation. For the purpose of combating algorithmic discrimination, it is possible to regulate the properties that are taken into account in the decision-making process, but this is often ineffective and contrary to the very purpose of using such systems.



5.1 Uvod

Umetna inteligenca je tehnologija, ki rapidno razvija in fascinira širšo javnost ne glede na obseg njihovega poznavanja delovanja in implementacije tovrstne tehnologije. Vpliv in uporaba te tehnologije je pogosta tema zabavne industrije in medijskih prispevkov, ki umetno inteligenco ljudem pogosto prikazujejo senzacionalistično in prekomerno futuristično. To vsekakor ne pomeni, da tehnologija ni presenetljivo sposobna. Ogromno sistemov umetne inteligence obdeluje osebne podatke z namenom razvoja algoritemskih modelov, na podlagi katerih lahko kasneje ob vnosu osebnih podatkov sklepajo o določenih lastnostih posameznika, na katerega se ti podatki nanašajo. S tem so tovrstni sistemi sposobni avtomatiziranega odločanja, tudi v kompleksnih primerih, ki zadevajo večje število lastnosti in predhodno nedoločenih kriterijev.¹ Sistemi, ki temeljijo na umetni inteligenci so za namen odločanja poleg ekonomskih razlogov privlačni zaradi domnevne natančnosti, objektivnosti in nepristranskosti. V očeh mnogih umetna inteligenca predstavlja možnost objektivnega in zanesljivega odločevalca. Vendar ali je to res? Ali umetna inteligenca resnično predstavlja alternativo človeškemu odločanju, s tem, da v odsotnosti čustev in predsodkov nudi pravične odločitve?

V tem prispevku obravnavam tveganje diskriminacije pri avtomatiziranem odločanju, ki temelji na uporabi umetne inteligence. Za razumevanje problematike prvotno opredelim diskriminacijo in se osredotočim na vzrokediskriminacije pri tovrstnih sistemih. Poleg drugih ukrepov za boj zoper diskriminatorno avtomatizirano odločanje se posebej osredotočim na pravico, da za posameznika ne velja odločitev, ki temelji na avtomatizirani obdelavi podatkov. Namen tega prispevka je določiti, kdaj je odločitev avtomatiziranega sistema diskriminatorna, kaj so dejavniki, ki povzročajo takšno odločanje, kakšni so možni načini preprečitve takšnih odločitev in ali jih je mogoče oziroma smiselno s zakonodajno regulacijo preprečiti, ter kdaj avtomatizirane odločitve za posameznika, na katerega se nanašajo, sploh veljajo.

¹ Sartor, G.: The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, European Parliamentary Research Service, Bruselj, 2020, str. 1

5.2 Prepoved diskriminacije

5.2.1 Prepoved diskriminacije v Evropski konvenciji o varstvu človekovih pravic

Za določitev prepovedi diskriminacije je najbolj smiselno začeti na samem vrhu hierarhije veljavnih pravnih aktov, zato določitev prepovedi diskriminacije začenjam z Evropsko konvencijo o varstvu človekovih pravic²(v nadaljevanju EKČP), ki na podlagi 56. člena EKČP vsekakor velja na območju Evropske unije (v nadaljevanju EU). Prepoved diskriminacije je natančneje določena v 14. členu EKČP, ki določa, da je uživanje pravic in svoboščin določenih s EKČP »zagotovljeno vsem ljudem brez razlikovanja glede na spol, raso, barvo kože, jezik, vero, politično ali drugo prepričanje, narodnostni ali socialni izvor, pripadnost narodni manjšini, lastnino, rojstvo ali kakšne druge okoliščine.« Za razumevanje prepovedi diskriminacije je zelo pomembna praksa Evropskega sodišča za človekove pravice (v nadaljevanju Evropsko sodišče). V zadevi *Thlimmenos*³ je na primer Evropsko sodišče odločilo, da je ravnanje lahko diskriminatorno, če sta dve osebi obravnavani enako, kljub temu da je njun položaj popolnoma različen.

5.2.2 Prepoved diskriminacije na ravni Evropske unije

Prepoved diskriminacije in načelo enake obravnave sta trdno zasidrani v samih temeljih EU. V 2. členu Pogodbe o Evropski uniji (v nadaljevanju PEU)⁴ je določeno, da EU »temelji na vrednotah spoštovanja človekovega dostojanstva, svobode, demokracije, enakosti, pravne države in spoštovanja človekovih pravic, vključno s pravicami pripadnikov manjšin«⁵, ter da to skupno družbo držav članic označujejo pluralizem, nediskriminacija, strpnost, pravičnost, solidarnost ter enakost moških in žensk. Tovrsten princip je nadalje poudarjen v 2. odstavku 3. člena PEU, kjer je določeno, da se EU bori proti izključenosti in diskriminaciji.

² Zakon o ratifikaciji Konvencije o varstvu človekovih pravic in temeljnih svoboščin, spremenjene s protokoli št. 3, 5 in 8 ter dopolnjene s protokolom št. 2, ter njenih protokolov št. 1, 4, 6, 7, 9, 10 in 11 (Uradni list RS – Mednarodne pogodbe, št. 7/94).

³ *Thlimmenos v Greece*, št. 34369/97, ECLI:CE:ECHR:2000:0406JUD003436997.

⁴ Pogodba o Evropski uniji, Uradni list Evropske unije, C 326, 26.10.2012, str. 13–390.

⁵ *Ibidem*, člen 2.

Pravo EU, ki temelji na nedeljivih in univerzalnih vrednotah človekovega dostojanstva, svobode, enakopravnosti in solidarnosti⁶, prepoved diskriminacije kot temeljno pravico prebivalcev EU konkretizira v Listini Evropske unije o temeljnih pravicah (LEUTP)⁷. Prvi člen določa, da je človekovo dostojanstvo nedotakljivo, in da ga je potrebno spoštovati in varovati. Nadalje 20. člen LEUTP določa, da smo pred zakonom vsi enaki. Prepoved diskriminacije je neposredno določena v prvem odstavku 21. člena LEUTP, ki zapoveduje, da je prepovedana »vsakršna diskriminacija na podlagi spola, rase, barve kože, etničnega ali socialnega porekla, genetskih značilnosti, jezika, vere ali prepričanja, političnega ali drugega mnenja, pripadnosti narodnosti manjšini, premoženja, rojstva, invalidnosti, starosti ali spolne usmerjenosti.«⁸ Ta člen ni pomemben zgolj zaradi neposredne prepovedi diskriminacije, temveč zaradi jasne opredelitve lastnosti, na podlagi katerih bazirajo skupine, ki jih je v boju zoper diskriminacijo potrebno zaščititi (v nadaljevanju zaščitene skupine).

Prepoved diskriminacije je natančneje urejena v več direktivah: Direktiva sveta o izvajanju načela enakega obravnavanja oseb ne glede na raso ali narodnost,⁹ Direktiva sveta o splošnih okvirih enakega obravnavanja pri zaposlovanju in delu¹⁰, Direktiva Evropskega parlamenta in sveta o uresničevanju načela enakih možnosti ter enakega obravnavanja moških in žensk pri zaposlovanju in poklicnem delu¹¹. Direktiva sveta o izvajanju načela enakega obravnavanja moških in žensk pri dostopu do blaga in storitev ter oskrbi z njimi¹². Vsebina teh direktiv je v Slovenski pravni red prenesena z Zakonom o varstvu pred diskriminacijo (v nadaljevanju ZVarD).

⁶ Preambula Listine Evropske unije o človekovih pravicah, Uradni list Evropske unije, C 83/389, str. 391–407.

⁷ Ibidem.

⁸ Ibidem člen 21, prvi odstavek.

⁹ Direktiva Sveta 2000/43/ES z dne 29. junija 2000 o izvajanju načela enakega obravnavanja oseb ne glede na raso ali narodnost, Uradni list Evropske unije, L 180, 19.7.2000, str. 22–26.

¹⁰ Direktiva Sveta 2000/78/ES z dne 27. novembra 2000 o splošnih okvirih enakega obravnavanja pri zaposlovanju in delu, Uradni list Evropske unije, L 303, 2.12.2000, str. 16–22.

¹¹ Direktiva 2006/54/ES Evropskega parlamenta in Sveta z dne 5. julija 2006 o uresničevanju načela enakih možnosti ter enakega obravnavanja moških in žensk pri zaposlovanju in poklicnem delu (preoblikovano), Uradni list Evropske unije, L 204, 26.7.2006, str. 23–36

¹² Direktiva Sveta 2004/113/ES z dne 13. decembra 2004 o izvajanju načela enakega obravnavanja moških in žensk pri dostopu do blaga in storitev ter oskrbi z njimi, Uradni list Evropske unije, L 373, 21.12.2004, str. 37–43.

5.2.3 Prepoved diskriminacije v slovenski zakonodaji

Ustava Republike Slovenije v 14. člen določa, da »so vsakomur zagotovljene enake človekove pravice in temeljne svoboščine, ne glede na narodnost, raso, spol, jezik, vero, politično ali drugo prepričanje, gmotno stanje, rojstvo, izobrazbo, družbeni položaj, invalidnost ali katerokoli drugo osebno okoliščino.«¹³ Drugi odstavek istega člena dodaja da smo pred zakonom enaki vsi. S tem določilom je prepoved diskriminacije zapečaten v URS. V primeru, da je oseba deležna diskriminacije ima v 22. členu URS prav tako kakor kdor koli drug zagotovljeno enako varstvo pravic v postopku pred sodiščem in pred drugimi državnimi organi, organi lokalnih skupnosti in nosilci javnih pooblastil, ki odločajo o njegovih pravicah, dolžnostih ali pravnih interesih. Skladno s obravnavano temo je potrebno dodati, da je v 63. členu URS prepovedano vsakršno spodbujanje k narodni, rasni, verski ali drugi neenakopravnosti. S tem je na ustavni ravni tako kot diskriminacija prepovedano tudi vsakršno spodbujanje k diskriminaciji.

Kot sem omenil prej je vsebina Direktiv EU, ki urejajo prepoved diskriminacije, v slovenski pravni red vpeljana v ZVarD. »Ta zakon določa varstvo vsakega posameznika in posameznice (v nadaljnjem besedilu: posameznik) pred diskriminacijo ne glede na spol, narodnost, raso ali etnično poreklo, jezik, vero ali prepričanje, invalidnost, starost, spolno usmerjenost, spolno identiteto in spolni izraz, družbeni položaj, premoženjsko stanje, izobrazbo ali katero koli drugo osebno okoliščino«¹⁴ V 4. členu ZVarD je diskriminacija opredeljena kot »vsako neupravičeno dejansko ali pravno neenako obravnavanje, razlikovanje, izključevanje ali omejevanje ali opustitev ravnanja zaradi osebnih okoliščin, ki ima za cilj ali posledico oviranje, zmanjšanje ali izničevanje enakopravnega priznavanja, uživanja ali uresničevanja človekovih pravic in temeljnih svoboščin, drugih pravic, pravnih interesov in ugodnosti.« V drugem odstavku 4. člena prav tako piše, da je prepovedana diskriminacija zaradi katere koli osebne okoliščine. Enako obravnavanje je v 5. členu opredeljeno kot odsotnost neposredne ali posredne diskriminacije zaradi katere koli osebne okoliščine. Poleg neposredne in posredne diskriminacije ZVarD v 7. členu določa tudi druge oblike diskriminacije, namreč nadlegovanje in spolno nadlegovanje, navodila za diskriminacijo, pozivanje k

¹³ 14. člen Ustave Republike Slovenije (Uradni list RS, št. 33/91-I, 42/97 – UZS68, 66/00 – UZ80, 24/03 – UZ3a, 47, 68, 69/04 – UZ14, 69/04 – UZ43, 69/04 – UZ50, 68/06 – UZ121,140,143, 47/13 – UZ148, 47/13 – UZ90,97,99, 75/16 – UZ70a in 92/21 – UZ62a).

¹⁴ 1. člen prvi odstavek Zakona o varstvu pred diskriminacijo (Uradni list RS, št. 33/16 in 21/18 – ZNOrg).

diskriminaciji in povračilne ukrepe oziroma viktimizacijo. V 12. členu prav tako določa hujše oblike diskriminacije, za katere štejejo večkratna diskriminacija (oseba diskriminirana zaradi več osebnih okoliščin hkrati), množična diskriminacija (kadar je diskriminiranih več oseb hkrati), dolgotrajna oziroma ponavljajoča se diskriminacija in diskriminacija, ki vsebuje ali bi lahko vsebovala težko popravljive posledice za diskriminirano osebo glede povzročitve škode njenemu pravnemu položaju, pravicam ali obveznostim, zlasti če je storjena v razmerju do otrok ali drugih slabotnih oseb.

5.3 Kdaj je odločitev diskriminatorna?

Pojem tako neposredne kot tudi posredne diskriminacije je podan v drugem odstavku 2. člena Direktive sveta o izvajanju načel enakega obravnavanja oseb ne glede na raso ali narodnost, ki po določilu prvega odstavka tega člena obe predstavljata kršitev načela enakega obravnavanja, v smislu te direktive.

Za neposredno diskriminacijo se šteje, če se ena oseba obravnava manj ugodno, kakor se obravnava, se je obravnavala ali pa bi se obravnavala druga oseba v primerljivem položaju na podlagi rase ali narodnosti. Za posredno diskriminacijo se šteje, če bi na videz nevtralna določba, merilo ali praksa postavila osebe neke rase ali narodnosti v posebno neugoden položaj v primerjavi z drugimi osebami, razen če to določbo, merilo ali prakso objektivno upravičuje zakonit cilj in če so sredstva za doseganje tega cilja ustrezna in *potrebna*.¹⁵ Ne samo da je poleg neposredne diskriminacije prepovedana tudi posredna diskriminacija, ki je sicer pri algoritemski diskriminaciji bolj pogosta, za diskriminacijo ni potreben namen storilca. Zelo pomembno je poudariti, da prej omenjen člen določa, da ne gre za posredno diskriminacijo, če določbo, merilo ali prakso objektivno upravičuje zakonit cilj in če so sredstva za doseganje tega cilja ustrezna in potrebna. Kot je določeno v 3. členu, se ta direktiva uporablja v okviru pristojnosti, dodeljenih Skupnosti, za vse osebe v javnem in zasebnem sektorju.

Za določitev tako neposredne kakor posredne diskriminacije je potreben *primerjalnik*, posameznik ali skupina, katere situacija je podobna domnevni žrtvi diskriminacije, katere obravnavo lahko primerjamo z obravnavo domnevne žrtve diskriminacije, za ugotovitev oziroma utemeljitev diskriminacije. Na podlagi takšne

¹⁵ Direktiva 2000/43/ES člen 2, odstavek 2a

primerjave je Evropsko sodišče presojalo v zadevi *Moustaquim v Belgium*¹⁶, kjer je ugotovilo, da položaj Maročana, ki bi naj bil deportiran iz Belgije, ni primerljiv s položajem Belgijca, saj le ta na podlagi 3. člena Protokola št. 4 k EKČP, ki določa, da nihče ne sme biti izgnan iz ozemlja države, katere državljan je, ne sme biti izgnan. Priznalo pa je, da je položaj Maročana primerljiv s položajem državljanov EU. V zadevi *Test Achats*¹⁷ je Sodišče Evropske Unije (v nadaljevanju Sodišče) poudarilo, da v skladu z ustaljeno sodno prakso Sodišča načelo enakega obravnavanja terja, da se primerljivi položaji ne obravnavajo različno in da se različni položaji ne obravnavajo enako, razen če je tako obravnavanje objektivno upravičeno in da je primerljivost položajev potrebno presojati glede na cilj in namen akta Unije, s katerim je uvedeno zadevno razlikovanje.

ZVarD neposredno in posredno diskriminacijo opredeljuje v 6. členu. Določa namreč, da neposredna diskriminacija obstaja, »če je oseba ali skupina oseb zaradi določene osebne okoliščine bila, je ali bi lahko bila v enakih ali podobnih situacijah obravnavana manj ugodno, kot se obravnava, se je obravnavala ali bi se obravnavala druga oseba ali skupina oseb.«¹⁸ Za posredno diskriminacijo je določeno, da obstaja, »kadar je oseba ali skupina oseb z določeno osebno okoliščino bila, je ali bi lahko bila zaradi navidezno nevtralne določbe, merila ali prakse v manj ugodnem položaju kot druge osebe, razen če ta določba, merilo ali praksa objektivno temelji na legitimnem cilju in so sredstva za doseganje tega cilja ustrezna in *nujno potrebna*.« Tukaj vidimo, da ZVarD postavlja višji prag za presojo odsotnosti posredne diskriminacije, saj namesto da so *potrebna* (kot je določeno v prej omenjeni direktivi EU) za sredstva za doseganje legitimnega cilja (zaradi katerega je oseb s določenimi lastnostmi v manj ugodnem položaju) določa, da morajo biti *nujno potrebna*.

Na podlagi pojma posredne diskriminacije in v nadaljevanju opisanih načinov, kako do diskriminacije pri algoritemskem odločanju lahko pride, se na prvi pogled zdi, da je vsaka odločitev, ki temelji na statističnem modelu diskriminatorna, saj je za skoraj vsako upoštevano lastnost mogoče trditi, da je zaradi tovrstnega merila neka oseba, na podlagi določene osebne okoliščine v manj ugodnem položaju. To je vsekakor res, saj se v vsakem izboru oziroma odločanju, ki vpliva na posameznike, upoštevajo lastnosti teh posameznikov (in njim podobnih posameznikov), zaradi česar je skoraj neizbežna manj ugodna obravnava nekaterih, na podlagi določenih osebnih

¹⁶ *Moustaquim v Belgium*, št. 12313/86, ECLI:CE:ECHR:1991:0218JUD001231386.

¹⁷ Zadeva C-236/09, *Test-Achats*, ECLI:EU:C:2011:100

¹⁸ ZVarD 6. člen 1. odstavek.

okolščin. To ni zgolj problem avtomatiziranega odločanja, temveč tudi človeškega. Vselej kadar so postavljeni določeni kriteriji, nekateri posamezniki teh ne izpolnjujejo oziroma jih izpolnjujejo v manjši meri. Sam namen kriterijev namreč je tovrstna selekcija. Za resnično razumevanje, kdaj je odločanje na podlagi avtomatiziranega sistema, ozirom kdaj so merila, na podlagi katerih ta sistem presoja posredno diskriminatorna, se moremo osredotočiti na drugi del definicije posredne diskriminacije¹⁹, ki izključuje obstoj posredne diskriminacije v primeru, kadar domnevno sporno določbo, merilo ali prakso objektivno upravičuje zakonit cilj in kadar so sredstva za doseganje tega cilja ustrezna in potrebna.

Koncept posredne diskriminacije in zakonitega cilja je Sodišče obravnavalo v zadevi *Achbita*²⁰, kjer je presojalo ali je delodajalčeva nevtralna prepoved nošenja simbolov političnih, filozofskih ali verskih prepričanj na delovnem mestu diskriminatorna zoper zaposleno, ki je muslimanka in iz verskih razlogov nosi naglavno ruto. Sodišče je v tem primeru izključilo neposredno diskriminacijo, vendar kljub nevtralnemu pravilu, ki sicer zadeva vse vere in druga prepričanja, ni izključilo posredne diskriminacije, saj je tovrstna omejitev lahko za nekatere zaposlene bolj intruzivna kakor za druge (nekateri verski simboli se nosijo zgolj z namenom izkazovanja pripadnosti določeni veri, med tem ko ima nošenje nekaterih simbolov močno versko in moralno vrednost) s čimer lahko takšna omejitev nekatere zaposlene obravnava manj ugodno. S tem so utemeljili pogoj za obstoj posredne diskriminacije. Naslednje vprašanje je bilo, ali je delodajalec s tem pravilom zasledoval zakonit cilj. Delodajalec je to pravilo postavil z namen ohranitve nevtralnega izgleda podjetja, ob stiku zaposlenih s strankami. Sodišče je priznalo, da je tovrsten zasledovan cilj na prvi pogled zakonit, saj izhaja iz pravice do svobodne gospodarske pobude, določene v 16. členu LEUTP. Izpostavilo je pogoj, da je sredstvo za doseganje tega cilja ustrezno in potrebno. Glede ustreznosti je Sodišče odločilo, da bi jo bilo potrebno presojati glede na to, ali je pravilo bilo implementirano generalno in enako za vse, ter poudarilo, da je tovrstno pravilo nediskriminatorno kadar velja zgolj za delavce, ki dejansko so v stiku s stranki. Le tedaj je takšno pravilo potrebno (prepoved nošenja verskih simbolov pri zasledovanju cilja nevtralne podobe ni potrebna pri zaposlenih, ki pri delu niso v stiku s strankami).

¹⁹ 2000/43/ES člen 2, odstavek 2a

²⁰ C-157/15, *Achbita*, ECLI:EU:C:2017:203

Problem predpisa, ki prepoveduje posredno diskriminacijo je, da posredna diskriminacija pogosto ni jasna. Potrebno je dokazati, da sicer nevtralnno pravilo neproporcionalno prizadene določeno zaščiteno skupino. Po navadi se za dokazovanje neproporcionalnega učinka uporablja statistika.²¹ Prvi odstavek 8. člena Direktive sveta o izvajanju načel enakega obravnavanja oseb ne glede na raso ali narodnost določa, da morajo države skladno s svojim pravnim sistemom sprejeti ukrepe, s katerim zagotovijo, da mora sicer toženec dokazati, da ni bilo kršeno načelo enakega obravnavanja, vendar pod pogojem, da tožeča stranka predstavi dejstva, na podlagi katerih se lahko domneva, da je do tovrstne kršitve prišlo. V skladu s tem tudi ZVarD določa obrnjeno dokazno breme. Od diskriminirane osebe se zahteva zgolj, da izkaže dejstva, na podlagi katerih se lahko domneva, da je bila kršena prepoved diskriminacije. V tem primeru mora domnevni kršitelj dokazati, da ni kršil očitane prepovedi diskriminacije.

5.4 Kaj je algoritemaska diskriminacija?

Pojem algoritem je pogosto uporabljen v kontekstu umetne inteligence, kot »algoritemsko odločanje« ali v našem primeru »algoritemaska diskriminacija«. Koncept algoritma je sicer širši od umetne inteligence in sicer vključuje vsako zaporedje nedvoumno opredeljenih navodil za izvajanje nalog, zlasti, vendar ne izključno z matematičnimi izračuni. Med tem ko vsak algoritem ne vključuje umetne inteligence, vsak sistem umetne inteligence vključuje več algoritmov.²²

Algoritemsko diskriminacijo lahko, za namen tega prispevka, definiramo kot pojav, ki nastane, kadar avtomatiziran program za odločanje, ki temelji na umetni inteligenci, posameznika obravnava manj ugodno kakor drugega na primerljivem položaju, in je tovrstna negativna obravnava pogojena z ali v močni korelaciji z določeno njegovo osebno lastnostjo. Kadar je negativna obravnava neposredno pogojena oziroma povezana s osebno lastnostjo (se neposredno upošteva in negativno vrednoti) obravnavanega posameznika, gre za neposredno diskriminacijo. Kadar se pri odločanju upoštevajo lastnosti, ki so v močni korelaciji s posameznikovo osebno lastnostjo, ter to vodi do negativne obravnave posameznikov s to lastnostjo, tovrstno odločanje lahko predstavlja posredno

²¹ Borgesius, F. Z.: Discrimination, artificial intelligence, and algorithmic decision-making, Directorate General of Democracy, Council of Europe, Strasbourg, 2018, str. 19

²² Sartor, G.: The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, European Parliamentary Research Service, Bruselj, 2020, str. 17

diskriminacijo, če se te lastnosti ne presojuje z namenom zasledovanja zakonitega cilja in če niso nujne.

Urejanje virov nenamerne diskriminacije pri algoritemskem odločanju in odprava povezanih pomanjkljivosti v pravu bo tehnično, pravno in politično težavna. Diskriminacija je lahko produkt podatkovnega rudarjenja in ne tega, da programerji določenim pogojem določajo diskriminatorne vrednosti. Podatkovno rudarjenje je postopek odkrivanja zanimivih vzorcev in znanja na podlagi velikih količin podatkov²³ Ta nevarnost je pogosto spregledana s strani akademikov in zakonodajne oblasti, ki se pogosto bolj bojijo diskriminatornega programiranja tovrstnih sistemov. Podatkovno rudarjenje lahko predstavlja večjo grožnjo, saj kadar se diskriminacija pojavi zaradi podatkov, ki so posledica predhodne namerne diskriminacije, pogosto ni primerne metode za popravek historičnih predsodkov znotraj teh podatkov. Tudi kadar so podatkovni rudarji skrbni, lahko pride do diskriminatornih rezultatov z uporabo modelov, ki uporabljajo posredne podatke, ki odražajo zaščitene skupine. V nekaterih primerih bo pri podatkovnem rudarjenju nemogoče odpravljati napake brez vprašanja, kakšna stopnja neenakosti je sprejemljiva v določenem kontekstu tovrstnih sistemov. Podatkovno rudarjenje je vselej vrsta statistične in tako na prvi pogled racionalne diskriminacije. Smisel tega procesa je ustvariti racionalno podlago na kateri lahko posameznike ločimo in jih klasificiramo v skupine na podlagi določenih lastnosti. Vendar s tem ta proces omogoča nepravilno obravnavo z zakonom zaščitene skupin. Z razliko od klasičnih podatkovnih metod, kjer analiza preprosto vrednoti določene podatke ali ustvari statistiko, podatkovno rudarjenje skuša najti statistične povezave znotraj baze podatkov. Gre za avtomatiziran proces odkrivanja uporabnih vzorcev, na podlagi katerih se kasneje lahko tvorijo odločitve. Skupek odkritih povezav se imenuje model. S tem ko algoritem strojnega učenja izpostavimo primerom zanimanja, algoritem odkrije lastnosti teh primerov, na podlagi katerih lahko kasneje prepozna in oceni podobne ali enake primere.²⁴

Uporaba avtomatiziranih sistemov za odločanje ni privlačna zgolj zaradi tega, ker so napovedi in odločitve teh sistemov cenejše, temveč zato, ker so bolj precizne in nepristranske kakor človeške, saj ti sistemi niso podrejeni tipičnim človeškim psihološkim »napakam« in so lahko podvrženi rigoroznemu nadzoru. Kljub naravi

²³ Han, J. Pei, J. Kamber, M.: *Data Mining: Concepts and Techniques*, 2. izdaja, Morgan Kaufman Publishers, San Francisco, 2006, str. 6

²⁴Barocas, S., Selbst, A., D.: *Big Data's Disparate Impact*, v: *California Law Review*, 104 (2016) 617, str. 672-678

teh sistemov, so njihove odločitve lahko napačne in diskriminatorne. V tem primeru lahko gre za reproduciranje človeške pristranskosti ali ustvarjanje novi pristranskosti.²⁵

Vselej je potrebno upoštevati, da sistemi umetne inteligence sami po sebi ne nastanejo in sami po sebi nimajo negativnega vpliva na človekove pravice temveč ima takšen vpliv njihova implementacija in uporaba za človeško interakcijo. Za vsako odločitvijo avtonomnega sistema stoji fizična ali pravna oseba, ki ta sistem uporablja za določen namen.²⁶ Uporaba algoritmov za namen odločanja nadomesti samovoljno človeško odločanje s strogimi pravili, konstruiranimi iz podatkov. Prav tako je za razliko od človeškega odločanja, ekskluzija določenih podatkov (posameznikova rasa, spol, itd.) v primeru algoritemskega odločanja lahko zagotovljena.²⁷

5.5 Vzroki algoritemske diskriminacije

Preden lahko govorimo o preprečevanju algoritemske diskriminacije moremo razumeti kako do tovrstne diskriminacije pride. Algoritemska diskriminacija je lahko posledica človekovega vpliva na postopek podatkovnega rudarjenja na pet različnih načinov.

5.5.1 Določanje ciljne spremenljivke in razredov

Med tem ko pri podatkovnem rudarjenju ciljna spremenljivka določa tisto, kar želimo ugotoviti oziroma nas zanima (ali je kandidat za delovno mesto dober ali slab), razredi razdelijo vse možne vrednosti ciljne spremenljivke v ekskluzivne kategorije (izobrazba, izkušnje,...). Pri podatkovnem rudarjenju skušamo problem rešiti tako, da postavimo vprašanje, katerega odgovor predstavlja ciljna spremenljivka. Pomembno je, da vprašanje opredelimo tako, da ustreza izmerljivim vrednostim. Ker je postopek določitve tega vprašanja subjektivne narave, se lahko zgodi, da tisti, ki to določajo nenamerno postavijo vprašanje, ki slabše obravnava člane zaščitene skupin. Ciljna spremenljivka in razredi določajo, kaj bo postopek

²⁵ Sartor, G.: The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, European Parliamentary Research Service, Bruselj, 2020, str. 5

²⁶ Borgesius, F. Z.: Discrimination, artificial intelligence, and algorithmic decision-making, Directorate General of Democracy, Council of Europe, Strasbourg, 2018, str.9

²⁷ Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: University of Chicago Law Review, 86 (2019) 459, str. 460-463.

podatkovnega rudarjenja odkril. Med tem ko se mnogi kritiki podatkovnega rudarjenja osredotočajo na problem nepravilne klasifikacije (lažna pozitivna ali negativna določitev), vsaj takšno ali celo večjo nevarnost diskriminacije predstavlja slaba in diskriminatorna določitev razredov in s tem vrednotenje učnih primerov oziroma njihovih lastnosti, na podlagi katerih se tvorijo pravila.²⁸ Nekateri sicer na prvi pogled popolnoma objektivni razredi lahko odsevajo zgodovinski ali trenutni slabši položaj nekaterih skupin in s tem negativno vplivajo na odločanje glede posameznikov teh skupin.²⁹

Problem določanja primernih ciljnih spremenljivk je, da vsak argument za superiornost določene ciljne spremenljivke v primerjavi drugimi izzove spor glede kontradiktornih in ne združljivih vrednot.

5.5.2 Učni primeri za strojno učenje

Sistem podatkovnega rudarjenja se uči na podlagi učnih primerov, natančneje tistih, katerim je bil izpostavljen. Učne primere tvorijo podatki, ki jih sistem podatkovnega rudarjenja analizira in se na njihovi podlagi uči. Do diskriminatornega rezultata lahko pride, ko se sistem uči iz pristranskih učnih primerov. Do tega lahko pride na dva načina. Prvi je, ko se sistem uči na podlagi primerov, ki so plod diskriminacije ali predsodkov. To je problematično, saj sistem podatke, iz katerih se uči tretira kot resnične oziroma pravilne, zaradi česar bo pravilo, ki bo nastalo na podlagi tovrstnih primerov zrcalilo diskriminacijo iz teh primerov. Drugi način oziroma vzrok je nereprezentativnost podatkov iz katerih se sistem uči. Če podatki, ki tvorijo učne primere predstavljajo le del populacije oziroma ta del predstavljajo prekomerno, je rezultat lahko diskriminatoren zoper del populacije, ki je v teh podatkih pomanjkljivo ali prekomerno predstavljen.³⁰

Problematičen je lahko tudi proces označevanja primerov, pri katerem se učnim primerom manualno dodelijo razredi. Tudi ta proces je subjektiven in lahko na rezultat oziroma pravilo vpliva v zadostni meri, da bodo posledično odločitve na podlagi tega pravila nadaljnje primere karakterizirale na podoben način. To je pomembno, saj kadar odločitve, ki so posledica predsodkov služijo kot primer

²⁸ Barocas, S., Selbst, A., D.: Big Data's Disparate Impact, v: California Law Review, 104 (2016) 617, str. 678-680

²⁹ Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: University of Chicago Law Review, 86 (2019) 459, str. 471

³⁰ Barocas, S., Selbst, A., D.: Big Data's Disparate Impact, v: California Law Review, 104 (2016) 617, str. 680-681.

pravih odločitev, sistem podatkovnega rudarjenja ustvarja pravila, ki odsevajo te predsodke in so diskriminatorna.³¹

Pri podatkovnem rudarjenju ne gre zgolj za nevarnost podedovanja predhodne diskriminatorne pristranskosti, temveč lahko odseva trenutno obstoječe predsodke in diskriminatorno stanje v družbi. Ker se postopek podatkovnega rudarjenja zanaša na učne primere kot resnično podlago, bo v primeru, da so ti primeri zastrupljeni s predsodki ali odsevajo diskriminatorno stanje, rezultat nezanesljiv ali celo diskriminatoren.³² Pri tem se lahko pojavi konstantno dopolnjevanje zaradi česar je rezultat nezanesljiv ali celo diskriminatoren.

5.5.3 Zbiranje podatkov

Veliko pozornosti je potrebno posvetiti samemu zbiranju podatkov, ki služijo kot učni primeri. Učenje na podlagi napačnih, delnih ali nereprezentativnih podatkov lahko vodi do odločitev, ki imajo diskriminatoren učinek zoper zaščitene skupine. Določena baza podatkov lahko zajema neproporcionalno manjše število podatkov o posameznikih iz določene zaščitene skupine ali pa so podatki, ki jih zajema o članih te skupine nižje kvalitete. Nekateri so v teh bazah podatkov nenamerno pozabljeni, saj zaradi revščine, lokacije ali načina življenja o njih obstaja manjša količina podatkov, zaradi česar so nekatere skupnosti in državljani masovno spregledani.³³ Tukaj vidimo problem, namreč tovrstne odločitve ne zadevajo vedno zgolj posameznikov, katerih podatki so se zbirali oziroma, ki so prispevali podatke, temveč tudi tistih, katerih podatki se niso zbrali.

Podatkovno rudarjenje je še posebej občutljivo iz vidika statistične pristranskosti, saj vzorce oziroma pravila, ki jih tovrsten sistem odkriva uporabniki tretirajo kot generalna odkritja kljub temu, da analizirani podatki vključujejo zgolj delni vzorec relevantne populacije, zajet v določenem omejenem časovnem obdobju. Reprezentativnost teh podatkov se pogosto preprosto domneva, vendar je ta domneva pogosto napačna. Efektivnost podatkovnega rudarjenja je pogojena z kvaliteto podatkov, ki služijo kot učni primeri. Če baza podatkov ne zajema kvalitetnega in korektnega vzorca podatkov, ki zadevajo določeno skupino, je rezultat lahko diskriminatoren zoper to skupino. Prav tako je lahko škodljiva

³¹ Ibidem, str. 681-682.

³² Ibidem, str. 682-683.

³³ Ibidem, str. 684-685.

prekomerna reprezentacija določene skupine v bazi učnih primerov, saj lahko vodi do neproporcionalno visoke količine negativnih odločitev za člane te skupine.³⁴ Povedano drugače, kadar določeno skupino ljudi opazujemo bolj pogosto oziroma dalj časa, pri njej zabeležimo večje število opazovanih lastnosti.

Kadar gre za pomanjkanje podatkov o določeni skupini ljudi lahko podatkovni rudarji proaktivno pomanjkanje podatkov nadomestijo z pretiranim vzorčenjem te skupine in s tem nekoliko zmanjšajo pristranskost rezultata. Podobno je mogoče sicer storiti tudi retroaktivno, kadar zaznamo, da je rezultat diskriminatoren zaradi nereprezentativnosti. Ker sta proaktivno pretirano vzorčenje in retroaktivno popravljanje baze podatkov mogoči, nudi kljub mnogim pomanjkljivostim tovrsten pristopom pozitivne manipulacije učnih primerov najbolj perspektivni in učinkoviti način odpravljanja negativnih rezultatov, ki nastanejo v vseh mehanizmih podatkovnega rudarjenja, ki sem jih predstavil.³⁵ V primerih pretirane ali borne reprezentativnosti podatkov o določeni skupini je izključitev (zaščitene) lastnosti, ki definira to skupino, za namen odločanja, lahko škodljiva, saj v tem primeru druge lastnosti ocenjujemo skupno za predstavnike zaščitene skupine in druge, kljub temu, da so modeli ustvarjeni na gmoti podatkov, ki so za zaščiteno skupino nereprezentativni. S tem ko algoritem upošteva lastnost, ki določa zaščiteno skupino, lahko prilagodi težo lastnosti, ocena katerih temelji na analizi nereprezentativnih podatkov.³⁶

5.5.4 Določitev analiziranih lastnosti

Za proces podatkovnega rudarjenja je potrebno izbrati, katere lastnosti mora sistem opazovati, saj na njih temelji odločitev. Ta del postopka je zelo pomemben, saj je subjektivne narave in izbor lastnosti, ki ugaja določenim skupinam ali zgolj večini, lahko vodi do diskriminacije zaščitene skupine. Problem je zasidran v dejstvu, da so obravnavani podatki reducirana predstavitev resničnega življenja in zato pogosto pomanjkljivi za ugotovitev kritičnih lastnosti. Nemogoče je zajeti in v modele vključiti vse lastnosti posameznika in vse okoljske vidike. Odločitve, ki temeljijo na statističnih vendar neuniverzalnih generalizacijah so racionalne, vendar hkrati pogosto nepošteno.³⁷ Možna je formalna izločitev prepovedanih lastnosti, na primer

³⁴ Ibidem, str. 686-887

³⁵ Ibidem, str. 718-719

³⁶ Gillis, T. B.: False Dreams of Algorithmic Fairness: The Case of Credit Pricing, str. 50

³⁷ Barocas, S., Selbst, A., D.: Big Data's Disparate Impact, v: California Law Review, 104 (2016) 617, str. 688

rase ali spola, s čimer bi izključili vplive teh lastnosti, ki niso povezani z drugimi lastnostmi.³⁸ Tovrstna prepoved uporabe določenih lastnosti pri odločanju je še posebej privlačna, saj jih je pri avtomatiziranih sistemih mogoče formalno izključiti, med tem ko to ni mogoče pri človeškem odločanju, pri katerem je določena lastnost, kljub prepovedi, dejansko lahko upoštevana. Gre za stalno problematiko protidiskriminacijske zakonodaje. Pri človeškem odločanju je namreč težavno dokazati, da se določena lastnost ni upoštevala.

Problem je tudi to, da sistem umetne inteligence lahko razvije nove diskriminatorne skupine z določenimi lastnostmi, ki sicer ne spadajo pod neposredno zaščitene skupine, vendar vseeno nepravilno razlikujejo ljudi in povzročajo socialno neenakost.³⁹ Opozoriti moremo, da ima izključitev za odločanje občutljivih lastnosti (na podlagi katerih temeljijo zaščitene skupine) omejen učinek v smislu preprečevanja diskriminacije, in zadostuje zgolj omejenemu vidiku protidiskriminacijskega prava, saj so drugi podatki lahko v močni korelaciji s tovrstnimi občutljivimi lastnostmi⁴⁰(več v nadaljevanju o posrednih podatkih). Izključitev občutljivih lastnosti morda ni popolnoma zaželjena, saj vključitev tovrstnih lastnosti lahko omogoči algoritmu, da popravi diskriminatoren vpliv drugih podatkov, ki na zaščiteni skupino vplivajo pristransko.⁴¹ Omejitev analiziranih lastnosti je prav tako problematična, saj znižuje natančnost napovedi.⁴²

Poseben pojav oziroma vrsto diskriminacije, ki jo moramo tukaj izpostaviti je tako imenovan »redlining«. Redlining označuje zavrnitev storitev osebi, ki živi na določenem območju (ali na podlagi druge navidezno nediskriminatorne lastnosti), kljub temu, da posameznik izpolnjuje pogoje za tovrstno storitev (neodobritev kredita sicer kreditno sposobni osebi, ker živi v ekonomsko šibkejši soseski). Izraz se navezuje na domnevno prakso hipotekarnih posojilodajalcev, po kateri naj bi na zemljevidih z rdečimi črtami označevali soseske, katerih prebivalcem ne želijo odobriti kreditov.⁴³ Posebni primeri redlininga so primeri racionalnega rasizma. Gre za primere, kjer se za podatkovno rudarjenje rasa kot analizirana lastnost eksplicitno upošteva, vendar ne

³⁸ Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: University of Chicago Law Review, 86 (2019) 459, str. 468

³⁹ Borgesius, F. Z.: Discrimination, artificial intelligence, and algorithmic decision-making, Directorate General of Democracy, Council of Europe, Strasbourg, 2018, str., 5

⁴⁰ Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: University of Chicago Law Review, 86 (2019) 459, str. 464

⁴¹ Ibidem, str. 471-472

⁴² Gillis, T. B.: False Dreams of Algorithmic Fairness: The Case of Credit Pricing, str. 43

⁴³ Federal Fair Lending Regulations and Statutes Fair Housing Act, str 1

zlonamerno. V teh primer je razlog za vključitev rase kot analizirane lastnosti v tem, da se v tej lastnosti nahaja določena statistično generalizirana lastnost, do katere uporabniki sistema na ravni posameznika nimajo dostopa ali je ta težaven (zdravstvena kartoteka, statistično je določena skupina bolj izpostavljena določenim zdravstvenim stanjem). Gre za uporabo hitro dosegljivih generaliziranih podatkov, saj je dostop do specifičnih podatkov otežen ali onemogočen.⁴⁴ Ekonomisti tovrstno uporabo imenujejo tudi statistična diskriminacija. Pri takšni uporabi zaščitena lastnosti sama po sebi ni stvar interesa, temveč algoritem presoja na podlagi lastnosti, ki so v visoki statistični korelaciji s to lastnostjo.⁴⁵

5.5.5 Posredni podatki ali »proxy«

Pri posrednih podatkih govorimo o tistih podatkih, ki so sami po sebi nevtralni, vendar lahko delijo ljudi v razrede. Določena lastnost, ki se opazuje je lahko manj ali bolj pogosta v določenih zaščitene skupinah. Gre za problem, ki mu raziskovalci rečejo odvečni zapis. Govorimo o primerih, ko je posameznikova pripadnost določeni skupini prepoznavna iz nekega drugega podatka. Do tega pride kadar je določen podatek oziroma lastnost v močni korelaciji z določeno skupno.⁴⁶ Informacija glede posameznikove zaščitene lastnosti je vtisnjena v njegovih drugih podatkih, zaradi česar je algoritmu lahko znana (oziroma jo ta predvideva), tudi kadar je formalno izključena.⁴⁷

Efektivni način preprečevanja diskriminacije zaradi posrednih podatkov je zmanjšanje natančnosti njihovih determinacij.⁴⁸ Podatki, ki so povezani z zaščitnimi lastnostmi, lahko služijo kot relevantne informacije za odločanje in hkrati kot posredni podatki, za zaščiteno in formalno izključeno lastnost. Sposobnost algoritma, da odkrije zaščiteno lastnost na podlagi drugih podatkov je manj problematična, kadar upoštevanje zaščitene lastnosti služi zgolj za obravnavo drugih lastnosti (prej omenjena statistična diskriminacija). Kadar zaščitena lastnost ni stvar presoje sama po sebi oziroma ni stvar interesa, bo algoritem neposredno iskal posredne podatke za določitev podatkov, za določitev katerih se sicer uporablja korelacija z zaščiteno lastnostjo, ter s tem ignoriral odkrito zaščiteno lastnost. Zaradi

⁴⁴ Barocas, S., Selbst, A., D.: Big Data's Disparate Impact, v: California Law Review, 104 (2016) 617, str. 690

⁴⁵ Gillis, T. B.: False Dreams of Algorithmic Fairness: The Case of Credit Pricing, str. 46

⁴⁶ Barocas, S., Selbst, A., D.: Big Data's Disparate Impact, v: California Law Review, 104 (2016) 617, str. 691-692

⁴⁷ Gillis, T. B.: False Dreams of Algorithmic Fairness: The Case of Credit Pricing, str. 42.

⁴⁸ Borgese, F. Z.: Discrimination, artificial intelligence, and algorithmic decision-making, Directorate General of Democracy, Council of Europe, Strasbourg, 2018, str. 13.

tega bo z razvojem tovrstnega sistema postala uporaba zaščitenih lastnosti nepotrebna.⁴⁹

Še ena perspektivna metoda preprečevanja algoritemske diskriminacije zaradi posrednih podatkov je ortogonalizacija vhodnih podatkov. Gre za statistično metodo, ki v tem kontekstu pomeni, da vstopni podatki, ki so v močni korelaciji z zaščitenimi lastnostmi, ne služijo kot posredni podatki za te zaščitene lastnosti. Takšna metoda je še posebej učinkovita, kadar je nek podatek v korelaciji tako s zaščitenimi lastnostjo, kakor tudi s neko drugo lastnostjo, ki je objektivno pomembna za natančno napoved. Pri takšnem statističnem pristopu ločimo fazo učenja in fazo odločanja. Pri fazi učenja zaščiteni lastnosti (npr. raso) upoštevamo kot opazovano lastnost. Na podlagi rezultata te faze ocenimo težo, ki jo ta lastnost ima pri odločanju. Pri postopku odločanja pa te zaščitene lastnosti ne upoštevamo, oziroma se algoritem te lastnosti ne zaveda. Na ta način, kljub temu, da se je algoritem učil s upoštevanjem zaščitenih lastnosti, pri odločanju te lastnosti ne upošteva. S tem je zmanjšan negativen učinek uporabe posrednih podatkov, saj ta metoda izloči neposreden vpliv zaščitenih lastnosti na odločitev. Na žalost je uporaba te metode v kontekstu strojnega učenja težavna, saj temelji na tem, da algoritmu lažemo oziroma ga zavajamo. Pri tej metodi namreč od algoritma zahtevamo, da optimizira napoved ob upoštevanju določene lastnosti, kasneje pri fazi odločanja pa ga prikrajšamo podatkov glede te lastnosti. To ni problem v fazi, ko zgolj ugotovljamo učinek te lastnosti na odločitev, vendar je problematično kasneje v fazi odločanja, saj je napoved algoritma optimizirana za uporabo te lastnosti.⁵⁰

5.5.6 Zakrivanje namerne diskriminacije

Velik problem podatkovnega rudarjenja je, da se z uporabo katerega koli od prej naštetih mehanizmov lahko zakrije namerna diskriminacija. Vsaka od prej omenjenih oblik diskriminacije je namreč lahko namerno izvedena. Možna je na primer uporaba redlininga brez kakršnih koli ekonomskih razlogov, namreč iz popolnoma diskriminatornih razlogov. Ob uporabi s strani nekoga, ki želi odločiti diskriminatorno, je podatkovno rudarjenje nevarno orodje, saj lahko odkrije, da je posameznik član določene skupine, kljub temu, da ta lastnost nikjer ni navedena. To je mogoče z analizo podatkov, ki imajo močno korelacijo z določeno skupino,

⁴⁹ Gillis, T. B.: False Dreams of Algorithmic Fairness: The Case of Credit Pricing, str. 46.

⁵⁰ Ibidem, str. 63-68.

oziroma posrednimi podatki. Analiza teh podatkov omogoča namerno diskriminacijo, tudi kadar upravljalci sistema nimajo dostopa do občutljivih informacij. Prav tako prikrije dejstvo, da so odločevalci odločali na podlagi občutljivih lastnosti.⁵¹

5.6 Obstoječi in možni zaščitni ukrepi za boj zoper algoritemsko diskriminacijo

Sistemi umetne inteligence vsekakor lahko delujejo diskriminatorno, vendar to ne pomeni, da iz tega vidika odločajo slabše kot ljudje. V algoritemskem odločanju pogosto vidimo možnost odločanja, ki formalno ne bazira na zaščitnih lastnostih, vendar ima tovrstno odločanje pomembne omejitve.⁵² Sistem umetne inteligence pogosto diskriminira zato, ker je bil učen na podlagi podatkov, ki odsevajo človeško diskriminacijo. Potrebno je postaviti vprašanje, ali je naš cilj, da tovrsten sistem odloča kot človek ali popolnoma nediskriminatorno.⁵³ Menim, da je načeloma boljše strmeti k popolnoma nediskriminatornemu sistemu, še posebej kadar je sam cilj uporabe tovrstnega sistema najti odgovor, ki je na podlagi izmerljivih kvalitet, iz vidika opazovanih lastnosti objektivni in natančen. Vendar zaradi človeške vloge pri ustvarjanju teh sistemov, oziroma kot sem pojasnil prej, neizbežnega dejstva, da so primeri iz katerih se sistem uči pogosto zastrupljeni s človeško subjektivnostjo, ki pogosto temelji na diskriminatornih predsodkih in pristranskosti, je obstoj popolnoma nediskriminatornega sistema vprašljiv. Kljub temu, da stremimo k sistemom, ki ne diskriminirajo, moremo predvideti tovrstno diskriminacijo, podobno kot pri ljudeh, ter v ta namen implementirati določene preventivne regulacije, sisteme varnostnih mrež in možnosti restitucije.

Najbolj pogosti in učinkoviti orodji za preprečevanje diskriminacije s strani umetne inteligence sta protidiskriminacijska zakonodaja in zakonodaja za zaščito podatkov. Protidiskriminacijske določbe so pomanjkljive v smislu, da določajo omejeno količino zaščitnih skupin, med tem ko umetna inteligenca ustvarja nove diskriminatorne skupine. Zakonodaja o zaščiti podatkov zmanjšuje tveganje tovrstne diskriminacije, na primer tako, da zahteva transparentnost glede obdelave

⁵¹ Barocas, S., Selbst, A., D.: Big Data's Disparate Impact, v: California Law Review, 104 (2016) 617, str. 692-693.

⁵² Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: University of Chicago Law Review, 86 (2019) 459, str. 468.

⁵³ Borgesius, F. Z.: Discrimination, artificial intelligence, and algorithmic decision-making, Directorate General of Democracy, Council of Europe, Strasbourg, 2018, str. 18.

osebnih podatkov. Smernice so prav tako dobro orodje saj se hitro spreminjajo in so lahko specifične.⁵⁴

Problem preprečevanja diskriminacije v digitalni dobi je, da protidiskriminacijska zakonodaja temelji na zastareli paradigmi kavzalnosti, pri čemer kot relevantno opazujemo (in dokazujemo) ali je imela posameznikova zaščitena lastnost (oziroma upoštevanje te lastnosti) vpliv na odločitev. To je problematično saj strojno učenje temelji na korelaciji in ne kavzalnosti. Pri uporabi algoritmov za napoved, je cilj čim večja natančnost napovedi in le ta določa kvaliteto sistema. Zaradi tega protidiskriminacijska regulacija ne more temeljiti na tradicionalni analizi vzročnosti.⁵⁵ Ta ni učinkovita zaradi možnosti formalne izključitve lastnosti, ki določajo skupino, ki je domnevno diskriminirana, med tem ko s tovrstno izključitvijo neugodna obravnava ni izključena, zaradi korelacije te skupine z drugimi upoštevanimi lastnostmi. Dokazovati smiselno oziroma za diskriminacijo relevantno korelacijo je zelo težko, saj je vpogled v samo odločanje avtomatiziranih sistemov pogosto tehnično nemogoč. Domnevni žrtvi diskriminacije s tem preostane zgolj dokazovanje na podlagi statistike odločitev, vendar ta ne more služiti kot dokaz, saj je lahko plod obravnavane populacije in ne diskriminatornega postopka odločanja.

Velik del protidiskriminacijske pravne doktrine se osredotoča na regulacijo vhodnih podatkov, kar sledi logiki tradicionalne protidiskriminacijske zakonodaje, s tem da omejuje to, kar gre v algoritem. Tovrstni ukrepi so neprimerni zaradi treh razlogov. Pogosto ne izpolnjujejo lastnih ohlapnih ciljev pravičnosti, ni jih mogoče praktično implementirati oziroma niso primerni za sisteme strojnega učenja in so lahko za zaščitene skupine celo škodljivi in prispevajo k diskriminatorni obravnavi.⁵⁶ Kljub temu, da strmino k avtomatiziranim sistemom z željo transparentnega in objektivnega odločanja, se moremo zavedati, da ima zgolj regulacija upoštevanih lastnosti omejen učinek. Res je, da pri tovrstnih sistemih lahko dosežemo dejansko izključitev določenih lastnosti za namen odločanja, vendar če nas resnično skrbi rezultat teh sistemov, moremo iz vidika protidiskriminacijske zakonodaje strmeti onkraj preprostih omejitev upoštevanih lastnosti, saj imajo te omejitve omejen ali celo negativen učinek.⁵⁷ Meja med empirično relevantnim in pravno dopustnim nikoli ne izgine, med tem ko postopoma izginja razlika med sistemom, ki se zaveda

⁵⁴ Ibidem, str. 5-33.

⁵⁵ Gillis, T. B.: False Dreams of Algorithmic Fairness: The Case of Credit Pricing, str. 43.

⁵⁶ Ibidem, str. 42.

⁵⁷ Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: University of Chicago Law Review, 86 (2019) 459, str. 473.

zaščitene lastnosti in tistim, ki se ne, zaradi česar je prepoved vhodnih informacij nesmiselna.⁵⁸

Možen način regulacije vhodnih podatkov je določitev dopustnih lastnosti za odločanja. Za razliko s prej omenjeno regulacijo lastnosti, ki se jih za namen odločanje ne bi smelo upoštevati, tovrstna regulacija določa lastnosti, ki se lahko uporabljajo za odločanje in prepoveduje uporabo drugih lastnosti. Namesto določitve nedopustnih lastnosti, določa dopustne. Pri tem konceptu se ponovno pojavi problem določitve dopustnih lastnosti. Če je namreč namen regulacije uporaba zgolj relevantnih informacij, je takšna omejitev nesmiselna, saj sistem, katerega cilj je ustvarjati čim bolj natančne napovedi, za razliko od človeka, sam najboljše presodi, katere značilnosti so najbolj relevantne za natančno napoved. Omejitev analiziranih lastnosti iz tega vidika zgolj zmanjšuje kvaliteto napovedi.⁵⁹ Zaradi specifične narave odločitev za različne namene bi za takšno regulacijo bili primerni zgolj sektorski regulatorji.

5.7 Pravica, da za posameznika ne velja odločitev, ki temelji na avtomatki obdelavi podatkov

Zelo pomembna pravica za boj zoper algoritemsko diskriminacijo je določena v prvem odstavku 22. člena Uredbe o varstvu posameznikov pri obdelavi osebnih podatkov in o prostem pretoku takih podatkov ter o razveljavitvi Direktive 95/46/ES⁶⁰ (v nadaljevanju GDPR), namreč pravica, da za posameznika, na katerega se nanašajo posebni podatki ne velja odločitev, ki temelji zgolj na avtomatizirani obdelavi podatkov, vključno z oblikovanjem profilov, ki ima pravne učinke v zvezi z njim ali na podoben način nanj znatno vpliva.⁶¹ Oblikovanje profila generalno pomeni določitev določenih lastnosti posameznika ali skupine in obravnava tega posameznika ali skupine obziroma na te lastnosti. Pravica, ki jo določa 22. člen GDPR, primarno vpliva na drugi del, oblikovanje profila in s tem služi kot varovalo pred kategorizacijo, oceno in diskriminacijo posameznikov.⁶²

⁵⁸ Gillis, T. B.: False Dreams of Algorithmic Fairness: The Case of Credit Pricing, str. 49.

⁵⁹ Ibidem, str. 58-60.

⁶⁰ Uredba (EU) 2016/679 Evropskega parlamenta in Sveta z dne 27. aprila 2016 o varstvu posameznikov pri obdelavi osebnih podatkov in o prostem pretoku takih podatkov ter o razveljavitvi Direktive 95/46/ES (Splošna uredba o varstvu podatkov) (Besedilo velja za EGP), Uradni list Evropske unije, L 119, 4.5.2016, str. 1–88.

⁶¹ 22. člen GDPR.

⁶² Mendoza, I., Bygrave, L. A., The Right Not To Be Subject to Automated Decisions Based on Profiling, v: Universitz of Oslo Faculty of Law Legal Studies Research Paper Series, (2017) 2017-20, str. 1

Preprosta kategorizacija te določbe kot pravice ni popolnoma natančna, saj sooznačuje pravico, katero lahko uveljavlja nosilec pravice, oziroma oseba na katero se nanašajo osebni podatki. Iz stroge besedne razlage prvega odstavka 22. člena vidimo, da prej opredeljen posameznik nima pravice nečesa storiti, temveč pravico, da se ga na določen način ne obravnava, kar vodi do razumevanja tega člena kot določene prepovedi in ne pravice, ki jo zaščiten posameznik mora uveljavljati. Če bi to določilo razlagali kot posameznikovo pravico do pritožbe na tovrstno odločanje, bi njen učinek bil odvisen od uveljavljanja pravice s strani posameznika, če pa določilo razlagamo kot kvalificirano prepoved, jo tisti, ki izvaja tovrstno odločitev more spoštovati neodvisno od nosilca pravice.⁶³ V primeru, da bi bilo zahtevano uveljavljanje te pravice s strani posameznika, na katerega se osebni podatki v obdelavi nanašajo, bi tovrsten zaščitni ukrep imel manjšo moč, saj bi bil odvisen ne le od posameznikovega poznanja te pravice, temveč tudi njegovega zavedanja, da odločitev na podlagi njegovih osebnih podatkov sprejema avtonomen sistem. Tovrstno interpretacijo je potrdila tudi delovna skupina WP29⁶⁴ in zagotovila, da ta prepoved, ki se navezuje na avtomatizirano določanje, posameznika ščiti ne glede na njegovo uveljavljanje te pravice.

Kot lahko izberemo iz prvega odstavka 22. člena, je ta pravica določena pod tremi pogoji:

1. Sprejeta mora biti odločitev
2. Pri določanju so bili uporabljeni osebni podatki
3. Odločitev temelji zgolj na avtomatizirani obdelavi, vključno z oblikovanjem profilov

Prvi odstavek 22. člena zadeva odločitve, ki so sprejete zgolj na podlagi avtomatizirane obdelave, kar pomeni, da je do odločitve prišlo brez človeške udeležbe. Tej določbi se ne da izogniti zgolj s fiktivno človeško udeležbo, na primer tako, da človek odločitev zgolj posreduje, saj v tem primeru še vedno gre za odločitev, ki temelji na avtomatizirani obdelavi. Kriterij človeške udeležbe je zadoščen, kadar upravljalec zagotovi, da je človeški nadzor na postopek odločanja smiseln in izveden s strani nekoga, ki je kompetenten in ima avtoriteto odločitev

⁶³ Ibidem.

⁶⁴ Delovna skupina ustanovljena na podlagi 29. člena Direktive 95/46/ES, Article 29 Data Protection Working Party.

spremeniti. Primerno bi bilo, da upravljalec kot del ocene učinka v zvezi z varstvom podatkov (v nadaljevanju DPIA⁶⁵) določi in zabeleži tovrstno človeško udeležbo.⁶⁶

Pri razlagi drugega dela tega pogoja pride do vprašanja obsega tega določila, saj je pomembno določiti kaj iz vidika te direktive pomenita avtomatizirana obdelava in oblikovanje profilov, ter kakšna je njivna korelacija iz vidika tega določila. Problematičen je del določila »vključno z«, saj poraja vprašanje, ali dodatek oblikovanja profilov obseg določila oži s tem da konkretizira, da se določilo nanaša zgolj na avtomatizirano obdelavo, ki vključuje tudi oblikovanje profilov, ali obseg določala širi oziroma konkretizira, s tem da dodaja nov proces za sprejemanje odločitev oziroma z namenom jasnosti konkretizira oblikovanje profilov, kot za to določilo relevantno obliko avtonomne obdelave podatkov. Razlika je v tem, ali se določilo nanaša samo na odločitve sprejete na podlagi avtomatizirane obdelave, pri kateri je bilo prisotno oblikovanje profilov, ali se nanaša na odločitve sprejete na podlagi vsake avtomatizirane obdelave (osebnih) podatkov. Med tem, ko v slovenskem prevodu GDPR besedna interpretacija vodi do druge možnosti, je v primeru nekaterih drugih jezikov manj jasna (ang. Includes, v tem primeru je možna dvojna interpretacija).

Pojma obdelava in oblikovanje profilov sta opredeljena v 4. členu.

- „obdelava“ pomeni vsako dejanje ali niz dejanj, ki se izvaja v zvezi z osebnimi podatki ali nizi osebnih podatkov z avtomatiziranimi sredstvi ali brez njih, kot je zbiranje, beleženje, urejanje, strukturiranje, shranjevanje, prilagajanje ali spreminjanje, priklic, vpogled, uporaba, razkritje s posredovanjem, razširjanje ali drugačno omogočanje dostopa, prilagajanje ali kombiniranje, omejevanje, izbris ali uničenje⁶⁷
- „oblikovanje profilov“ pomeni vsako obliko avtomatizirane obdelave osebnih podatkov, ki vključuje uporabo osebnih podatkov za ocenjevanje nekaterih osebnih vidikov v zvezi s posameznikom, zlasti za analizo ali predvidevanje uspešnosti pri delu, ekonomskega položaja, zdravja,

⁶⁵ Kratica izvira iz angleškega pojma *Data Protection Impact Assessment*.

⁶⁶ Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679, str. 20-21.

⁶⁷ 4(2). člen GDPR.

osebnega okusa, interesov, zanesljivosti, vedenja, lokacije ali gibanja tega posameznika⁶⁸

Kot vidimo v opredelitvi oblikovanja profilov (»pomeni vsako obliko avtomatizirane obdelave osebnih podatkov«) je oblikovanje profilov vrsta avtomatske obdelave, kar odraža prej pojasnjeno konkretizacijo obsega določila. Vendar kot pojasnjujeta Mendoza in Bygrave, recital 71 vodi do interpretacije, da je oblikovanje profilov element avtomatizirane obdelave, v kontekstu tega določila. V omenjenem recitalu namreč piše: » Taka obdelava vkjučuje „oblikovanje profilov“ v kakršni koli obliki avtomatizirane obdelave osebnih podatkov, na podlagi katerih se ocenjujejo osebni vidiki v zvezi s posameznikom,...«⁶⁹. Vlogo vključenosti pojma »oblikovanje profilov« pojasnjujeta kot kriterij za to določbo relevantne avtomatizirane obdelave podatkov. Na podlagi tega se zavzemata za ožji obseg te določbe, namreč v odnosu zgolj do avtomatizirane obdelave, katere del je obvezno tudi oblikovanje profilov. To pojasnjujeta tudi s tem, da se pripravna dela GDPR fokusirajo pretežno na nevarnosti oblikovanja profilov in da je tovrstna interpretacija bolj smiselna iz vidika ozadja 22. člena. S tem bi domet tega člena prav tako bil širši kakor domet njegovega prednika(DPD člen 15).⁷⁰

Vendar, kot je pojasnjeno v smernicah delovne skupine WP29, se četrti odstavek 4. člena GDPR, ki določa oblikovanje profilov nanaša na vsako obliko avtomatizirane obdelave in ne na obdelavo, ki je zgolj oziroma popolnoma avtomatizirana. To pomeni, da kljub temu, da je uporaba avtomatizirane obdelave podatkov pogoj za oblikovanje profilov v smislu GDPR, človeška vključenost v ta postopek ne pomeni, da ne gre za oblikovanje profilov. ⁷¹Oblikovanje profilov pogosto vključuje prediktivne elemente, zaradi česar ustvarjeni profili pogosto niso natančni, k čemur prav tako prispeva nepravilen zajem obdelovanih podatkov, obdelava irelevantnih podatkov ali pa so podatki obdelani izven konteksta. Upoštevati moremo tudi možne napake v samem algoritmu glede prepoznavanja korelacij. ⁷²

⁶⁸ 4(4). člen GDPR.

⁶⁹ Recital 71 GDPR.

⁷⁰ Mendoza, I., Bygrave, L. A., The Right Not To Be Subject to Automated Decisions Based on Profiling, v: University of Oslo Faculty of Law Legal Studies Research Paper Series, (2017) 2017-20, str. 13-14.

⁷¹ Article. 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679, str. 6.

⁷² Ibidem, str. 24.

Kot je v smernicah v nadaljevanju pojasnjeno se odločitve na podlagi avtomatizirane obdelave podatkov sprejemajo na podlagi podatkov, ki so jih neposredno zagotovili obravnavani posamezniki, podatkov zajetih z opazovanjem posameznikov in izvedenih podatkov (me te sodijo tudi profili, ki so bili že ustvarjeni).⁷³Ta opredelitev je ključnega pomena iz vidika določila 22(1) člena, saj pozornost iz oblikovanj profila kot postopka, osredotoči na rezultat tega postopka. Kot je v smernicah nadalje pojasnjeno, odločanje na podlagi avtomatizirane obdelave podatkov lahko poteka z ali brez oblikovanja profilov in da lahko gre za ločena postopka ali ne. Prav tako opozarja, da se postopek oblikovanja profilov, oziroma njegov rezultat uporablja tudi za odločanje, ki ni avtomatizirano, temveč ga vrši človek.⁷⁴

Za razumevanje obsega tega določila je prvotno treba poudariti, da avtomatizirani sistemi obdelave podatkov ne sprejemajo odločitve. Odločitev sprejema lahko zgolj človek. Avtomatiziran sistem obdelave podatkov lahko človeku nudi zgolj odgovor (predlog ali navodilo), kako naj odloči. To smiselno izhaja tudi iz same formulacije določila 22(1) člena, ki govori o odločitvah, ki temeljijo zgolj na avtomatizirani obdelavi podatkov in ne na odločitvah, ki jih ti sistemi sprejmejo, saj le ti tega iz pravnega vidika niso sposobni. Smiselno s predhodnim pojasnilom pogoja, da odločitev temelji zgolj na avtomatizirani obdelavi podatkov, lahko »odgovor«, v smislu 22. člena določimo kot informacijo, ki popolnoma določa človekovo odločitev, brez njegove smiselne interakcije. To pomeni, da je človekova odločitev identična odgovoru in temelji zgolj na njem (v tem primeru gre za navodilo, kako odločiti). Takšen odgovor lahko delimo na neposreden in posreden odgovor. Za tovrstno delitev moremo postopek odločanje obravnavati kot iskanje odgovora na določeno vprašanje. Preprost primer bi bilo v postopku odločanja o odobritvi kredita. V tem primeru bi za odločanje bilo zastavljeno vprašanje, ali kreditojemalcu odobriti kredit. Možna odgovora, ki prav tako označujeta odločitev sta DA ali NE. Program, ki ga odločevalec za namen odločitve uporabi, lahko ustvari posreden ali neposreden odgovor na zastavljeno vprašanje. Neposredni odgovorov bi bil DA ali NE med tem ko bi posreden odgovor bil drugačen na primer, da je kreditojemalec kreditno spodoben. Med tem ko takšen odgovor ni identičen kateremu od možnih odgovorov za odločitev ima močnejšo korelacijo z enim iz med njiju (odgovor kreditno sposoben usmerja odločevalca proti odgovoru DA). Tukaj je potrebno

⁷³ Ibidem, str. 7-8.

⁷⁴ Ibidem.

opozoriti, da ne glede na to, ali gre za posredni ali neposredni odgovor, ni nujno, da sodi v domet 22. člena. Odločevalec lahko tovrstni avtomatiziran postopek uporabi zgolj kot nasvet (odloča sam vendar odločitev zgolj preveri s tovrstnim sistemom). Takšna uporaba je praktično problematična, saj bi v tovrstnem primeru resnično vlogo takšnega sistema lahko določali zgolj na podlagi primerjave večje količine odločitev odločevalca z odgovori sistema in njegove svobode arbitrarnega odločanja. Posredni odgovor prav tako lahko služi zgolj kot izpolnitev enega izmed več kriterijev na podlagi katerih odločevalec odloča, lahko pa neposredno določa odločitev, če izpolnjuje edini kriterij odločanja ali nosi takšno težo, da drugi kriteriji odločitve ne morejo spremeniti in če odločevalec ne more odločiti kontradiktorno temu odgovoru. (Primer: odgovor sistema je , da je kreditjemalec kreditno sposoben, odločevalec pa glede na podana navodila kredite odobrava zgolj pod pogojem kreditne sposobnosti).

Oblikovanje profilov moramo za lažje razumevanje tega določila obravnavati ločeno kot postopek obdelave podatkov in kot izveden podatek oziroma profil, ki ga ta postopek ustvari. Pomen tovrstne ločitve je v tem, da vloga oblikovanja profilov ni samo odločanje, temveč ustvarjanje izvedenih podatkov (kot izhaja iz 4(4) člena oceno nekaterih osebnih vidikov), na podlagi katerih (vključno z ali brez drugih podatkov) z avtomatizirano obdelavo sprejemajo odločitve. Postopek oblikovanja profila je lahko združen z avtomatizirano obdelavo podatkov, na kateri temelji odločitev ali pa ločen postopek. Kot ločen postopek je lahko namenjen neposrednemu odločanju ali zgolj za namen oblikovanja profila.

Oblikovan profili glede na uporabo lahko služi zgolj kot izveden podatek (oseba A je dober voznik, vendar to ni edina kvalifikacija na podlagi katere se presoja premija zavarovanja) ali pa kot predhodno opisan posredni odgovor (oseba A je dober voznik, ta lastnost je odločilna). V tej delitvi najdemo smisel eksplicitne omembe oblikovanja profilov v prvem odstavku 22. člena GDPR, saj je oblikovanje profilov, ki je samostojno določen pojem v GDPR lahko vrsta avtomatizirane obdelave podatkov, zgolj na kateri temelji odločitev.

Kot sem omenil prej, lahko v procesu oblikovanja profila sodeluje človek. Na podlagi tega bi bila možna preprosta razlaga, da omejitve, določena v prvem odstavku 22. člena GDPR velja za postopke oblikovanja profila, če oblikovan profil služi kot predhodno opisan posredni odgovor in če v postopku oblikovanja ni sodeloval človek. Vendar menim, da je tovrstna interpretacija neustrezna, saj sama

po sebi ne zahteva učinkovite človeške udeležbe, kot sem jo pojasnil predhodno pri razlagi, kdaj odgovor temelji zgolj na podlagi avtomatizirane obdelave podatkov. Da tovrsten profil ne sodi pod prepoved, ki jo določa prvi odstavek 22. člena GDPR, more človek v procesu oblikovanja profila sodelovati na način, ki je smiseln in izveden s strani nekoga, ki je kompetenten in ima avtoriteto vplivati na dejanski profil.

Na podlagi pojasnjene je razumna interpretacija tega določila v okviru širšega obsega, namreč, da prepoved iz 22(1) člena zajema vse odločitve, ki temeljijo zgolj na podlagi avtomatizirane obdelave, tako tiste, ki vključujejo oblikovanje profilov kot del oziroma vrsto obdelave podatkov, kot tudi tiste, ki tega ne vključujejo.

Širša razlaga, ki štiti posameznika tudi v odsotnosti oblikovanja profila, je prav tako smiselna zato, ker ima nenazadnje tudi zgolj avtomatizirana obdelava podatkov lahko sporne, celo diskriminativne učinke zaradi omejenega obsega presojanih lastnosti ter potencialno sporne postavitve kriterijev in meril odločanja. Nenazadnje, kot bom pojasnil v nadaljevanju, to določilo zadeva odločitve, ki imajo pravne ali podobne učinke. V tem smislu gre za odločitve, ki so enako pomembne iz vidika zaščite posameznika. Določitev enake mere pozornosti pri preprostem avtomatiziranemu odločanju, tudi če temelji samo na objektivnih parametrih, je pomembna zaradi možnosti napak pri različnih fazah, pomembnih za tovrstno obdelavo (napake pri zajemu podatkov, sortiranju teh podatkov, itd) Preprost primer je določitev kazni za prehitro vožnjo na podlagi podatkov o izmerjeni hitrosti in identifikaciji vozila z registersko tablico, ki jih je zabeležila kamera ob cestišču in obdelal avtomatiziran sistem. Pri obeh podatkih lahko pride do napak (napačen avtomatiziran prepis tablice, ki temelji na računalniškem vidu in izmerjeni hitrosti, ki je zaradi tehničnih problemov lahko absurdno nepravilna, kar bi človek ob pregledu opazil, sistem avtomatizirane obdelave pa morda ne bi (npr. če sistem določa kazen ob prekoračitvi hitrosti 50km/h in nima *priučene* razumne zgornje meje hitrosti, ki jo avti lahko dosežejo, bo izdal kazen v primeru, kjer bo izmerjena hitrost 6000000 km/h, kljub temu, da avto takšne hitrosti ne more doseči in lahko opravičljivo sklepamo, da je zabeležena hitrost posledica tehnične napake merilne naprave ali drugih okoljskih vplivov, ki napravo motijo). Človek, ki bi takšno meritev pregledal, bi takoj videl, da je z meritvijo nekaj narobe. Seveda lahko v sistem opisan v našem primeru vključimo razumno zgornjo mejo izmerjene hitrosti, vendar je vprašljivo ali tovrsten sistem lahko optimiziramo, da je odporen na, oziroma da zazna, prisotnost vseh vplivov, ki lahko vodijo do napačne meritve. To je še veliko

težje v primeru kompleksnih sistemov avtomatizirane obdelave podatkov, ki presojujejo na podlagi več lastnosti in podatke črpajo iz več virov.

5.7.1 Odločitev ima pravne učinke v zvezi z njim ali na podoben način nanj znatno vpliva

Prepoved določena v 22. členu posameznika štiti, če ima odločitev pravne učinke v zvezi z njim ali na podoben način nanj znatno vpliva. Ta pogoj je izpolnjen, če odločitev delno ali popolnoma določi ali spremeni posameznikove pravice ali dolžnosti oziroma ima posledice, ki so več kot malenkostne za dobrobit posameznika.⁷⁵

Kot pojasnjuje delovna skupina WP29, gre v smislu 22. člena za pravni učinek, če odločitev vpliva na posameznikove pravice (pravica do združevanja), kot tudi če odločitev vpliva na posameznikov pravni status (npr. pridobitev državljanstva) ali na njegove pravice v pogodbenem razmerju (prenehanje pogodbe). Tudi če učinki odločitve ne vplivajo na posameznikove pravice ali obligacije, sodijo pod to določilo, če ga prizadenejo primerljivo s tistimi, ki imajo pravni učinek.⁷⁶

Delovna skupina WP29 posebej izpostavlja učinke, ki znatno vplivajo na posameznikove okoliščine, vedenje in odločitve, ki imajo na njega dolgotrajen učinek in ki v skrajnih primerih vodijo do diskriminacije. Kot primere znatnih učinkov navajajo tiste, ki vplivajo na posameznikovo finančno stanje (dostop do kredita), njegov dostop do zdravstvenih storitev, zaposlitve in izobraževanja. Opozarja, da ima tovrstni učinek lahko tudi targetirano oglaševanje, ki temelji na podlagi oblikovanja profilov, glede na njeno intruzivnost (če vključuje sledenje uporabnikom širom različnih strani, naprav in storitev), glede na pričakovanja in želje posameznika na katerega se navezuje, glede na način oglaševanja in kadar uporablja šibkosti targetiranega posameznika. Poudarjajo primere targetiranega oglaševanja visoko obrestnih kreditov posameznikom s šibkim finančnim stanjem in cenovnega spreminjanja ponudb za storitve in izdelke, če tovrstno prilagajanje onemogoči dostop zaščitenih skupin do teh storitev in izdelkov. Opozarjajo, da dejanja in stanja drugih prav tako lahko močno vplivajo na učinke zoper obravnavanega posameznika

⁷⁵ Mendoza, I., Bygrave, L. A., The Right Not To Be Subject to Automated Decisions Based on Profiling, v: University of Oslo Faculty of Law Legal Studies Research Paper Series, (2017) 2017-20, str. 12, 20.

⁷⁶ Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, str. 21.

(pogoji posojila se presojajo glede na kreditno sposobnost drugih prebivalcev območja, kjer posameznik živi.⁷⁷

Kot opozarjata Bygraves in Mendoza ne smemo spregledati recital 38 GDPR, ki odraža, da kadar avtomatizirano sprejemanje odločitev zadeva otroka, je tovrstna odločitev znatna v smislu 22. člena.⁷⁸ Recital pojasnjuje potrebo posebnega varstva otrok, saj se ti morda ne zavedajo zadevnih tveganj, posledic in zaščitnih ukrepov, ter pojasnjuje, da privolitev »nosilca starševske odgovornosti ne bi smela biti potrebna v okviru storitev preventive ali svetovanja, ki se nudijo neposredno otroku.«⁷⁹

5.7.2 Izjeme prepovedi določila 22. člena GDPR

Drugi odstavek 22. člena določa tri izjeme prepovedi iz prvega odstavka, namreč če je odločitev:

1. Nujna za sklenitev ali izvajanje pogodbe med posameznikom, na katerega se nanašajo osebni podatki, in upravljavcem podatkov⁸⁰

Uporaba postopkov za avtomatizirano odločanje za namen sklepanja pogodb je pogosto zaželenja saj je pogosto bolj praktična in včasih edina izvedljiva možnost (sklepanje velikih količin pogodb). Za uporabo teh procesov iz vidika te izjeme, mora upravljalec podatkov biti sposoben utemeljiti, da so tovrstni procesi za ta namen nujni, kar pa niso, če obstajajo drugi učinkoviti in manj intruzivni načini za doseg tega namena. Upoštevati je potrebno tudi pomen in uporabo avtomatiziranih sistemov za odločanje v predhodni fazi sklenitve pogodbe (na primer filtriranje kandidatov za delovno mesto, ki ne ustrezajo objektivnim pogojem, v primeru velike količine kandidatov).⁸¹

⁷⁷ Ibidem, str. 21-22.

⁷⁸ Mendoza, I., Bygrave, L. A., The Right Not To Be Subject to Automated Decisions Based on Profiling, v: University of Oslo Faculty of Law Legal Studies Research Paper Series, (2017) 2017-20, str. 12.

⁷⁹ Recital 38 GDPR.

⁸⁰ 22(2a). člen GDPR.

⁸¹ Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679, str. 23.

2. Dovoljena v pravu Unije ali pravu države članice, ki velja za upravljavca in določa tudi ustrezne ukrepe za zaščito pravic in svoboščin ter zakonitih interesov posameznika, na katerega se nanašajo osebni podatki,⁸²

V recitalu 71 GDPR poudarjeno, da bi sprejemanje odločitev, na podlagi avtomatizirane obdelave podatkov moralo biti dovoljeno »kadar ga izrecno dovoljuje pravo Unije ali pravo države članice, ki velja za upravljavca, tudi za namene spremljanja in preprečevanja zlorab in davčnih utaj v skladu s predpisi, standardi in priporočili institucij Unije ali nacionalnih nadzornih teles ter zagotavljanje varnosti in zanesljivosti storitve, ki jo zagotavlja upravljavec...«⁸³ Na podlagi tega lahko sklepamo, da je ta izjema bila sprejeta pretežno za namen uporabe tovrstnih procesov za namen izvajanja funkcij državnih organov in institucij.

3. Utemeljena z izrecno privolitvijo posameznika, na katerega se nanašajo osebni podatki⁸⁴

Ker avtomatizirana obdelava podatkov za namen odločanja predstavlja veliko tveganje za posameznika, na katerega se ti podatki nanašajo, je primerna višja raven nadzora tega posameznika nad uporabo njegovih osebnih podatkov.⁸⁵ Privolitev posameznika, na katerega se nanašajo osebni podatki pomeni »vsako prostovoljno, izrecno, informirano in nedvoumno izjavo volje posameznika, na katerega se nanašajo osebni podatki, s katero z izjavo ali jasnim pritrdilnim dejanjem izrazi soglasje z obdelavo osebnih podatkov, ki se nanašajo nanj«⁸⁶. Kot izhaja iz definicije je privolitev ustrezna če ustreza štirim pogojem:

4. Je dana prostovoljno

Ta pogoj je izpolnjen kadar je posameznik resnično imel svobodo in nadzor izraziti svojo voljo. Kadar posameznik ni imel dejanske izbire, je bil k soglasju privolitvi primoran ali bi bil v primeru ne privolitve deležen negativnih posledic, privolitev za izpolnitev pogojev GDPR ni ustrezna. Privolitev prav tako ni ustrezna, če je del privolitev pogojev uporabe, o katerih se posameznik ne more pogajati. Načelno

⁸² 22(2b). člen GDPR.

⁸³ recital 71 GDPR.

⁸⁴ 22(2c). člen GDPR.

⁸⁵ Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679, str. 24.

⁸⁶ 4(11). člen GDPR.

velja, da privolitev ni prostovoljna, če je posameznik ne more umakniti ali zavrniti brez škode.⁸⁷

5. Je izrecna

Privolitev posameznika mora biti izrecna za eno ali več določenih namenov. Namen te zahteve je zagotoviti posamezniku ustrezno mero nadzora nad obdelavo njegovih podatkov.⁸⁸ Ta zahteva je poudarjena tudi v prvem odstavku 6. člena GDPR, kjer je kot pogoj, da je obdelava zakonita določeno, da »posameznik, na katerega se nanašajo osebni podatki, je privolil v obdelavo njegovih osebnih podatkov v enega ali več določenih namenov«⁸⁹.

6. Je dana informirano

GDPR zahteva, da je privolitev informirana. Informiranost posameznikov je za njihovo privolitev pomembna, saj zgolj tako lahko sprejmejo informirano odločitev, razumejo, k čemu soglašajo in zgolj tako lahko učinkovito uveljavljajo svojo pravico do zadržanja soglasja. Da je izpolnjen ta pogoj mora posameznik biti seznanjen vsaj z identiteto upravljalca podatkov, namena vsakega postopka obdelave, za katerega upravljaec skuša pridobiti privolitev, kakšne vrste podatkov bodo zbrane in uporabljene, z informacijami o uporabi podatkov za avtomatizirano odločanje v skladu s 22(2)(c) členom GDPR (ki določa izjemo v primeru privolitve, ki jo trenutno obravnavamo) in v primeru prenosov podatkov, za katere se uporabljajo ustrezni zaščitni ukrepi, s temi, kot so določeni v 46. členu GDPR.⁹⁰ V zadevi *Planet49*⁹¹ je Sodišče potrdilo, da »mora jasna in izčrpna obvestitev uporabniku omogočiti, da z lahkoto ugotovi posledice morebitne dane privolitve, in zagotoviti, da je ta privolitev dana ob popolni seznanjenosti z dejstvi. Ta obvestitev mora biti jasno razumljiva in dovolj podrobna, da uporabnik lahko razume delovanje uporabljenih piškotkov«. V tej zadevi je sodišče prav tako pojasnilo, da je namestitev piškotkov vrsta obdelave osebnih podatkov, skladno s določbami GDPR, zato takšna pojasnilna dolžnost smiselno velja splošno za obdelavo osebnih podatkov.

⁸⁷ European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, version 1.1, 2020, str. 7.

⁸⁸ Ibidem, str. 13-14.

⁸⁹ 95/46/ES, 6(1a). člen GDPR.

⁹⁰ European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, version 1.1, 2020, str. 15-16.

⁹¹ Zadeva C-673/17, *Planet 49*, ECLI:EU:C:2019:801.

Prav tako je odločilo, da morajo informacije o namestitvi piškotkov zajemati trajanje delovanja piškotkov in z možnost oziroma nezmožnost tretjih oseb, da dostopajo do teh piškotkov.

7. Dana z nedvoumna izjavo volje

Za privolitev je potrebna izjava ali jasno pritrdilno dejanje, ki mora biti s strani posameznika dano na aktiven način. Privolitev k določeni obdelavi more biti očitna. Jasno pritrdilno dejanje torej pomeni aktivno dejanje, s katerim posameznik izraža privolitev za določeno obdelavo.⁹² V recitalu 32 GDPR je nadalje določeno, »da je posameznik, na katerega se nanašajo osebni podatki, prostovoljno, specifično, ozaveščeno in nedvoumno izrazil soglasje k obdelavi osebnih podatkov v zvezi z njim, kot je s pisno, tudi z elektronskimi sredstvi, ali ustno izjavo. To lahko vključuje označitev okenca ob obisku spletne strani, izbiro tehničnih nastavitvev za storitve informacijske družbe ali katero koli drugo izjavo ali ravnanje, ki v tem okviru jasno kaže na to, da posameznik, na katerega se nanašajo osebni podatki, sprejema predlagano obdelavo svojih osebnih podatkov. Molk, vnaprej označena okenca ali nedejavnost zato ne pomenijo privolitve.«⁹³ Glede vnaprej označenih okenc je Sodišče odločilo v predhodno omenjeni zadevi *Planet49*⁹⁴, namreč, da ne gre za veljavno privolitev, kadar je potrditveno polje predhodno označil ponudnik storitve in ki bi ga moral uporabnik, da zavrne svojo privolitev odznačiti. Dejstvo, da je posameznik želel dostopati do določene storitve (v primeru te zadeve sodelovanje v nagradni igri) ne more zadostovati kot veljavna privolitev, v tem primeru za namestitvev piškotkov.

Pri spletnem dostopanju do storitev vseh vrst praktično stalno dajemo privolitev v obliki sklepanja pogodb s enim klikom (označitev okenca, pritisk gumba s napisano: Se strinjam). Problem sklepanja spletnih pogodb z enim klikom je, da ljudi primora k avtomatskemu ravnanju, brez razmisleka. Ljudje se pri sklepanju spletnih pogodb le redko vprašajo glede vsebine pogodb in strank, s katerimi te pogodbe sklepajo. Problem je v tem da svoboda do sklenitve pogodbe zahteva tudi obratno, svobodo ne skleniti pogodbo in ko sklepanje postane avtomatično izgineta obe, saj več ne govorimo o svobodi in avtonomiji.⁹⁵

⁹² European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, version 1.1, 2020, str. 18.

⁹³ Recital 32 GDPR.

⁹⁴ C-673/17, *Planet 49*, ECLI:EU:C:2019:801.

⁹⁵ Povzeto po: <<http://cyberlaw.stanford.edu/blog/2019/02/electronic-contracts-and-illusion-consent>> (13.4.2020).

5.7.2 Omejitev izjem

V tretjem odstavku 22. člena je določeno, da v primerih odločitev, ki so nujne za sklenitev ali izpolnitev pogodbe in odločitev, ki so utemeljene z izrecno privolitvijo, nosi upravljalec podatkov dolžnost da »izvede ustrezne ukrepe za zaščito pravic in svoboščin ter zakonitih interesov posameznika, na katerega se nanašajo osebni podatki, vsaj pravice do osebnega posredovanja upravljavca, do izražanja lastnega stališča in izpodbijanja odločitve.«⁹⁶ Pravice navedene v tem odstavku ne predstavljajo vseh možnih zaščitnih ukrepov temveč zgolj minimalne zahteve.⁹⁷ Posebej pomembna je pravica do osebnega posredovanja upravljavca, saj kot je pojasnila delovna skupina WP29, mora pregled odločitve speljati oseba, ki ima avtoriteto in sposobnost odločitev spremeniti. Pri pregledu mora upoštevati vse relevantne podatke, vključno z vsemi podatki, ki mu jih dodatno zagotovi posameznik, na katerega se osebni podatki nanašajo. Upravljalec podatkov je dolžan obravnavanim posameznikom zagotoviti uveljavljanje iz 3. odstavka 22. člena GDPR.⁹⁸

Omenjene izjeme iz 2. odstavka 22. člena so na podlagi 4. odstavka tega člena prav tako omejene, kadar odločitve temeljijo na podlagi posebnih vrst osebnih podatkov, kot so določeni v prvem odstavku 9. člena GDPR. Osebni podatki posebnih vrst so podatki, »ki razkrivajo rasno ali etnično poreklo, politično mnenje, versko ali filozofsko prepričanje ali članstvo v sindikatu, in obdelava genetskih podatkov, biometričnih podatkov za namene edinstvene identifikacije posameznika, podatkov v zvezi z zdravjem ali podatkov v zvezi s posameznikovim spolnim življenjem ali spolno usmerjenostjo.«⁹⁹ Odločitev, ki temelji zgolj na avtomatizirani obdelavi podatkov, vključno z osebnimi podatki posebnih vrst po tem takem velja, kadar gre za primer, ki ga določa ena iz med prej opisanih izjem iz 2. odstavka 22. člena GDPR in če so izpolnjeni pogoji, ki jih določata točki a in g, drugega odstavka 9. člena.¹⁰⁰ Ti določata, da je uporaba posebnih vrst osebnih podatkov dovoljena, kadar:

⁹⁶ 22(3). Člen GDPR.

⁹⁷ Mendoza, I., Bygrave, L. A., The Right Not To Be Subject to Automated Decisions Based on Profiling, v: University of Oslo Faculty of Law Legal Studies Research Paper Series, (2017) 2017-20, str. 15.

⁹⁸ Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679, str. 27.

⁹⁹ 9(1). Člen GDPR.

¹⁰⁰ Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679, str. 24.

- posameznik, na katerega se nanašajo osebni podatki, je dal izrecno privolitev za obdelavo navedenih osebnih podatkov za enega ali več določenih namenov, razen kadar pravo Unije ali pravo države članice določa, da posameznik, na katerega se nanašajo osebni podatki, ne sme odstopiti od prepovedi iz odstavka 1 (člena 9 GDPR)¹⁰¹
- obdelava je potrebna iz razlogov bistvenega javnega interesa na podlagi prava Unije ali prava države članice, ki je sorazmerno z zastavljenim ciljem, spoštuje bistvo pravice do varstva podatkov ter zagotavlja ustrezne in posebne ukrepe za zaščito temeljnih pravic in interesov posameznika, na katerega se nanašajo osebni podatki.¹⁰²

Poleg teh pogojev, za veljavnost odločitev, ki temeljijo na osebnih podatkih posebnih vrst, se zahteva, da se izvajajo ustrezni ukrepi za zaščito pravic in svoboščin ter zakonitih interesov posameznika, na katerega se nanašajo osebni podatki, kakor je določeno sicer že v 3. odstavku 22. člena GDPR. Razlika je seveda v tem, da se 4. odstavek nanaša na vse tri izjeme iz 2. odstavka, se pravi tudi kadar je uporaba avtomatizirane obdelave osebnih podatkov za odločanje dovoljena s strani EU ali države članice. Pomembno je tudi poudariti, da za razliko od 3. odstavka, 4. odstavek ne določa, da more upravljalec podatkov zagotoviti omenjene zaščitne ukrepe, temveč zgolj, da se tovrstni ukrepi izvajajo. Tovrstna formulacija določbe smiselno nakazuje na dolžnost EU ali držav članic, da zagotovi izvajanje tovrstnih zaščitnih ukrepov, kadar je uporaba avtomatiziranega odločanja dovoljena s strani EU ali državne članice (izjema) in kadar se osebni podatki posebnih vrst obdelujejo iz razlogov bistvenega javnega interesa na podlagi prava Unije ali prava države članice (kakor je določeno v 9 (2g)).

5.8 Ocena učinka v zvezi z varstvom podatkov

DPIA je postopek namenjen opisu obdelave, oceni njene nujnosti in proporcionalnosti in kot pomoč pri odpravi tveganj, ki zadevajo pravice in svoboščine posameznikov, ki izvirajo iz obdelave osebnih podatkov, s tem da tovrstna tveganja oceni in določi ukrepe za njihovo odpravo. Povedano drugače, DPIA je orodje za doseganje in dokazovanje skladnosti z določili GDPR.¹⁰³

¹⁰¹ 9(2a). člen GDPR.

¹⁰² 9(2g). člen GDPR.

¹⁰³ Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679, 2017, str. 4.

V 35. členu GDPR je določena DPIA, ki jo upravljalec mora opraviti kadar je možno, »da bi lahko vrsta obdelave, zlasti z uporabo novih tehnologij, ob upoštevanju narave, obsega, okoliščin in namenov obdelave povzročila veliko tveganje za pravice in svoboščine posameznikov.«¹⁰⁴V točki a 3. odstavka 35. člena je nadalje določeno, da se DPIA zahteva v primeru »sistematičnega in obsežnega vrednotenja osebnih vidikov v zvezi s posamezniki, ki temelji na avtomatizirani obdelavi, vključno z oblikovanjem profilov, in je osnova za odločitve, ki imajo pravne učinke v zvezi s posameznikom ali nanj na podoben način znatno vplivajo«.¹⁰⁵ Pozorni moramo biti na formulacijo določila saj se navezuje na vrednotenje osebnih vidikov, ki temelji na avtomatizirani obdelavi podatkov in ne zgolj na avtomatizirani obdelavi podatkov. Zaradi tega se to določilo nanaša na odločanje, vključno s oblikovanjem profilom, s pravnim ali podobnim učinkom, ki je delno ali popolnoma avtomatizirano.¹⁰⁶

5.9 Prepoznavanje diskriminacije in predhodni preizkus sistemov umetne inteligence

Algoritemsko odločanje predstavlja možnost transparentnost, saj teoretično nudi možnost rekonstrukcije postopka odločanja, ki pri ljudeh ni mogoča. Ta transparentnost je vendarle še vedno omejena s sposobnostjo razumevanja delovanja teh sistemov. Možen način interpretacije algoritmičnih odločitev je na primer analiza upoštevanih podatkov.¹⁰⁷

Velik problem boja zoper algoritemsko diskriminacijo je zaznavanje tovrstne diskriminacije. Zaradi zapletenosti sistemov, ki temeljijo na strojnem učenju, je pogosto težko razumeti, kako ta deluje in sprejema odločitve. Upravljalec podatkov mora posamezniku, na katerega se osebni podatki navezujejo postopek sprejemanja odločitev pojasniti na preprost način, ki ne rabi vsebovati temeljite razlage ali razkritja uporabljenega algoritma. Tovrstno pojasnilo mora biti za posameznika, katerega osebni podatki se obdelujejo smiselno.¹⁰⁸ Kadar sistem vključuje tako

¹⁰⁴ 35(1). Člen GDPR.

¹⁰⁵ 35(3a). člen GDPR.

¹⁰⁶ Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679, str. 29.

¹⁰⁷ Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: University of Chicago Law Review, 86 (2019) 459, str. 474-475.

¹⁰⁸ Article. 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679, str. 14.

oblikovanje profilov, kakor tudi avtomatizirano odločanje na podlagi teh profilov, pojasnilna dolžnost velja glede obeh procesov.¹⁰⁹

Na podlagi načela transparentnosti, kot ga določa GDPR, morajo upravljavci posameznikom, na katere se osebni podatki nanašajo, na razumljiv in preprost način razložiti, kako poteka postopek oblikovanja profilov in avtomatiziranega odločanja. Kadar se oblikovanje profila uporablja za odločanje (tako v obsegu 22. člena ali ne) mora biti posameznik, na katerega se osebni podatki v obdelavi nanašajo obveščen tako o postopku oblikovanja profila, kot tudi o namenu uporabe tega profila za odločanje.¹¹⁰ V recitalu 60 GDPR je eksplicitno določeno: »Načeli poštene in pregledne obdelave zahtevata, da je treba posameznika, na katerega se nanašajo osebni podatki, obvestiti o obstoju dejanja obdelave in njegovih namenih. Upravljavec bi moral posamezniku, na katerega se nanašajo osebni podatki, zagotoviti vse dodatne informacije, potrebne za zagotavljanje poštene in pregledne obdelave ob upoštevanju specifičnih okoliščin in okvira obdelave osebnih podatkov. Poleg tega bi moral biti ta posameznik obveščen o oblikovanju profilov in njegovih posledicah.«¹¹¹

Nedavna odkritja na področju računalništva in statistike ponujajo različne vidike glede tega, kdaj algoritem odloča pravično. Vendar se na žalost večina teh odkritij temelji na statistični analizi rezultatov algoritma in ne samih delov procesa algoritemskega odločanja.¹¹²

Prednost avtomatiziranega odločanja je, da tovrsten sistem lahko preizkusimo pred uporabo. S predhodnim preizkusom lahko ugotovimo, ali je postopek pravičen in s tem nediskriminatoren. V tovrstnem preizkusu bi se sistem uporabil za odločanje o hipotetični (testni) populaciji, s čimer bi se odkrilo razlikovanje odločanja glede na zaščitene skupine. Problem je, da je tovrsten preizkus praktično težaven, saj je preprosta analiza razlikovanja rezultatov glede na pripadnost posameznikov določenim skupinam *prima facie* primer različnega oziroma pristranskega odločanja. Oceniti, kdaj je rezultat resnično diskriminatoren je težavno, saj ne poznamo

¹⁰⁹ Ibidem, str. 23.

¹¹⁰ Ibidem, str. 16.

¹¹¹ Recital 60 GDPR.

¹¹² Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: University of Chicago Law Review, 86 (2019) 459, str. 465

kriterijev za primerjavo dveh posameznikov, pri določanju, ali sta bila obravnavana različno. S tem se tovrstna analiza rezultatov osredotoča na preprosto primerjavo.¹¹³

Smiselno bi bilo, da bi se za uporabo v javnem sektorju smeli uporabljati zgolj sistemi, ki bi bili predhodno preizkušeni za tveganje in bi bili sposobni pregleda in revizije njihovih odločitev oziroma rezultatov. Takšen pogoj bi morda bil koristen tudi za uporabo tovrstnih sistemov v zasebnem sektorju, kadar ima uporaba tovrstnih sistemov znatne učinke.¹¹⁴ Menim, da bi za presojo, kdaj bi tovrsten pogoj bil ustrezen za zasebni sektor najbolj učinkovito služil kriterij relevantnega učinka, kot je določen v prvem odstavku 22. člena, namreč kadar ima pravne učinke v zvezi posameznikom, čigar podatki se obdelujejo ali na podoben način nanj znatno vpliva. Seveda se moramo zavedati, da je tovrsten pogoj uporabe tehnično izredno zahteven in pogosto celo neizvedljiv. Pri obravnavi regulacije avtomatiziranega odločanja seveda strmimo k čim večji varnosti posameznikov, ki jih te odločitve prizadenejo in preglednosti postopka odločanja, vendar vselej moramo pri določanju tovrstnih zahtev postopati razumno. Pri presoji, kdaj določiti takšne ukrepe, moramo vselej upoštevati tehnično izvedljivost samih ukrepov in ali določen sistem, v kontekstu uporabe oziroma glede na to, kdo ga uporablja, predstavlja tako veliko tveganje, daje tovrstna zahteva utemeljena. Pri tem menim, da je tovrstna zahteva vsekakor bolj potrebna in s tem primerna za sisteme, ki so v uporabi v javnem sektorju. Tako drastične zahteve bi bilo razumljivo zahtevati zgolj pri postopkih avtomatiziranega odločanja, ki imajo znatni učinek na ljudi, katere odločitve teh sistemov zadevajo.

Kot sem pojasnil prej, imajo po prepovedi iz prvega odstavka 22. člena GDPR tovrstne odločitve veljavo zgolj v primeru opisanih izjem, med katerimi bo praktično najbolj pogosta uporaba na podlagi izrecnega soglasja. To sicer v primeru sistemov, ki se uporabljajo v zasebnem sektorju teoretično ni sporno. Problematična je uporaba, ki temelji na privolitvi v javnem sektorju, saj je, kot opozarja tudi delovna skupina WP29¹¹⁵ vprašljivo, če je v odnosu do obdelave podatkov v javnem sektorju sploh možno podati prostovoljno privolitev, saj zaradi pomembnosti storitev, ki jih nudi javni sektor in pomanjkanja alternativ, posameznik, na katerega se podatki navezujejo tovrstne obdelave dejansko ne more zavrniti. Recital 43 GDPR namreč

¹¹³ Ibidem, str. 479-481.

¹¹⁴ Borgesius, F. Z.: Discrimination, artificial intelligence, and algorithmic decision-making, Directorate General of Democracy, Council of Europe, Strasbourg, 2018, str. 35.

¹¹⁵ European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, version 1.1, 2020, str. 8

določa, da za »zagotovitev, da je privolitev dana prostovoljno, privolitev ne bi smela biti veljavna pravna podlaga za obdelavo osebnih podatkov v posebnem primeru, ko obstaja očitno neravnotežje med posameznikom, na katerega se nanašajo osebni podatki, in upravljavcem, zlasti kadar je upravljavec javni organ in je zato malo verjetno, da je bila privolitev dana prostovoljno v vseh okoliščinah te specifične situacije.«¹¹⁶ Za avtomatizirano odločanje v kontekstu javnega servisa, ki ima učinke, ki so potrebni za veljavo prepovedi prvega odstavka 22. člena GDPR je zatorej smiselno, da je uporaba avtomatiziranega odločanja dopustna zgolj v primeru, da je dovoljena v pravu države članice EU. V tem primeru morejo biti določeni tudi ustrezni ukrepi za zaščito pravic in svoboščin ter zakonitih interesov posameznikov, na katere se nanašajo osebni podatki. Menim, da je v takšnem primeru potrebna še večja zaščita posameznikov, čigar osebni podatki se obdelujejo in da so v takšnem primeru prej opisani pogoji uporabe ne le utemeljeni temveč potrebni.

5.10 Zaključek

Med tem ko umetna inteligenca omogoča precej natančne napovedi in formalno izključitev obravnave določenih lastnosti posameznikov, se moramo zavedati možnosti, da ima lahko uporaba tovrstne tehnologije za namen odločanja diskriminatoren učinek in k uporabi takšnih sistemov pristopati z vsaj takšno mero previdnosti, kakor pri človeškem odločanju. Odločanje, ki temelji na avtomatiziranih sistemih je lahko na prvi pogled vsekakor zastrašujoče, še posebej kadar gre za sisteme namenjene množičnemu odločanju, ki temeljijo na strojnem učenju in njihov postopek odločanja pogosto ni popolnoma razumljiv niti upravljavcem teh sistemov. Avtomatizirane odločitve lahko in pogosto imajo pomemben učinek na posameznika, katerega zadevajo, zaradi česar je popolnoma upravičen strah tega posameznika, da ta odločitev ne bo pravična in da bo zgolj del nepojasnljive statistike. Kljub temu, da je najbolj privlačna lastnost avtomatiziranega odločanja domnevna popolna nevtralnost tovrstnih sistemov, strah in nezaupanje do teh sistemov verjetno temelji na pomanjkanju človeškega stika, saj v tem primeru sklepamo, da je odločevalcu vsaj nekoliko mar in da nismo zgolj ime na seznamu. Seveda realno stanje po navadi ne ustreza takšni predstavi.

Glavni vir regulacije avtomatiziranih sistemov predstavlja GDPR. Ta sicer dopušča razvoj umetne inteligence, v uravnoteženem razmerju zaščite osebnih podatkov in

¹¹⁶ Recital 43 GDPR.

ekonomskih interesov, vendar nudi zgolj omejena navodila, kako to razmerje doseči. Vsebuje namreč obilje načel in odprtih standardov, zadostitev katerih pogosto zahteva tehtanje nasprotujočih si interesov. V primeru umetne inteligence je ta nejasnost še toliko bolj izrazita, zaradi novosti in kompleksnosti te tehnologije in širokega spektra njenih učinkov.¹¹⁷

Pogosta kritika prava je, da stalno zaostaja s tehnološkim razvojem, vendar je pomembno, da z regulacijo tehnologije ne prehitavamo. Postavljanje nerealnih pogojev za uporaba avtomatiziranih sistemov zgolj ovira razvoj in uporabo teh sistemov. Dodatna regulacija morda sploh ni rešitev problema. Glede na to, da je pretežni del uporabe tovrstnih sistemov pogojen z izrecnim soglasjem posameznikov, katerih podatki se obdelujejo, je morda čas, da tudi ti posamezniki prevzamejo smiselno odgovornost za zaščito lastnih interesov. Dodatna regulacija, ki temelji na odprtih standardih in načelih (bolj natančna regulacija je zaradi specifične narave obravnavane tehnologije izjemno težavna) zaradi nejasnosti zgolj zmanjšuje konkurenčnost trga EU za razvoj tovrstne tehnologije. Informacijska ozaveščenost in zaščita osebnih podatkov sta temi, ki sta večini ljudi tuji in dolgočasni, saj se ne zavedajo mogočih implikacij njihovih »odločitev« pri uporabi digitalnih storitev in deljenju osebnih podatkov. Vprašanje je, kdaj bosta ti temi za mnoge postali zanimivi. Za marsikoga morda prepozno ali celo nikoli.

Seznam literature in virov

Članki in poglavja iz knjig

- Barocas, S., Selbst, A., D.: Big Data's Disparate Impact, v: *California Law Review*, 104 (2016) 617, str. 672-678
- Borgesius, F. Z.: Discrimination, artificial intelligence, and algorithmic decision-making, Directorate General of Democracy, Council of Europe, Strasbourg, 2018.
- Gillis, T. B., Spiess, J.: Big Data and Discrimination, v: *University of Chicago Law Review*, 86 (2019) 459.
- Gillis, T. B.: False Dreams of Algorithmic Fairness: The Case of Credit Pricing, <<https://projects.iq.harvard.edu/fintechlaw/publications/false-dreams-algorithmic-fairness-case-credit-pricing>>
- Han, J. Pei, J. Kamber, M.: *Data Mining: Concepts and Techniques*, 2. izdaja, Morgan Kaufman Publishers, San Francisco, 2006, str. 6
- Mendoza, I., Bygrave, L. A., The Right Not To Be Subject to Automated Decisions Based on Profiling, v: *Universitz of Oslo Faculty of Law Legal Studies Research Paper Series*, (2017) 2017-20.
- Sartor, G.: The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, European Parliamentary Research Service, Bruselj, 2020.

Pravni viri

¹¹⁷ Sartor, G.: The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, European Parliamentary Research Service, Bruselj, 2020, str. 7.

- Ustava Republike Slovenije (Uradni list RS, št. 33/91-I, 42/97 – UZS68, 66/00 – UZ80, 24/03 – UZ3a, 47, 68, 69/04 – UZ14, 69/04 – UZ43, 69/04 – UZ50, 68/06 – UZ121,140,143, 47/13 – UZ148, 47/13 – UZ90,97,99, 75/16 – UZ70a in 92/21 – UZ62a).
- Zakon o varstvu pred diskriminacijo (Uradni list RS, št. 33/16 in 21/18 – ZNOrg).
- Zakon o ratifikaciji Konvencije o varstvu človekovih pravic in temeljnih svoboščin, spremenjene s protokoli št. 3, 5 in 8 ter dopolnjene s protokolom št. 2, ter njenih protokolov št. 1, 4, 6, 7, 9, 10 in 11 (Uradni list RS – Mednarodne pogodbe, št. 7/94).
- Pogodba o Evropski uniji, Uradni list Evropske unije, C 326, 26.10.2012, str. 13–390.
- Listina Evropske Unije o človekovih pravicah, Uradni list Evropske unije, C 83/389, str. 391–407.
- Direktiva Sveta 2000/43/ES z dne 29. junija 2000 o izvajanju načela enakega obravnavanja oseb ne glede na raso ali narodnost, Uradni list Evropske unije, L 180, 19.7.2000, str. 22–26.
- Direktiva Sveta 2000/78/ES z dne 27. novembra 2000 o splošnih okvirih enakega obravnavanja pri zaposlovanju in delu, Uradni list Evropske unije, L 303, 2.12.2000, str. 16–22.
- Direktiva Sveta 2004/113/ES z dne 13. decembra 2004 o izvajanju načela enakega obravnavanja moških in žensk pri dostopu do blaga in storitev ter oskrbi z njimi, Uradni list Evropske unije, L 373, 21.12.2004, str. 37–43.
- Direktiva 2006/54/ES Evropskega parlamenta in Sveta z dne 5. julija 2006 o uresničevanju načela enakih možnosti ter enakega obravnavanja moških in žensk pri zaposlovanju in poklicnem delu (preoblikovano), Uradni list Evropske unije, L 204, 26.7.2006, str. 23–36.
- Uredba (EU) 2016/679 Evropskega parlamenta in Sveta z dne 27. aprila 2016 o varstvu posameznikov pri obdelavi osebnih podatkov in o prostem pretoku takih podatkov ter o razveljavitvi Direktive 95/46/ES (Splošna uredba o varstvu podatkov) (Besedilo velja za EGP), Uradni list Evropske unije, L 119, 4.5.2016, str. 1–88.
- Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016 679.
- Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679, 2017
- European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, version 1.1, 2020

Sodna praksa

- C-673/17, *Planet 49*, ECLI:EU:C:2019:801.
- C-236/09, *Test-Achats*, ECLI:EU:C:2011:100
- C-157/15, *Abbita*, ECLI:EU:C:2017:203
- Moustaquim v Belgium*, št. 12313/86, ECLI:CE:ECHR:1991:0218JUD001231386.
- Thlimmenos v Greece*, št. 34369/97, ECLI:CE:ECHR:2000:0406JUD003436997.

Spletni viri

- <<http://cyberlaw.stanford.edu/blog/2019/02/electronic-contracts-and-illusion-consent>> (13.4.2020).