

# ZBORNİK POVZETKOV

Inštitut za kriminologijo pri Pravni fakulteti v Ljubljani  
v okviru Ciljnega raziskovalnega projekta  
Človekove pravice in regulacija umetne inteligence vredne zaupanja  
organizira konferenco

## UMETNA INTELIGENCA, ČLOVEKOVE PRAVICE IN DRUŽBENA ŠKODA



15. junij 2021 | 9:30 - 15:30 | Zoom  
Obvezne predhodne prijave do ponedeljka, 14. junija 2021, do 12:00

## O projektu

### Človekove pravice in regulacija umetne inteligence vredne zaupanja

Umetna inteligenca (UI) prinaša številne prednosti za posameznike, podjetja in države. Primeri segajo od situacij, ko družbena omrežja priporočajo prijatelje, algoritmi, ki osmišljajo veliko podatkovje opravijo del transakcij na borzah ali pa organom odkrivanja in pregona pokažejo, kje in kdaj naj bi se zgodil prihodnji zločin. Evropska unija je z vseevropsko pobudo o sodelovanju na področju umetne inteligence, ki se ji je pridružila tudi Slovenija začrtala prizadevanja, da je smiselno, da z velikim podatkovjem, strojnim učenjem oz. UI poskušamo reševati tudi številne družbene probleme: revščino, klimatske spremembe, družbeno neenakost in kriminaliteto. Nujno pa je, da se tehnološki razvoj UI odvija v ustreznem pravno-etičnem okvirju, ki bo okrepil obstoječe ravni spoštovanja človekovih pravic in temeljnih svoboščin vsakega posameznika in utrdil legitimne demokratične procese.

Naše raziskovalno delo je razdeljeno na tri dele. V prvem delu, *Atlas uporabe UI*, analiziramo rabo UI v družbenih domenah in družbenih podsistemih, kjer UI lahko vpliva na temelje demokratične družbene ureditve in pravne države. V tem delu se med drugim ukvarjamo tudi s preučevanjem obstoječih uporab UI v javni politikah. Na podlagi študij primerov dobrih praks uporabe UI v javnih

politikah v tujini in analize znanja iz že obstoječih platform želimo oblikovati smernice kako dobre prakse in znanja uspešno prenesti in implementirati v sfero javnih politik v Republiki Sloveniji.

V drugem delu projekta, *Učinki UI na človekove pravice in posameznika*, je naš cilj oblikovati kvalitativne analize rabe UI v različnih domenah. Dva ključna učinka UI, ki jih projekt naslavlja, sta vplivi UI na temeljne človekove pravice (pravni vidiki) ter vplivi na človekovo subjektivnost (nepravni oz. psiho-socialni vidiki). Izhodišče projekta je, da mora vpeljava UI temeljiti na spoštovanju temeljnih človekovih pravic in temeljnih svoboščin, demokracije, pravne države in zagotavljanju gospodarske in politične stabilnosti. Raznoliki primeri obstoječe uporabe UI kažejo, da ob številnih koristih novi sistemi okrepljeni z UI lahko vodijo tudi v kršitve temeljnih človekovih pravic. Grožnje so specifične za posamično domeno – npr. zavarovalniško, bančno, sodno, policijsko – nekatere pa so skupne delovanju UI. Primeri uporabe UI kažejo, da ta prinaša številne pravne in etične izzive. Na mikro ravni UI uporabljena v sistemih odločanja o posameznikovih pravicah in obveznostih vpliva na številne individualne človekove pravice, na primer na zasebnost in varstvo osebnih podatkov, načelo enakosti in socialne pravičnosti, na avtonomijo posameznika in na njegovo dostojanstvo. Na mezo in makro družbeni ravni pa prihaja do vplivov na delovanje demokracije in povečanje ekonomske neenakosti ter vodi v družbeno razvrščanje prebivalstva.

Zavedamo se, da UI vstopa v specifičen socialno, kulturno-političen prostor obenem pa je tudi produkt specifične družbeno-ekonomske situacije, ki jo povratno tudi sama pomaga sooblikovati. UI je posledično lahko uporabljena tako za krepitev »negativnih« družbenih sprememb, npr. povečevanje ekonomske neenakosti na globalni ravni, širitev družbeno razdiralnih in sovražnih politik kot tudi sredstvo odpravljanja takšnih družbenih sprememb. V okviru projekta je naš cilj tako analizirati pravne dileme na izbranih pravnih področjih in identificirati nad-domenske, tj. skupne problematične značilnosti uporabe AI v odločevalskih procesih, kot so npr. netransparentnost sistemov odločanja z UI; težave povezane s kvaliteto in pristranostjo podatkov; samouresničujoče se prerokbe pri preveliki odvisnosti od preteklih učnih primerov; nevarnosti podatkovnega maksimiranja; nesorazmerje moči, ki nastaja zaradi uporabe UI; nevarnosti podatkovnega determinizma in diskriminacije; zastraševalni učinek na pravice; ustvarjanje »odmevnih soban« pri personaliziranih storitvah ipd.

V tretjem delu, *Regulacija in upravljanje UI*, predhodno analizirane rabe UI postavljamo v etični in pravni okvir. V tem delu preučujemo različne etike in vrednote, ki jih imamo v mislih, ko govorimo o etiki v umetni inteligenci. Ena imed osrednjih tematik s katero se bomo ukvarjali je tudi varstvo osebnih podatkov, saj je UI doživela preporod ravno v času sprememb v intenzivnosti in obsegu zbiranja, obdelave in dostopa do osebnih podatkov. Temeljna načela varstva osebnih podatkov v EU so se razvila v obdobju pred digitalno revolucijo, zato pogosto nasprotujejo realnosti zbiranja in obdelave podatkov v dobi velikega podatkovja in UI.

Ker gre pri UI za globalne spremembe, ki jih ni mogoče regulirati samo na nacionalni ravni, se vprašanja odgovornosti UI, transparentnosti in posledic za človekove pravice že obravnavajo na mednarodnih forumih, npr. v okviru OECD, UNESCO, G20, EU, vendar celovit model upravljanja AI zanekrat še ni razvit. Mednarodno sodelovanje lahko zagotovi učinkovito globalno upravljanje, vendar pa se je obenem potrebno zavedati, da na področju UI hiter tehnološki razvoj in nesorazmernost informacij med oblikovalci politik na eni strani in skupnostjo, ki UI razvija in uporablja na drugi, resno ogrožata možnost vseh vpletenih deležnikov, da bi vprašanja, ki se odpirajo učinkovito naslovili. V okviru projekta se med drugim ukvarjamo z vprašanjem, kako zgraditi celovit in zanesljiv globalni režim upravljanja UI po zgledu modelov dobrega upravljanja na drugih področjih. Ker je učinkovito in odgovorno upravljanje mogoče izvesti le z dobro razvitimi kazalniki razvitosti UI je eden izmed ciljev našega projekta tudi vzpostavitev načrta enotnega informacijskega vira, ki bi ponujal celoviti vpogled v razvojne trende znotraj zapletenega področja UI.

Ciljni raziskovalni projekt (CRP) Človekove pravice in regulacija umetne inteligence vredne zaupanja (V5-1930) je financiran s strani Javne agencije za raziskovalno dejavnost Republike Slovenije (RS), Ministrstva za zunanje zadeve RS, Ministrstva za javno upravo RS, Ministrstva za izobraževanje, znanost in šport RS ter Ministrstva za pravosodje RS.



# Program konference

## 1. Sekcija (9:30- 11h)

Moderira: Katja Simončič

- GREGOR STROJIN: Pristopi mednarodne skupnosti k regulaciji UI
- ALEŠ ZAVRŠNIK: Dostop do podatkov in umetna inteligenca: primer računalniškega vida
- TIM HORVAT: Vpliv političnega mikrociljanja na človekove pravice in demokratični proces
- KATJA SIMONČIČ: Dileme uporabe umetne inteligence v javni upravi
- Debata

Odmor za kavo 30 min

## 2. Sekcija (11:30h- 13:h)

Moderira: Pika Šarf

- MAJA BOGATAJ JANČIČ: Ali je lahko umetna inteligenca avtor avtorskega dela?
- MATIJA DAMJAN: Uporaba umetne inteligence za moderiranje spletnih vsebin
- MATJAŽ JAGER: Notranje napetosti med aktualnimi pristopi k etičnemu programiranju avtonomnih vozil
- LJUPČO TODOROVSKI: Algoritmčna etika
- PIKA ŠARF: Prvi obris regulacije umetne inteligence v EU
- Debata

Odmor za kosilo

## 3. Sekcija (14h-15:30h)

Moderira: Marko Drobnjak

- LARA DULAR: Tehnologija obrazne prepoznavne skozi prizmo človekovih pravic
- JAN ČEJVANOVIČ: Pravna jamstva pri uporabi umetne inteligence v kazenskem postopku skozi relevantno sodno prakso
- ANA BABNIK: Reševanje bagatelne kriminalitete preko algoritmov umetne inteligence
- MARKO DROBNJAK: Odločitve Vrhovnega sodišče Republike Slovenije kot jezikovni vir za procesiranje naravnega jezika: prepovedan sadež ali prvi korak k umetni inteligenci vredni zaupanja
- Debata

## **Povzetki prispevkov**

### **ANA BABNIK: Reševanje bagatelne kriminalitete preko algoritmov umetne inteligence**

Ker bagatelna kriminaliteta predstavlja znatno obremenitev za organe odkrivanja in tudi za organe pregona, je namen prispevka predstaviti, kako bi jim uporaba umetne inteligence lahko pomagala pri odkrivanju in reševanju takšne oblike kriminalitete.

V prispevku bodo zato kratko predstavljene nekatere rešitve, ki organom ponekod že olajšujejo delo in njihova uporabljivost z vidika pregona bagatelne kriminalitete (npr. predvidevanje t.i. žariščnih točk («hot spots»), zaznavanje kriminalitete v realnem času, tehnologije obrazne prepoznave ter pojav robota odvetnika). Kot še posebej zanimiva pa bo izpostavljena rešitev Republike Estonije, ki je zasnovala t.i. avtomatiziranega sodnika, ki lahko brez posebnega posredovanja sodnika rešuje spore majhne vrednosti v višini do 7.000 evrov.

### **MAJA BOGATAJ JANČIČ: Ali je lahko umetna inteligenca avtor avtorskega dela?**

V času hitrega tehnološkega napredka se morebiti že bližamo trenutku, ko bo umetna inteligenca lahko samostojno generirala stvaritve brez kakršnegakoli sodelovaja človeka. Zato je povsem razumljivo, da je zmeraj bolj aktualno vprašanje ali so lahko stvaritve, ki jih samostojno ustvari umetna inteligenca, varovane z avtorsko pravico? Ali je za varstvo dovolj, da rezultat, ki ga ustvari stroj, izgleda enako, kot stvaritev, ki jo ustvari človek? Ali umetna inteligenca sploh potrebuje nagrado oziroma spodbudo za ustvarjanje? Ali lahko umetno inteligenco štejemo za "avtorje", ki ustvarjajo "izvirna" dela. Veljavna avtorsko-pravna zakonodaja da povsem jasen odgovor, saj določa, da brez kreativnih odločitev človeka, ki se odražajo v izvirnosti avtorskega dela, ne moremo govoriti o stvaritvah, ki si zaslužijo avtorsko- pravno varstvo. V prispevku bodo tako predstavljeni glavni izzivi, ki jih razvoj umetne inteligence zastavlja tradicionalnim konceptom avtorskih pravic, in predstavljene smeri možnih rešitev.

### **JAN ČEJVANOVIČ: "Pravna jamstva pri uporabi umetne inteligence v kazenskem postopku skozi relevantno sodno prakso"**

COMPAS je v Združenih državah Amerike uveljavljen računalniški program, ki je sodniku v pomoč pri ugotavljanju kakšna stopnja verjetnosti je podana pri posameznem obdolžencu, da bo v določenem časovnem obdobju ponovil kaznivo dejanje oziroma se kazenskemu postopku izmikal. Na primeru navedenega avtomatiziranega sistema se bom v prispevku opredelil do nekaterih glavnih zadržkov pri vpeljavi umetne inteligence v kazenski postopek.

Glavni poudarek prispevka bo na sodni praksi ameriških sodišč. COMPAS je bil namreč že večkrat predmet obravnave pred sodišči različnih ameriških zveznih držav, ki so s svojimi odločitvami vzpostavila pravne standarde v zvezi s tem, kaj je potrebno zagotoviti, da so spoštovana pravna jamstva pri uporabi umetne inteligence v kazenskem postopku.

### **MATIJA DAMJAN: Uporaba umetne inteligence za moderiranje spletnih vsebin**

Zakonodaja Evropske unije od upravljalcev spletnih platform in družbenih omrežij vse bolj zahteva proaktivno ravnanje z visoko skrbnostjo, da v okviru svojih storitev preprečijo dostop do nezakonitih uporabniških vsebin. Takšne določbe najdemo v Direktivi o avtorski pravici na enotnem digitalnem trgu, v Uredbi o preprečevanju razširjanja terorističnih spletnih vsebin in v osnutku nove uredbe o digitalnih storitvah. Glede na ogromne količine informacij, ki jih vsakodnevno objavljajo spletni uporabniki, si je težko zamisliti, da bi odstranjevanje nezakonitih vsebin potekalo brez uporabe avtomatiziranih programskih orodij. Pri tem se algoritmi umetne inteligence pogosto predstavljajo kot rešitev, ki bo znala razpoznati sovražni govor ali teroristične vsebine in sploh natančno razločevati med zakonitimi in nezakonitimi objavami. Prispevek bo razpravljal o tehnični izvedljivosti takšnega razlikovanja, pa tudi o načelnih pomislekih zoper vnaprejšnje algoritmsko filtriranje objav, ki presegajo vprašanje točnosti zaznavanja nezakonitih vsebin.

### **MARKO DROBNJAK: Odločitve Vrhovnega sodišče Republike Slovenije kot jezikovni vir za procesiranje naravnega jezika: prepovedan sadež ali prvi korak k umetni inteligenci vredni zaupanja**

Pisni viri so osrednji vir raziskovanja sodne prakse, medtem ko percepcija govora in njen vpliv na percepcijo posameznikov v sodnih postopkih in nenazadnje na transkripcijo izrečenega na narokih pogosto ostane spregledana. Avtor prispevka bo zato sprva v uvodnem

delu predstavil, kako percepcija govora vpliva na sodno odločanja ter posredno ali neposredno tudi na poznejše pisne vire in sodno odločanje. V trenutku, ko je celoten sodni postopek zreduciran zgolj na končno odločitev, je namreč vsebina izrečenega na ustnih narokih lahko izgubljena med vrsticami oz. omejena na interpretirane ali dobesedne transkripte. Z zavedanjem predstavljenih omejitev bo avtor v nadaljevanju predstavil potek luščenja podatkov iz spletnega arhiva sodne prakse in slednje uporabil kot jezikovni vir za procesiranja naravnega jezika ter raziskovanje sodnih odločitev. V sklepnem delu bodo predstavljeni preliminarni podatki z distribucijo besedišča raziskovalnega obdobja odločitev Vrhovnega sodišča Republike Slovenije, kar predstavlja obdobje od leta 1992 do 2020 z več kot 50.000 odločitvami. Dodana bo tudi analiza z najpogostejšo zakonodajo, na katero se sklicujejo odločitve ter prikaz izhodišča za raziskovanje vpliva senatov na končne odločitve. Razprava bo temeljila na predstavljenih rezultatih in bo poskušala nasloviti vprašanje: ali so lahko odločitve sodišč primeren vir za procesiranja jezika, ki bo pripomogel k prvemu koraku do umetne inteligence vredne zaupanja, ali gre za prepovedan sadež.

### **LARA DULAR: Tehnologija obrazne prepoznave skozi prizmo človekovih pravic**

Tehnologija obrazne prepoznave v zadnjih nekaj letih doživlja intenziven razvoj in hkrati prodira na vsa področja naših življenj (javno in zasebno). Številna pravna in etična vprašanja sproža predvsem uporaba tovrstnih orodij v javnem prostoru. Na kratko bom predstavila tehnologijo obrazne prepoznave kot obliko biometričnega sistema, napake, do katerih privedejo pristranosti podatkov v biometričnih sistemih, nato pa se bomo sprehodili po zemljevidu pravne regulacije v Sloveniji, Evropi in globalno (pravna regulacija je v tem smislu zelo raznolika) ter osvetlili zorne kote skozi prizmo prepovedi diskriminacije, pravice do zasebnosti in varstva osebnih podatkov ter pravice do poštenega sojenja. Omenjen bo tudi vpliv tehnologije na ranljive skupine. V zaključnem delu bom na kratko predstavila sklepne misli avtoric glede prihodnosti tehnologije obrazne prepoznave v luči sprememb zaradi pandemije covid-19 in načrtov EU, da na ravni Unije postavi pravila glede njene uporabe v sklopu regulacije umetne inteligence.

### **TIM HORVAT: Vpliv političnega mikrociljanja na človekove pravice in demokratični proces**

Politični akterji na Zahodu dandanes vse pogosteje uporabljajo tehnologijo političnega mikrociljanja za doseganje političnih ciljev. Ta tehnologija združuje veliko podatkovje z behaviorističnimi tehnologijami in digitalnim oglaševanjem na družbenih platformah, tako da lahko posamezniku pošilja personalizirana sporočila na podlagi prediktivnih modelov, ki uporabljajo baze osebnih podatkov o posameznikih, sporočila pa se stalno prilagajajo odzivu posameznika. Prispevek poveže politično mikrociljanje s filtrirnimi mehurčki in fenomenom lažnih novic, saj si delajo skupne lastnosti – uporabo velikega podatkovja, behavioristične tehnologije in suspenz statusa resnice – vendar ga obravnava posebej. V prispevku je najprej prikazano delovanje političnega mikrociljanja, nato pa so analizirani njegovi vplivi na pravico do zasebnosti, svobodo izražanja, pravico do obveščenosti in varstvo posameznikove avtonomije. Posebej je obravnavan vpliv političnega mikrociljanja na demokratični proces, kot ga ureja pravo, pokazane pa so tudi njegove inherentne omejitve. Na koncu so predstavljene možnosti regulacije, ki segajo od zgolj zagotavljanja transparentnosti političnih oglasov do omejitev financiranja kampanj, trgovanja s podatki ter političnega profiliranja, ter naposled do vsebinskega omejevanja izražanja na spletu, ki je kritično analizirano. Avtor problematiko političnega mikrociljanja umesti tudi v širši kontekst krize predstavniške demokracije in statusa Resnice.

### **MATJAŽ JAGER: Notranje napetosti med aktualnimi pristopi k etičnemu programiranju avtonomnih vozil**

V svojem prispevku bom na kratko skiciral tri aktualne pristope k etičnemu programiranju popolnoma avtonomnih vozil. Ti pristopi so: 1) od zgoraj navzdol, 2) od spodaj navzgor in 3) pristop, ki kombinira prva dva. Vse tri pristope lahko opazujemo na primeru Zvezne republike Nemčije, ki je ena najambicioznejših držav v iskanju rešitev na tem področju. Cilj vseh prizadevanj je družbeno soglasje o rešitvah, ki bodo vgrajene v avtonomna vozila. Pokazal bom, kako je pot do zastavljenega cilja veliko bolj makadamska, kot je morda kdo optimistično predvideval na začetku tega procesa. V katero smer bi lahko šle rešitve?

### **KATJA SIMONČIČ: Vpliv umetne inteligence na človekove pravice in na nastanek družbene škode v javni upravi**



Izhodišče poglavja je vseprisotnost uporabe umetne inteligence (UI) v različnih družbenih domenah ter dejstvo, da se njene razširjenosti velik del državljanov sploh ne zaveda. Poznavanje razširjenosti uporabe UI je ključna za posameznike na čigar življenja vpliva malodane na vsakem koraku, saj zgolj osveščeni državljani od vlad lahko zahtevajo ustrezne zakonodajne rešitve, ki zmanjšujejo tveganja, povezana s splošno rabo UI. Cilj tega poglavja je tako (1) predstaviti pojavnost tehnologije UI na področju javne uprave in orisati tveganja za človekove pravice in potencialno škodo, do katerih lahko pripelje pravna neurejenost uporabe UI.

### **PIKA ŠARF: Prvi obris regulacije umetne inteligence v EU**

Evropska Komisija je 21. aprila 2021 predstavila predlog uredbe, ki postavlja harmonizirana pravila o umetni inteligenci, ki je kljub temu, da ne gre za končno verzijo zakonodajnega besedila, ponudil prvi konkretniji vpogled v prihodnost regulacije umetne inteligence v EU. Predlog Uredbe UI sledi ciljem, ki jih je leta 2020 načrtala Komisija v Beli knjigi o umetni inteligenci in so utemeljeni na načelih, ki jih je izoblikovala Strokovna skupina na visoki ravni o umetni inteligenci, predvsem načelih človeškega nadzora, tehnične robustnosti in varnosti, spoštovanju načel varstva osebnih podatkov, transparentnosti, pravičnosti in nediskriminatornosti ter doslednem uveljavljanju odgovornosti. V prispevku bodo predstavljeni njeni ključni poudarki in kritično ovrednotene rešitve, ki jih predlaga Komisija

### **LJUPČO TODOROVSKI: Algoritmična etika (Computational Ethics)**

V članku, objavljenem v reviji *Artificial Intelligence and Society*, Samuel Segun zagovarja ustanovitev področja algoritmične etike. V nasprotju s prevladujočim diskurzom tradicionalne robotske etike na eni strani in novejšje strojne etike na drugi strani bi se algoritmična etika osredotočala na praktično vprašanje algoritmične izračunljivosti etičnih odločitev (Segun 2020). Tako je cilj algoritmične etike zagotoviti računalniškimi znanstvenikom in inženirjem tehnični okvir za gradnjo inteligentnih sistemov, ki so sposobni sprejemati etične odločitve. Tak okvir bi moral vključevati tudi minimalne standarde, ki jim mora zadostovati računalniški agent z etičnim obnašanjem.

Algoritmična etika je zato bolj pragmatična in ima ožji, bolj tehnični poudarek kot bolj splošni etični okvirji robotov in strojev. Prvi naslavlja širše vprašanje o vplivu razvoja robotov

na etično in socialno interakcijo z ljudmi: tak okvir obravnava pragmatično vprašanje zasnove etičnih robotov le posredno, implicitno (Kuipers 2018). Drugo povezani okvir strojne etike se pa ukvarja s preučevanjem močne povezave med inteligentnimi agenti in naše pravno, etične, socialne in ekonomske krajine (Torrance 2013).

Številnim prejšnjim raziskavam in pregledom na temo robotske in strojne etike, glej npr. (Anderson in Leigh Anderson 2011), navkljub, so pregledi pristopov na specifičnem področju algoritmične etike redki (Segun 2020). Cilj tega prispevka je definirati obseg področja algoritmične etike in ilustrirati njen pomen skozi predstavitev izbranih pristopov.

### **ALEŠ ZAVRŠNIK: Dostop do podatkov in umetna inteligenca: primer računalniškega vida**

Pri zbirkah podatkov, potrebnih za strojno učenje, je pomembno, da so te nastale na *pravno sprejemljive in dopustne načine*. V raznovrstnih domenah (npr. računalniški vid, jezikovne tehnologije) so te zbirke nastale na samoiniciativno računalničarjev in v dobri veri, da z njihovo pomočjo lahko učijo algoritme in trenirajo statistične modele. Zbirke so pripravili znanstveniki za lastne potrebe in najboljši med njimi so jih dali na voljo (svetovni) znanstveni skupnosti, v javno dobro, v obliki odprtega dostopa, ki naj pospeši razvoj (računalniške) znanosti. Zbirka *Labeled Faces in the Wild*, podatkovna zbirka fotografij obrazov sestavljena za študij problemov, povezanih z neomejenim prepoznavanjem obrazov slavnih in javnih oseb, ki so bile pridobljene iz časopisov, filmov, posnetkov javnih dogodkov, je na primer prejela kar najvišjo strokovno nagrado *Mark Everingham Award* za služenje znanstveni skupnosti za računalniški vid. Ali so številne zbirke oblikovane z nedvomno plemenitimi nameni – pospešiti razvoj znanosti in izdelati nove tehnologije, ki bodo olajšala številna dela, povečala učinkovitost itn. – pravno skladne z vidika prava varstva osebnih podatkov. Konflikt pri zbiranju podatkov za namene strojnega učenja ne nastaja samo pri računalniškem vidu, temveč pri vseh področjih, kjer gre za obdelavo osebnih podatkov. V prispevku bo prikazano na primeru računalniškega vida, kot paradigmatičnega primera, kako prihaja do trkov med znanstvenim napredkom na področju umetne inteligence na eni strani in obstoječim pravnim režimom prava varstva osebnih podatkov na drugi strani.

Zbornik pripravila: Katja Simončič, Inštitut za kriminologijo pri Pravni fakulteti v Ljubljani