

Odgovorni razvoj i primjena umjetne inteligencije

Javorović, Luka

Undergraduate thesis / Završni rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences / Sveučilište Josipa Jurja Strossmayera u Osijeku, Filozofski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:142:155019>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-16**



Repository / Repozitorij:

[FFOS-repository - Repository of the Faculty of Humanities and Social Sciences Osijek](#)



Sveučilište Josipa Jurja Strossmayera Osijek

Filozofski fakultet Osijek

Preddiplomski studij Informatologije

Luka Javorović

Odgovorni razvoj i primjena umjetne inteligencije

Završni rad

doc. dr. sc. Milijana Mičunović

Osijek, 2022.

Sveučilište Josipa Jurja Strossmayera Osijek

Filozofski fakultet Osijek

Odsjek za informacijske znanosti

Preddiplomski studij Informatologije

Luka Javorović

Odgovorni razvoj i primjena umjetne inteligencije

Završni rad

Društvene znanosti, informacijske i komunikacijske znanosti, informacijski
sustavi i informatologija

doc. dr. sc. Milijana Mičunović

Osijek, 2022.

IZJAVA

Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napravio te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citati s napisanim izvorom odakle su preneseni. Svojim vlastoručnim potpisom potvrđujem da sam suglasan da Filozofski fakultet Osijek trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta Osijek, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

U Osijeku, datum

21.09.2022.



0125151185

ime i prezime studenta, JMBAG

Sadržaj

1. Uvod.....	1
2. Odgovorni razvoj umjetne inteligencije.....	2
2.1 Općenito o umjetnoj inteligenciji.....	2
2.2 Regulacija umjetne inteligencije.....	7
3. Pravilnici i smjernice za regulaciju umjetne inteligencije.....	12
3.1 Pravilnici i smjernice državnog (javnog) sektora.....	12
3.2 Pravilnici i smjernice privatnog sektora.....	17
4. Zaključak.....	18
5. Literatura.....	20

Sažetak

Cilj je ovog radi dati općeniti pregled pravilnika, smjernica i praksi za odgovorni razvoj umjetne inteligencije (UI) u okviru europskih, odnosno međunarodnih projekata. Odgovorni razvoj UI ima potencijal postati novim standardom u razvoju i primjeni UI, posebno u kontekstu Četvrte industrijske revolucije. No, potreban je oprez u razvoju takvih sveobuhvatnih sustava koji imaju veliki potencijal za inovacije i utječu na, odnosno potiču transformaciju svih aspekata društva. Stvaranje pravilnika i smjernica koje bi regulirale odgovoran razvoj i primjenu UI jedna je od najvažnijih tema u znanstvenoj i stručnoj zajednici u području razvoja UI, robotike i pametnih strojeva i sustava. Zahvaljujući razvoju strojnog učenja, inovacija u području UI i implementaciji inteligentnih tehnologija koje zajedno povećavaju učinkovitost i produktivnost sustava UI, pojavljuje se potreba za definiranjem strožih standardiziranih pravilnika i smjernica. Ovaj rad donosi pregled, opis i analizu pravilnika i smjernica u području odgovornog razvoja UI, odnosno pregled pravilnika i smjernica javnog (državnog) sektora i pravilnika i smjernica privatnog sektora. Poseban naglasak će biti na tehnologijama visokog rizika, poput potencijalnih sigurnosnih problema izravno vezanih uz implementaciju inteligentnih rješenja i sustava (npr. pitanje pristranosti, privatnosti podataka i društvenih i etičkih načela razvoja). Stvaranje jedinstvenog, sustavnog i sveobuhvatnog pravilnika u području odgovornog razvoja UI predstavlja planirani i strateški pristup razvoju i implementaciji UI te tako na odgovoran i siguran način omogućuje znanstveni i tehnološki, odnosno ekonomski i gospodarski razvoj društva. Također, takav pristup jamči sigurnost krajnjeg korisnika i smanjuje (pa čak i isključuje) mogući dehumanizirajući učinak visokotehnoloških rješenja i sustava.

Ključne riječi: odgovorni razvoj, pravilnici, regulacija, umjetna inteligencija

1. Uvod

Kako bi razumjeli smjernice za odgovorni razvoj umjetne inteligencije (UI), potrebno je shvatiti što je UI. Postoji više definicija UI koje taj pojam promatraju s različitih stajališta, primjerice računalne znanosti, filozofije i psihologije. Tokom svog povijesnog razvoja koncept UI je većinom bio shvaćan kao sposobnost računala da reproducira ili simulira kognitivne procese ljudi. Stvaranje veze između kognitivnih sposobnosti stroja i čovjeka samo je usporavalo razvoj ove tehnologije, a samim time takva veza je nametala ograničenja i nedostatke biološke neuroarhitekture sustavu koji nije ograničen takvom neuroarhitekturom i posjeduje mogućnost konstantnog nadograđivanja. Mogućnost nesmetanog unaprjeđivanja na razini algoritma ili na razini komponenti dokazuje kolika je razlika između strojne i ljudske inteligencije i kognitivnih sposobnosti. Neke su definicije jasnije i preciznije od drugih, no za potrebe ovog rada, najprikladnija je ona koju nudi Europski parlament, predstavljena u dokumentu Draft Report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012 (NIL)) koja UI opisuje kao “sustav programa i algoritama koji prikupljaju, obrađuju i interpretiraju strukturirane ili nestrukturirane podatke, identificiraju uzorke i stvaraju modele u svrhu donošenja zaključka ili reagiranja u materijalnoj ili digitalnoj dimenziji na temelju donesenih zaključaka”.¹ Ova definicija upućuje na glavne karakteristike UI, odnosno na to kako bi inteligentni sustavi u bliskoj budućnosti trebali djelovati i ponašati se, no trenutnu stanje UI možda najbolje opisuju Luciano Floridi koji je 2014. godine izjavio kako se trenutni inteligentni strojevi temelje na sintaktičkim odnosima, odnosno da se obrada neprotumačenih podataka događa na temelju pravila koja donose algoritmi kojima razumijevanje, tj. interpretacija i značenje podataka (semantički odnosi) nije toliko važno. Prema mišljenju Floridija, sadašnja UI nije dovoljno samostalna da bi u promjenjivom okolišu mogla djelovati bez intervencije i pomoći čovjeka.² Navedeno upućuje na zaključak da je UI još uvijek na razini reproduciranja inteligentnog ponašanja, a ne njegova produciranja. Floridijevu ideju podržava i Nick Bostrom koji tvrdi da je UI i dalje inferiorna ljudskoj inteligenciji, tj.

¹ Garcia del Blanco, Iban. REPORT with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies. 8.10.2022.

URL: https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_EN.html (2022-09-21)

² Floridi, L. The 4th Revolution: How the Infosphere is Reshaping Human Reality. Oxford: University Press, 2014.

umjetna inteligencija se i dalje oslanja na intervenciju i pomoć čovjeka kako bi iskoristila potencijal vlastite neuroarhitekture i kako bi djelovala na inteligentan način.³

Model UI koji se ne oslanja na ljudsku asistenciju u velikoj mjeri i koji omogućuje automatizirano pametno upravljanje industrijskim procesima i sustavima predstavlja temelj za uspješno provođenje industrijske revolucije 4.0 (IR4.0) gdje se velik broj industrijskih, ekonomskih i društvenih procesa i sustava temelji na implementaciji i radu pametnih sustava. Stoga je važno već sada početi stvarati odgovarajući okvir koji će poticati i omogućavati razvoj i implementaciju UI i drugih pametnih sustava uzimajući pritom u obzir potencijalne i postojeće rizike, ali ne ograničavajući iskorištavanje njihovih mogućnosti. Takvi okviri odnose se na koncept odgovornog razvoja UI.

Naredna poglavlja bave se problematikom odgovornog razvoja UI, a predstaviti će se i postojeći pravilnici i smjernice za odgovorni razvoj UI. U poglavlju “Odgovorni razvoj umjetne inteligencije” bit će, prije svega, objašnjen pojam UI i dan njezin kratki povijesni razvoj, a potom će se opisati i objasniti koncept odgovornog razvoja UI. Poglavlje “Pravilnici i smjernice za regulaciju UI” donosi pregled odabranih pravilnika i smjernica za odgovorni razvoj UI na razini javnog, tj. državnog sektora, kao i na razini privatnog sektora. U “Zaključku” se donose glavne ideje i spoznaje po pitanju odgovornog razvoja UI, odnosno regulacije razvoja i primjene UI i pametnih sustava.

2. Odgovorni razvoj umjetne inteligencije

2.1 Općenito o umjetnoj inteligenciji

Koncept inteligentnog stroja proizašao iz sfere znanstvene fantastike, konceptualnih radova i misaonih eksperimenata koji su se bavili problematikom UI postupno je postao znanstvenom stvarnosti. Moderno poimanje UI smješta inteligentne strojeve i algoritme u javnu, društvenu, ekonomsku i političku sferu povećavajući zanimanje političkih tijela za razvoj ovih naprednih tehnologija i njihovu implementaciju u društvo.

Od skromnih začetaka koncepta UI, znanstvenici pokušavaju definirati što je to uistinu inteligentan stroj, odnosno kako možemo utvrditi i dokazati da je stroj inteligentan. Pojam UI je od samog početka nosio određenu konotaciju kako se radi o stroju koji može

³ Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: University Press.

izvoditi zadatke na način koji bi čovjek definirao inteligentnim. Nažalost, zbog takvog arhaičnog i pristranog poimanja inteligencije i inteligentnih strojeva, područje UI nije podržavalo njezino autonomno određenje i razvoj. Definicija UI navedena u Uvodu poima UI kao autonomnu, a inteligentne strojeve kao strojeve koji imaju sposobnost obrade organiziranih ili neorganiziranih podataka te mogu donositi odluke i djelovati u predvidivom i nepredvidivom digitalnom i fizičkom okolišu. Takva definicija izmješta koncept inteligentnog i složenog algoritma izvan granica čovjekove asistencije i pristranog poimanja inteligentnog ponašanja i djelovanja utemeljenog na ideji ljudske inteligencije.

Sam koncept UI utemeljen je u filozofiji i matematici., Pokušaji prevođenja logičkih procesa u matematičke funkcije omogućili su razvoj ideje o inteligentnim strojevima. Od samih početaka veliku su ulogu u razvoju UI imale zamisli koje su predstavili Hobbes, Leibniz i Descartes i koje se nadovezuju na Aristotelovu silogističku logiku, a prema kojima se racionalna misao može prevesti u matematički oblik. Iako su navedene pretpostavke razvojem računalne znanosti bile dokazane, prvi stvaran korak u razvoju računalne znanosti i UI predstavlja analitički stroj Charlesa Babbagea i prvi program koji je stvorila Ada Lovelace preko kojeg je dobila izračune za Bernoulijevu jednadžbu. Babbage i Lovelace ostvarili su prve uspjehe u području računalne arhitekture i automatskih računalnih sustava, a logiku je dalje razvio George Boole koji je dodatno razradio koncept prevođenja apstraktnih misli u matematički oblik stvorivši Booleovu logiku koja predstavlja pravu revoluciju u području računalne logike. Boole je uspio matematičke operatore prevesti u prave logičke operatore. Korištenjem tih operatora mogli su se opisati procesi, odnosno operacije koje se događaju u računalnim elektroničkim sklopovima, a koje se 'ponašaju' poput električne struje u strujnim prekidačima.⁴

Pravi put prema inteligentnim strojevima i modernom shvaćanju računalne znanosti započeli su Alan Turing sa svojim pretpostavkama o "Univerzalnom Turingovu stroju" kao novim pristupom u razvoju računala. Slijedom Turingovih pretpostavki stroj trebao biti napravljen tako da može izvršavati sve naredbe koje su mu upućene na način da ih razumije, što je predstavljalo velik konceptualni odmak od tadašnje paradigme o računskim strojevima. Uz Turinga potrebno je spomenuti i Claudea Shannona i njegovu matematičku teoriju informacija koja tvrdi kako je sve informacije moguće prevesti u binarni sustav i kojom je, uz utemeljenje informacijske znanosti, stvorio temelje i za razvoj napredne strojne logike.

⁴ McCorduck, Pamela. *Machines Who Think*, Natick, Massachusetts, USA. Peters, Ltd. 2004. Str. 37.-59.

Cijeli ovaj niz događaja rezultirao je stvaranjem nove grane računalne znanosti. Sam izraz „umjetna inteligencija” nastao je 1956. godine za vrijeme trajanja kongresa *Dartmouth summer research project on artificial intelligence* kojeg je organizirao John McCarthy, a u projektu su još sudjelovali Marvin Minsky (tvorac prvog stroja koji je imao neuralnu mrežu, SNARC) i Claude Shannon, tvorac informacijskih znanosti i moderne logike u području računalne znanosti. Ishod tog istraživanja predstavio je McCarthy koji je izjavio “kako se u teorijskom smislu svaki oblik učenja i svaki drugi oblik inteligencije može u načelu toliko detaljno objasniti da ga stroj može simulirati, na temelju čega će se provoditi istraživanja o ideji strojnog korištenja logičkih sklopova za rješavanje problema za čije je rješavanje inače potrebna ljudska inteligencija, odnosno istraživanja o tome kako će strojevi sami sebe unaprijeđivati i usavršavati”.⁵ Međutim ta definicija i dalje počiva na ideji biološke ljudske inteligencije. Uz nju, predstavljeno i sedam postulata UI koji glase:

- automatizirana računala – računala koja simuliraju više kognitivne procese. Ovo je revolucionarna zamisao budući da su tadašnja računala bila ograničena samo na obavljanje jednostavnih izračuna. Koristila su bušene kartice i jednostavne tranzistorske logičke sklopove i nisu bila sposobna izvršavati napredne algoritme.
- programiranje računala da koristi prirodne jezike – odnosi se na pitanje kako omogućiti računalu da koristi i razumije prirodan jezik. Prirodni jezici predstavljaju velik problem za računalo zbog svojstva kontekstualizacije, a veliki problem predstavlja i svojstvo takvih jezika da se vremenom mijenjaju. Računala preferiraju jezike koji su standardizirani.
- neuralne mreže – kreiranje uvjeta u kojem bi umjetni neuroni bili aktivirani na način da mogu shvaćati koncepte. Elektronski neuroni predstavljaju prvotne zamisli o simulaciji ljudskog uma u računalnom svijetu. Nažalost, tadašnji koncept elektronskog neurona nije točno opisivao elemente neuralnih mreža koje se koriste danas, ali sam koncept neuralnih mreža je rezultirao današnjim mrežama koje su temelj strojnog učenja.
- teorija opsega izračuna – omogućiti računalu da problem koji obrađuje shvati na odgovarajući način. Kontekstualizacija problema koji je postavljen računalu predstavlja višu razinu logike i rasuđivanja. Konceptualno shvaćanje zahtijeva

⁵ McCarthy John; Minsky Marvin; Rochester Nathaniel; Shanon Claude. A proposal for the Dartmouth summer research project on artificial intelligence. 1955, str. 2.
URL: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> (2022-09-21)

napredne algoritme koji bi mogli percipirati značenje svakog elementa problema u kontekstu koji se nalazi, što znači da bi model UI bio sposoban problem percipirati kao jednu cjelinu, a ne kao matematičku operaciju.

- sažetost – mogućnost algoritma da obrađuje kompleksne probleme i riješi ih sažimanjem problema u manje fragmente. Takva sposobnost fragmentiranja problemskog zadatka donosi mogućnost ubrzavanja procesa rješavanja problema. Kada se složene probleme rastavi na njegove sastavne, manje jedinice računalo ih može uspješnije i brže razriješiti.
- samostalan razvoj – mogućnost sustava da na temelju prijašnjih zaključaka samostalno unaprijedi svoj algoritam. Slobodan razvoj algoritma predstavlja mogućnost modela UI da sam sebe korigira, stvori prečace u razmišljanju i općenito postaje sve bolji u razrješavanju problemskih zadataka koji su mu zadani.
- kreativnost i slučajnost – mogućnost sustava da kreativno stvara i dolazi do slučajnih zaključaka koji ga razvijaju. Kroz ovakve kognitivne procese UI model može doći do bolje spoznaje samog sebe i svojih kognitivnih procesa, ovakav pristup omogućava napredan razvoj ove tehnologije kroz njeno samootkrivanje.⁶

U godinama koje su uslijedile bili su stvoreni i korišteni brojni algoritmi koji su dokazali da je stvaranje inteligentnog stroja moguće, no zbog nedostataka komponenti za izradu strojeva koji bi izvršavali te algoritme, kao i zbog nemogućnosti praktične primjene teorijskih pretpostavki, doći će do zastoja u razvoju UI. Rani pokušaji stvaranja umjetne opće inteligencije (eng. artificial general intelligence, AGI) ukazali su na sve nedostatke trenutne tehnologije koja se koristila za razvoj komponenti. Algoritmi koji su simulirali umjetnu opću inteligenciju (tzv. jaku UI) bili su zamijenjeni algoritmima koji su namijenjeni određenom obliku UI, poznatom kao umjetna uska inteligencija ili tzv. slaba UI (eng. artificial narrow intelligence, ANI). Kako su se ti ekspertni sustavi pokazali učinkovitima u poslovanju, ponovno se pokrenula inicijativa za financiranje takvih i sličnih projekata. Rezultat toga bili su strojevi koji su pobjeđivali šahovske majstore, smanjivali rashode u tvrtkama i uspješno iskorištavali potencijal tadašnje računalne tehnologije.⁷

Ujedno jedan od prvih testova koji su stvoreni za određivanje i mjerenje strojne inteligencije povezuje se s Alanom Turingom i naziva se Turingovim testom. Sam test je

⁶ McCarthy John; Minsky Marvin; Rochester Nathaniel; Shanon Claude. Nav. dj. str. 2

⁷ McCorduck, Pamela. Nav. dj., str. 111-171.

koncipiran tako da se ispitivača suoči sa strojem, koji za svoj cilj ima reproducirati ponašanje koje bi zavaralo ispitivača da razgovara s čovjekom. Takvim se koncipiranjem testa za određivanje strojne inteligencije stroj 'ograničava' na biološku ljudsku inteligenciju što predstavlja ograničavajući čimbenik u razvoju UI zbog toga što neuroarhitektura stroja i čovjeka nisu jednake. Ograničavanjem na strojnu reprodukciju ljudskog ponašanja ograničava se razvoj naprednih algoritama i logike UI, kao i daljnji razvoj i istraživanje u području UI.⁸

Danas postoji nekoliko vrsta UI, od kojih svaka ima svoje određene karakteristike koje ju razlikuju od drugih. Dva su temeljna tipa UI, tip UI temeljen na sposobnostima i tip UI temeljen na funkcijama. S obzirom na sposobnosti, postoje već spomenuta umjetna uska inteligencija (tzv. slaba UI) koja obuhvaća strojeve koji mogu funkcionirati samo u određenim uvjetima, ekspertni algoritmi UI koji su specijalizirani samo za jedno određeno područje, također već spomenuta umjetna opća inteligencija (tzv. jaka UI) koja obuhvaća strojeve koji mogu funkcionirati u raznolikim uvjetima i sposobni su djelovati i donositi odluke u promjenjivom fizičkom i digitalnom okolišu te umjetna super inteligencija (eng. artificial super intelligence, ASI) koja obuhvaća strojeve koji nadilaze ljudsku inteligenciju i kognitivne sposobnosti, a smatra ju se logičnim nastavkom u razvoju umjetne opće inteligencije.

Podjela UI prema funkcionalnosti daje dublji uvid u raznolikost sustava UI, kao i bolji uvid u sposobnosti određenih generacija inteligentnih strojeva. Prema funkcionalnosti razlikuju se sljedeći tipovi UI: reaktivni strojevi koji predstavljaju prvu fazu u razvoju inteligentnih strojeva, a sposobni su reagirati na unos vanjskih podataka; strojevi ograničene memorije koji, uz mogućnost reagiranja na unos vanjskih podataka, imaju i mogućnost učenja iz prijašnjih iskustava; strojevi koji simuliraju biološku ljudsku inteligenciju odnosno kognitivne sposobnosti čovjeka (postoje zamisli da bi se ti strojevi mogli kombinirati s algoritmima koji bi omogućili prepoznavanje i razumijevanje ljudskih osjećaja); i, u konačnici, samosvjesni strojevi koji ne simuliraju kognitivne sposobnosti čovjeka, već su svjesni svog postojanja i sposobni samostalno poimati svijet oko sebe te ulaziti u interakciju s njim.

Trenutno, najčešće korišteni sustavi UI obuhvaćaju specijalizirane inteligentne strojeve koji su već neko vrijeme uključeni u industrijske i poslovne procese automatizirajući

⁸ Isto, str. 504-505.

ih i ubrzavajući ih na nižim razinama. Komercijalna upotreba takvih sustava sve više približava koncept složenih i učinkovitih algoritama široj javnosti koja je, između ostalog, navikla na UI kamere na pametnim telefonima, algoritme koji prepoznaju lice vlasnika telefona ili prometne kamere koje znaju gdje se na automobilu nalaze registarske pločice i vozač. Također, komercijalizacija tehnologije UI stvara veće mogućnosti financiranja takvih projekata. Međutim, šira implementacija tehnologije UI ima i određene izazove. Zbog široke implementacije pametnih sustava u brojne komercijalne uređaje i njihove učinkovitosti i isplativosti, oni postupno postaju dijelom i većih i kompleksnijih uređaja i sustava, primjerice autonomnih vozila. Iako opća implementacija pametnih sustava može donijeti velik napredak ljudskom društvu, ekonomiji i poslovanju, potrebno je oprezno pristupiti njihovom razvoju i implementaciji. Primjerice, autonomna vozila predstavljaju jako zanimljiv okvir u kojem se može razvijati odgovoran model UI jer je promet dio svakodnevnog života. Tako prometne nesreće predstavljaju poseban izazov za odgovoran razvoj UI, odnosno implementaciju etičnosti i moralnih načela u djelovanje i odlučivanje stroja. Veliki izazov predstavlja stvaranje algoritma koji bi pravilno reagirao u slučaju prometne nesreće i potencijalno donio odluku koja će imati najmanje negativnih posljedica po osobe i infrastrukturu uključene u situaciju. Daljnji razvoj UI i pametnih sustava temelji se na rješavanju takvih i slični izazova koji uključuju pitanje sigurnosti, privatnosti i drugih prava. Upravo u tom kontekstu odgovorni razvoj UI i regulacija pametnih sustava ima veliku ulogu.

2.2 Regulacija umjetne inteligencije

Računalni sustavi, računalni programi i algoritmi koriste podatke kako bi došli do zaključaka, iako novi modeli, poput Alpha Go, računalnog programa koji pokazuje napredne vještine planiranja i zaključivanja pri igranju tradicionalne kineske društvene igre Go. Alpha Go je model UI, nastao u tvrtki *DeepMind* kao generacijski nasljednik Alpha Zero, koji se sastoji od velikog broja neuralnih mreža koje u sinergiji donose odluke na temelju vjerojatnosti i prijašnjih poteza protivnika⁹. Takvi modeli pokazuju napredne oblike kognitivnih procesa. Taj i slični primjeri pokazuju da je UI, uz podatke, počela obrađivati i informacije, tj. da su inteligentni strojevi u mogućnosti uspješno obrađivati višu razinu u DIKW (eng. Data-Information-Knowledge-Wisdom) hijerarhiji. Takav napredak inteligentnih sustava otvara brojne prilike za daljnji razvoj i inovacije u području UI. No, razvoj i inovacije u području UI

⁹ AlphaGo. URL: <https://www.deepmind.com/research/highlighted-research/alphago> (2022-09-21)

i njezina šira implementacija ne samo u industrijske i poslovne, već i društvene sustave nužno otvara pitanje stvaranja pravnog okvira za njezinu odgovornu i sigurnu implementaciju, odnosno zahtijeva regulaciju razvoja i primjene inteligentnih sustava što se najčešće odražava u oblikovanju pravilnika i smjernica za odgovorni razvoj UI. Dakako, potreban je oprez pri stvaranju pravnog okvira za odgovoran razvoj UI jer presnažan utjecaj pravnog okvira može potencijalno usporiti tehnički razvoj i inovacije. Stoga je potrebno stvaranje “fleksibilnih regulativa koje su usmjerene na budućnost”¹⁰, što znači da nadzorni sustavi trebaju imati moderne i neutralne algoritme koji će omogućiti donošenje odluka na dobrobit svih dionika društva.

Glavno uporište odgovornog razvoja umjetne inteligencije je stvaranje inteligentnih tehnologija usmjerenih na čovjeka što podrazumijeva inkluzivnost, tj. uključivanje čovjeka u procese i sustave i razvoj sustava koji će se ravnopravno odnositi prema svim članovima zajednice koja koristi određenu inteligentnu tehnologiju. Nadalje velika važnost u razvoju pridaje se transparentnosti procesa i sustava. Potrebno je navesti još jedan od temeljnih načela, a to je tzv. *faircoding*, koncept koji zahtijeva od osobe koja dizajnira model UI da kod napravi na što neutralniji način, što znači da u procesu donošenja odluke model UI ne smije ni na koji način diskriminirati osobe, tj. korisnike. Nažalost, *faircoding* u praksi ima određene nedostatke zbog toga što je nemoguće postići apsolutnu neutralnost, objektivnost i nediskriminaciju. Zbog svoje sveobuhvatnosti inteligentan sustav ima mogućnost diskriminiranja određenih grupa ljudi na temelju njihovih karakteristika što od razvojnog tima traži da oprezno pristupe stvaranju modela UI. Moguća diskriminacija može se negativno odraziti na grupu korisnika te na taj način značajno usporiti proces implementiranja pametnog sustava u društvo. U ovom hipotetskom primjeru možemo uvidjeti važnost i vrijednost odgovornog razvoja UI. Također, potrebno je shvatiti da zbog svoje sveobuhvatnosti u implementaciji, odgovornost ne snosi isključivo razvojni tim, već svi uključeni u proces planiranja, dizajniranja, programiranja, testiranja, implementacije, praćenja i evaluacije inteligentnih rješenja i pametnih sustava. To uključuje i donositelje odluka, kao i tijela odgovorna za nadzor rada sustava koji kroz praćenje i procjenu učinkovitosti sustava na temelju prikupljenih podataka odlučuju o njegovu daljnjem razvoju.¹¹

¹⁰ Garcia del Blanco, Iban. Nav. dj. str 10.

¹¹ Dignum, Virginia. Artificial Intelligence: Foundations, Theory, and Algorithms, Umeå, Sweden, Springer, 2019, str 119

Potrebno je spomenuti i prijetnje koje UI može predstavljati za pojedinca i društvo, a koje su potaknule ideju odgovornog razvoja UI. Mogućnosti primjene sustava UI u vojnom sektoru postale su izglednije zahvaljujući implementaciji autonomnih sustava i robota u oružane snage zemalja diljem svijeta. Primjerice, bespilotne letjelice i kopnene platforme koje mogu nositi naoružanje postale su toliko napredne da njihova integracija s UI postaje sve stvarnija prijetnja za društvo. Nadalje, stvaranjem autonomnih vozila javila se potreba za dizajniranjem sustava koji će uspješno donositi odluke u prometu i samostalno upravljati vozilom. Kao primjer vrlo se često navodi tzv. moralna dilema autonomnih vozila u sklopu koje je potrebno odlučiti treba li žrtvovati vozača ili pješake. Radi se o složenim pravnim, etičkim i sigurnosnim pitanjima, no budući da je odgovorni razvoj UI još uvijek u svojim začetcima, trenutno ne postoje jasni i jednoznačni na navedena pitanja. Napredak i nove spoznaje u razvoju UI i pametnih (autonomnih) sustava čine ideju odgovornog razvoja UI ne samo aktualnom, nego i nužnom. Navedeni izazovi nisu jedini, tu je i vrlo često pitanje pristranosti sustava u procesu donošenja odluka poput, primjerice, donošenja odluka unutar socijalnog ili zdravstvenog sustava. Sustav na temelju određenih karakteristika osobe (korisnika sustava) može o njemu donijeti diskriminatornu odluku i na temelju nje djelovati. Današnji koncept odgovorne UI nadilazi arhaični koncept Asimova tri zakona robotike, a jedan od najočitijih primjera jest pristranost sustava koja se temelji na odabiru i obradi podataka, djelovanju algoritama koji donose odluke i ugrožavanju privatnosti koja proizlazi iz prikupljanja i obrade osobnih podataka i informacija. To su samo neki od izazova koji se nastoje riješiti odgovornim razvojem UI kako bi se osigurale pravednost i inkluzivnost u društvu.

Inteligentni strojevi i pametni sustavi pružaju brojne mogućnosti, no donose i određene rizike. Kroz razvoj i praktičnu primjenu takvih sustava uočeni su problemi koje mnogi razvojni timovi nisu mogli predvidjeti, a koji ponajviše proizlaze iz činjenice da sustavi nisu orijentirani na čovjeka, kao i da su usmjereni na tehničko usavršavanje i komercijalno iskorištavanje sustava. Većina se problema ne može riješiti u razvojnoj ili testnoj fazi, već je nužna implementacija sustava u društvo koja će kroz njegovo korištenje pomoći identificirati određene probleme i zaštititi krajnjeg korisnika.

Novi koncept koji pokušava riješiti problem zaštite krajnjeg korisnika i promovirati tzv. antropocentrični karakter inteligentnih strojeva i pametnih sustava naziva se odgovorni razvoj umjetne inteligencije. Koncept je opisala Virginia Dignum 2019. godine kao model utemeljen na odgovornosti prema čovjeku, odnosno na razvoju UI koji je utemeljen na

načelima i vrijednostima ljudskog društva.¹² Takav koncept razvoja UI stavlja čovjeka u središte i umanjuje potencijalne rizike. Stvaranjem transparentnih, čovjeku razumljivih i nepristranih sustava jamči se njihova šira implementacija. Odgovorni razvoj UI dobio je podršku od strane različitih tijela javne i državne uprave na području EU, poput Europskog parlamenta koji je donio plan razvoja pravilnika i smjernica za odgovorni razvoj UI. Općeniti smjer razvoja ogleda se u definiciji Virginie Dignum, no konkretan popis općenitih načela za odgovorni razvoj UI može se pronaći u mnogim smjernicama kako javnog, tako i privatnog sektora. Jedan takav primjer jesu smjernice navedene na mrežnim stranicama 2021 AI, a koje uključuju pet glavnih načela:

1. Prvo načelo podrazumijeva transparentnost i razumljivost modela UI, a cilj mu je osigurati dizajniranje modela UI koji je razumljiv kako razvojnim timovima, tako i ulagačima i krajnjim korisnicima.
2. Drugo načelo podrazumijeva umanjivanje štete koju može prouzrokovati kvar sustava ili anomalija koja se dogodi pri obradi podataka. Ovo načelo štiti krajnjeg korisnika od potencijalne diskriminacije ili druge štete koja može nastati uslijed takvog nedostatka.
3. Treće načelo definira razinu odgovornosti za sustav koji se implementira, što znači da se razvojni tim koji je sudjelovao u razvoju, testiranju i implementaciji modela drži odgovornim za štetu koja potencijalno može nastati.
4. Četvrto načelo podrazumijeva poštivanje privatnosti podataka i način kako se upravlja tim podacima. UI modeli koriste baze podataka za učenje i razvijanje kognitivnih procesa pa je takve podatke potrebno zaštititi. Nadalje, privatni podaci koje će obrađivati model moraju biti zaštićeni od svih zlonamjernih napada. Sustav mora biti dizajniran na način da podaci, odnosno korisnici ostaju zaštićeni čak i u slučaju uspješnog napada.
5. Peto načelo uzima u obzir trajnost i sveobuhvatnost sustava. Ovakvi sustavi trebaju biti trajni i imati mogućnost dodatnog razvoja nakon njihove implementacije. Uz mogućnost naknadne nadogradnje sustava, potrebno je i osigurati sveobuhvatnost

¹² Dignum, Virginia. *Artificial Intelligence: Foundations, Theory, and Algorithms*, Umeå, Sweden, Springer, 2019, str 119

sustava, tj. sustav mora biti implementiran na način da se uključi u što više aspekata društva.¹³

Načela odgovornog razvoja UI stvorena su kao odgovor na brojne izazove s kojima se suvremeno društvo susreće pri implementaciji tehnologije UI. Jedan od najčešćih problema predstavljaju zaključci koje je sustav donio, a koji su nerazumljivi čovjeku. radi se o situacijama u kojima jedan inteligentan stroj, unatoč kontroliranom unosu podataka i kontroliranoj bazi podataka iz koje može učiti, nakon određenog vremena počne donositi zaključke koji nisu smisleni i razumljivi čovjeku. Takvi zaključci moraju biti detaljno promotreni i analizirani kako bi se pronašao izvor problema i ispravila greška koja se dogodila u procesu zaključivanja. Najčešća pogreška odnosi se na činjenicu da je sustav 'prekršio' pravila algoritma kako bi ubrzao radni proces i umjesto korištenja svih podataka počeo pronalaziti uzroke u podacima i putem njih donositi zaključke. Takva se greška prilično teško može uočiti u početnoj fazi kada model počinje kreirati nove strategije učenja. Još jedan od čestih problema s tehnologijama UI jest pitanje pristranosti koje se odnosi na anomalije pri donošenju zaključaka uzrokovane ili namjernom integracijom načela pristranosti u model ili 'slučajnom' pristranosti proizašloj iz podataka na temelju kojih model uči. Pristranost modela umjetne inteligencije nastaje iz dva razloga - kognitivne pristranosti osobe koja razvija model i pristranosti koja nastaje zbog nepotpunih podataka koji se koriste za učenje modela. U prvom slučaju pristranost je najčešće integrirana u model UI zbog pristranosti u umu dizajnera modela, a ta pristranost može pomoći u procesuiranju informacija o okolini. Pristranost nastala zbog nepotpunih podataka je greška koja se događa u slučaju da baza podataka iz koje model uči ima nepotpune i nedovoljne podatke zbog čega model neće biti sposoban razviti punu sliku o fenomenu i/ili entitetu o kojem uči. I ovaj problem je teško rješiv i teško ga je uočiti.¹⁴

Uz navedene izazove u razvoju i primjeni UI tu je i problem namjernog korištenja UI na destruktivan način. Usavršavanjem tehnika ratovanja razvile su se zračne i kopnene platforme opremljene naprednim sensorima i teškim naoružanjem koje izvršitelju daju veliku odgovornost i moć. Trenutno čovjek još uvijek donosi odluke, no čak i uz djelovanje čovjeka postoje slučajevi u kojima je napad takvim oružjem uzrokovao velike civilne žrtve, odnosno kolateralnu štetu na okolnoj infrastrukturi. Bepilotne letjelice koje kruže iznad bojišnice na

¹³ 2021 AI. 2021.AI's Responsible AI Principles. URL: <https://2021.ai/responsible-ai-principles/> (2022-09-21)

¹⁴ Roselli, Drew; Jeanna Matthews; Nisha Talagala. "Managing bias in AI." // Companion Proceedings of The 2019 World Wide Web Conference. San Francisco, CA USA, May 2019. Str. 539-544. URL: <https://dl.acm.org/doi/pdf/10.1145/3308560.3317590> (2022-09-21)

nekoliko stotina metara donose prednost strani koja ih koristi, poput, primjerice, korištenja sustava senzora koji lako otkrivaju mete, međutim dopuštaju i mogućnost ispaljivanja dalekometnih navođenih projektila, tj. mogućnost ratovanja na daljinu. Slično je i sa borbenim vozilima opremljenima sensorima i oružjem na daljinsko upravljanje. Djelomična implementacija UI, pak, podrazumijeva korištenje inteligentnih strojeva i pametnih sustava na način da olakšavaju i pojednostavljaju rad čovjeka kao izvršitelja borbenih aktivnosti, ali ne donose samostalne odluke, tj. ne djeluju autonomno. Takav sinergizam čovjeka i stroja osigurao bi uspješnost uz zadržavanje glavne odgovornosti čovjeka. Izazovi nastupaju automatizacijom u kojoj se izostavlja mjesto/uloga čovjeka što predstavlja ozbiljnu ugrozu sigurnosti društva.

Kao što je vidljivo iz navedenih primjera, UI koja nije temeljena na odgovornom razvoju predstavlja veliki rizik i opasnost za društvo. Regulacija i odgovorni razvoj UI logičan su nastavak u evoluciji UI i važan zadatak razvojnih timova. Suradnja struke, znanstvene zajednice, donositelja odluka, javnog, državnog i privatnog sektora važan je čimbenik uspješnosti odgovornog razvoja UI i digitalne transformacije društva koje je i dalje orijentirano na čovjeka. Pravne smjernice za zaštitu krajnjih korisnika inteligentnih rješenja i pametnih sustava važan su dio procesa stvaranja povjerenja u tehnologiju UI. Sigurnost krajnjeg korisnika znači i sigurno društvo. Moderan, fleksibilan i odgovoran sustava dizajniranja inteligentnih strojeva i sustava nudi mogućnost stvaranja pravednog i inkluzivnog društva koje će omogućiti jednake prilike za sve dionike društva.

3. Pravilnici i smjernice za regulaciju umjetne inteligencije

3.1 Pravilnici i smjernice državnog (javnog) sektora

Ovo poglavlje donosi pregled odabranih smjernica i pravilnika kreiranih od strane državnih i javnih tijela na razini Europske Unije (EU), odnosno na razini Ujedinjenog Kraljevstva (UK) i Hrvatske dajući uvid u regulatorne mjere u pogledu razvoja UI na razini EU, UK i Hrvatske. Specifičnost EU pravilnika jest ta da predstavlja europsku strategiju razvoja tehnologija u području UI te predstavlja polazište za kreiranje pravilnika i smjernica koji se primjenjuju u europskim tvrtkama i međunarodnim tvrtkama na području EU koje razvijaju sustave UI.

Prvi pravilnik, tj. smjernice koje će se analizirati u ovom poglavlju jest Izvješće s preporukama Komisiji o stvaranju radno okvira za etički razvoj UI, robotike i srodnih

tehnologija (2020/2012(INL)). Izvješće je doneseno na razini Europske Unije, izvjestitelj je Ibán García del Blanco i sastoji se od prijedloga rezolucije, detaljnih priloga, mišljenja dvaju odbora i detaljnog izvještaja s glasanja o prihvatanju ili odbijanju rezolucije. Sam prijedlog usvojen je 2020. godine. Izvješće donosi pregled potencijalnih rješenja za odgovorni razvoj i uvođenje sustava UI kroz nekoliko aspekata i na temelju nekoliko drugih izvješća koja se referiraju na isto ovo izvješće. Izvješće, prije svega, razmatra kreiranje okvira za razvoj sustava koji bi bio nepristran te bi tako umanjio mogućnost za diskriminaciju ili isključivanje društvenih grupa na temelju njihovih karakteristika, odnosno razvoj sustava koji bi se prema svim svojim korisnicima odnosio jednako pridonoseći jednakosti i pravednosti društva. Drugi dio izvješća odnosi se na stvaranje fleksibilnog pravnog oblika za razvoj tehnologija u području UI koji ne bi ograničavao i/ili onemogućavao tehnološki razvoj, ali bi jačao i promovirao sigurnost krajnjeg korisnika. Treći dio izvješća usmjeren je na pitanje klasifikacije intelektualnih djela koje je stvorio model UI te kako kategorizirati ta djela. Uz ova tri dijela izvješća postoji i nekoliko drugih dijelova koji se odnose na pitanja vojne i pravosudne upotrebu sustava UI, no sustavi UI u trenutnoj razvojnoj fazi predstavljaju opasnost i čimbenik rizika u, primjerice, vojnim i policijskim operacijama zbog brojnih ograničenja u sferi programa i komponenti koje se koriste za provedbu takvih programskih rješenja. Trenutno ne postoji pravni okvir u kojem bi se ovakva tehnologija mogla primijeniti, no još veću prepreku predstavlja razina tehnologije kako bi se ovakvi sustavi uveli u vojni i policijski sektor. Moderni senzori koji se trenutno koriste za operacije tog tipa i dalje zahtijevaju ljudsku kontrolu kako bi se koristili na efektivan i pravedan način. UI se u ovom Izvješću klasificira kao sustav visokog rizika, a cijeli skup srodnih tehnologija se jednako kategorizira. Naime, UI, roboti i srodne tehnologije u području inteligentnih strojeva i pametnih sustava koje uključuju određena softvere, algoritme i podatke koji se koriste u radu i/ili su rezultat rada takvih tehnologije, predstavljaju značajan rizik kršenja etičkih principa koji su prikazani u ovim smjernicama te se smatraju rizičnim tehnologijama. Iz navedenog proizlazi i težnja za fleksibilnijim pristupom u radu sa spomenutim visokorizičnim tehnologijama, odnosno potrebno je vršiti nadzor i procjenu visokorizičnih tehnologija u odnosu na smjernice za njihov razvoj i primjenu propisane u Izvješću. Nacionalna nadzorna tijela zadužena za provjeru i procjenu tehnologija i sustava trebaju biti sastavljena od stručnjaka različitih profila i specijalizacije kako bi se osigurao interdisciplinarni pristup. Nadzorna tijela bila bi ovlaštena procijeniti rizike koje predstavlja određeni projekt, a

procjena tog rizika mora biti u skladu s standardiziranim i unificiranim pravilima za procjenu rizika.

Smjernice predstavljene od strane Europskog parlamenta su sveobuhvatnije i uključuju više aspekata odgovornog razvoja, ujedno stavljajući veći fokus na umanjivanje (negativnih) posljedica po krajnjeg korisnika, primjerice, u slučaju kvara sustava ili određene situacije u kojoj rad sustava ne slijedi načela odgovornog razvoja. Postoji osam glavnih načela odgovornog razvoja:

1. Prvo načelo odnosi se na antropocentrični karakter UI, tj. na činjenicu da je UI model stvorio čovjek. To načelo podrazumijeva središnju ulogu čovjeka koji bi u radu s tehnologijama i sustavima UI trebao upravljati i odlučivati o radu sustava. To može, primjerice, značiti da čovjek upravlja sustavom u svim fazama razvoja, ili da u određenoj fazi razvoja može ponovno definirati rad i potencijalno ispraviti greške i anomalije koje su se pojavile.
2. Drugo načelo odnosi se na proces definiranja rizika modela UI koji se razvija. Definiranje rizika provode nadležna tijela koja mogu u slučaju kvara ili anomalije sustava promijeniti radni okvir ili smjernice koje se odnose na određeni model UI.
3. Treće načelo odnosi se na aspekt sigurnosti, transparentnosti rada i pitanje odgovornosti u slučaju kvara sustava. Načelo pojašnjava sigurnosne sustave i mehanizme koji se pokreću u slučaju kvara, a obuhvaća i aspekt transparentnosti i objašnjivosti odluka koje je donio sustav.
4. Četvrto načelo odnosi se na osiguravanje inkluzivnosti, pravednosti i raznolikosti u odnosu na sve uključene u rad sustava UI. To znači da sustav ne smije negativno ili pozitivno diskriminirati nikoga niti ograničavati ičiju slobodu.
5. Peto načelo odnosi se na društvenu odgovornost razvojnih timova koji pri razvoju sustava UI moraju obratiti pozornost na potencijalne anomalije koje mogu biti uzrokovane pristranošću sustava. Vrijednosti kao što su dobrobit društva, tj. društvena odgovornost moraju biti implementirane kako bi se osigurali temelji za razvijanje ravnopravnog i mnaperdnog društva.
6. Šesto načelo odnosi se na primjenu modernih materijala i razvoj novih rješenja za očuvanje okoliša. Kako bi se provelo ovo načelo potrebno je implementirati ekološki

neutralne izvore energije i materijale koji uopće neće ili će minimalno utjecati na okoliš u koji je implementiran sustav.

7. Sedmo načelo odnosi se na pravedno korištenje biometrijskih sustava i na zaštitu privatnosti krajnjeg korisnika. Biometrijski sustav prikuplja podatke o biološkim karakteristikama korisnika i takvi podaci bi morali biti primarni fokus zaštite podataka u sustavu. Ti podaci su izrazito vrijedni, a osobi koja bi ih koristila na zlonamjeren način mogu omogućiti različite ovlasti na brojnim sustavima. Nadalje, potrebno je poštovati privatnost korisnika koji samostalno unose svoje biometrijske podatke u sustav, no potrebno je omogućiti i pristup tim ipodacima u slučaju potencijalne ugroze društvenog poretka.
8. Osmo načelo odnosi se na osiguravanje pravne zaštite svakoj osobi koja prijavi nadzornim tijelima sve prijestupe u implementaciji smjernica i pravilnika za odgovorni razvoj. Naime, kako bi se zaštitio cjelokupan proces odgovornog razvoja UI, bilo je potrebno uvesti pitanje pravne zaštite u smjernice i pravilnike za odgovorni razvoj UI.¹⁵

Stvaranje takvog antropocentričnog okvira temelj je daljnjeg razvoja sustava UI koji će se, zahvaljujući tomu, puno lakše integrirati u društvo. Predsjednik radne komisije kojoj je bila povjerena implementacija navedenih smjernica i okvira definirao je ciljeve odgovornog razvoja UI i kako bi oni trebali utjecati na društvo, “Europa mora razviti umjetnu inteligenciju koja je pouzdana, koja otklanja predrasude i diskriminaciju, te služi javnom dobru, omogućujući napredak poslovanja, industrije te stvarajući ekonomsko blagostanje”.¹⁶ Navedena izjava ukazuje na put razvoja inteligentnih sustava i primarni cilj njihove primjene na području EU, a koji naglašava važnost njihove integracije u poslovanje, industriju, javni i državni sektor i druge aspekte društvenog djelovanja.¹⁷ Važno je spomenuti da na području EU djeluje i Europsko udruženje za umjetnu inteligenciju (The European AI Alliance¹⁸) koje je posvećeno stvaranju otvorene politike u području UI i Strategije za umjetnu inteligenciju (AI Strategy¹⁹) kroz suradnju s građanima, tijelima civilnog društva, organizacijama u

¹⁵ Garcia del Blanco, Iban. Nav. dj.

¹⁶ ECOPNET. AI rules: what the European Parliament want!. 2021.

URL: <https://www.ecopnet.com/post/ai-rules-what-the-european-parliament-wants> (2022-09-21)

¹⁷ Garcia del Blanco, Iban. Nav. dj.

¹⁸ The European AI Alliance. URL: <https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance> (2022-09-01)

¹⁹ Communication Artificial Intelligence for Europe. URL:

<https://digital-strategy.ec.europa.eu/en/library/communication-artificial-intelligence-europe> (2022-09-21)

poslovnom i trgovačkom sektoru, akademskom zajednicom, tijelima javne uprave i stručnjacima različitog profila.

Kao primjer pravilnika i smjernica donesenih na razini države, dana su dva primjera – primjer Ujedinjenog Kraljevstva, posebice Irske, te primjer Hrvatske. Ujedinjeno Kraljevstvo svoj program za razvoj UI počinje razvijati u sklopu Vijeća za istraživanje umjetnosti i humanističkih znanosti (Arts and Humanities Research Council, AHRC). Cilj je ovih smjernica i razvojnih programa stvoriti povjerenje među građanima Ujedinjenog Kraljevstva u sustave UI. U fokusu smjernica nalaze se pitanja obrade velikih podataka (eng. Big Data) u svrhu poslovanja, biometrijska identifikacija i upotreba UI u medicinskoj dijagnostici.²⁰ Takva sveobuhvatna priprema i planiranje za implementaciju UI u različite sfere društva predstavlja primjer dobre i odgovorne prakse odgovornog razvoja i primjene UI.

Irska već ima razvijen program za odgovorni razvoj UI. Donesen je 2021. godine, a po svojim je karakteristikama jako sličan programu koji je donijela Europska Unija. Temelji se na odgovornosti i zaštiti krajnjeg korisnika i sigurnoj implementaciji sustava UI u društvo. Plan je stvoriti radni okvir u kojem razvoj UIu nije ograničen pravnim pitanjima, ali opet zahtjeva ispunjavanje standarda propisanih strategijom.²¹

Republika Hrvatska ne posjeduje službeni plan za odgovorni razvoj UI koji se trenutno provodi, no kroz Hrvatsku udruhu za umjetnu inteligenciju, CroAI²², stvorena je strategija razvoja koja je poslana u sjedište Europske Unije na evaluaciju i potencijalno prihvaćanje. Udruga CroAI osnovana je u prosincu 2019. godine, a djeluje kao član Vladine radne skupine za definiranje buduće nacionalne strategije razvoja UI u Republici Hrvatskoj u kojoj bi se naglašavao razvoj UI orijentirane prema čovjeku. Udruga okuplja tvrtke i startupove iz područja UI u Hrvatskoj te je usmjerena na dijalog između poduzetnika i donositelja uloga u Hrvatskoj i EU. Više o radu Udruge, inicijativama i aktivnostima moguće je saznati na mrežnim stranicama CroAI.

²⁰ UKRI. £8.5 million programme to transform AI ethics and regulation

URL: <https://www.ukri.org/news/8-5-million-programme-to-transform-ai-ethics-and-regulation/> (2022-09-21)

²¹ Vlada Irske. AI- here for good. 2021

URL: [national-ai-strategy.pdf](https://www.enterprise.gov.ie/national-ai-strategy.pdf) (enterprise.gov.ie) (2022-09-21)

²² Hrvatska udruha za umjetnu inteligenciju CroAI. URL: <https://www.croai.org/regulation?lang=hr> (2022-09-21)

3.2 Pravilnici i smjernice privatnog sektora

Privatni sektor predstavlja posebnu zajednicu u kojoj tvrtke i startupovi stvaraju strategije razvoja i primjene UI. Zbog specifičnosti poslovanja i potreba, svaka tvrtka razvija svoje smjernice i pravilnik koji će slijediti pri razvoju i implementaciji takvih sustava. Zbog trenutne situacije u kojoj se još uvijek ne regulira razvoj inteligentnih sustava na sveobuhvatan i dosljedan način, postoje značajna odstupanja u pravilnicima i smjernicama različitih tvrtki, kao i odstupanja od europske strategije za odgovorni razvoj UI. No zbog sve veće aktualnosti pitanja odgovornog razvoja UI, brojne tvrtke koje su predvodnici u području UI počele su stvarati profesionalna udruženja koja svojim djelovanjem i aktivnostima nastoje doprinijeti odgovornom razvoju UI. Jedno od takvih udruženja je i *Partnership of AI (PAI)* među čijim su članovima Apple, Meta, Samsung, Sony, Intel.²³ Udruženje je orijentirano na pitanje utjecaja UI na ekonomiju i tržište rada, na načela pravednosti, transparentnosti i odgovornosti u području razvoja i primjene UI te na pitanje sigurnosti u društvu u kojem sustavi UI i slična inteligentna rješenja postaju čovjekovom svakodnevicom. Cilj je udruženja poticati razvoj UI koja podržava društvo jednakosti i obilja orijentirano prema čovjeku.

Kao konkretne primjene tvrtki koje su aktivne u području odgovornog razvoja UI, potrebno je navesti tvrtku Google koja kao jedan od predvodnika u polju ipak nema razrađenu cjelovitu strategiju odgovornog razvoja UI. Strategija uključuje tek nekoliko preporučenih praksi u području razvoja i primjene UI, a koje se uglavnom temelje na provođenju testiranja. Ne postoji niti jedna stavka u Googleovoj razvojnoj strategiji koja problematizira nužnost promjene pristupa razvoju UI, nužnost razvoja smjernica za primjenu faircoding metode ili metode za uklanjanje pristranosti sustava. Gotove sve preporuke i smjernice podređene su ideji osiguravanja tvrtkine vodeće pozicije u industriji.²⁴

Microsoft kao još jedna od vodećih tvrtki u svijetu razvoja računalnih rješenja, također nema striktno definirani radni okvir za odgovorni razvoj UI. Dostupni su samo informativni članci i video priloge o tome kako takav radni okvir primijeniti u razvoju UI. U fokusu članaka su tehnička rješenja kako uspješno testirati modele i sustave UI i kako riješiti potencijalne anomalije koje se u njima pojavljuju, poput pristranosti u kodu. Nažalost, nisu

²³ Partnership of AI

URL: [https://partnershiponai.org/partners/ \(2022-09-21\)](https://partnershiponai.org/partners/ (2022-09-21))

²⁴ Google, Responsible AI design

URL: [https://ai.google/responsibilities/responsible-ai-practices/ \(2022-09-21\)](https://ai.google/responsibilities/responsible-ai-practices/ (2022-09-21))

dostupne informacije o tome koga se štiti takvim kodiranjem i kakav utjecaj na zajednicu ima promjena okvira u kojem se takve tehnologije razvijaju.²⁵

Intel je još jedna tvrtka koja je fokusirana na razvoj UI u sklopu vlastitih proizvodnih programima. Njihovi ciljevi su fokusirani izravno na programe i sustave UI koji poboljšavaju čovjekovo zdravlje i osiguravaju njegovu. Programi poput inteligentnih invalidskih kolica ili modela UI koji su namijenjeni brzom dijagnostičiranju dijetetske retinopatije²⁶ imaju mogućnost snažnog utjecaja na ljudsko društvo.²⁷

Potrebno je naglasiti da velik broj radnih okvira i smjernica za odgovorni razvoj UI u privatnom sektoru djelomično ili u potpunosti slijedi načela europske strategije odgovornog razvoja UI. Iako navedene smjernice i okviri s pravne strane zadovoljava potrebu za njihovim donošenjem i primjenom, njihova nedovoljna sveobuhvatnost i razrada ostavljaju dojam da su upravo kreirani samo u svrhu zadovoljavanja pravnog okvira. Kako bi se osiguralo njihova stvarna primjena, bit će potrebno revidirati i prilagoditi navedene pravne okvire kako bi se osiguralo stvarno ispunjavanje sigurnosnih i razvojnih zahtjeva u području odgovornog razvoja UI koje je EU postavila. Temelji odgovornog razvoja UI trebaju biti uniformirano implementirani i provedeni pravni okviri i regulativni mehanizmi koji štite krajnjeg korisnika i socio-ekonomsku strukturu na kojoj počiva današnje društvo.

4. Zaključak

Pitanje moralnog djelovanja i etike problematiziralo se još od samih početaka UI, kako kroz popularna djela znanstvene fantastike, tako i kroz teoriju i praksu discipline UI. Veliki broj znanstvenika, inženjera i drugih stručnjaka iz područja UI smatra kako prava moralnost ne postoji u strojevima. Pokušaj definiranja etičnosti i moralnosti u sferi UI najbliži je funkcionalističkom pristupu prema kojemu stroj ne donosi (ispravne) odluke i ne djeluje svjesno, već ih donosi i djeluje jer je tako programiran. Povezivanje hladne racionalne logike, razboritosti stroja, odgovornosti, sigurnosti, moralnosti i etičnosti predstavlja veliki i dugotrajan izazov za disciplinu UI. Rezultat toga su brojne rasprave o odgovornom razvoju UI koji bi trebao dalje poticati razvoj UI i robotike. Zbog toga je potrebno imati čvrste i jasno definirane pravilnike i smjernice, odnosno pravni okvir za odgovoran razvoj UI kako bi se

²⁵ Microsoft. Responsible AI principles

URL: <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6> (2022-09-21)

²⁶ Bolest, tj. oštećenje retine koje nastaje uslijed kroničnih komplikacija šećerne bolesti.

²⁷ Intel. AI for good.

URL: <https://www.intel.com/content/www/us/en/artificial-intelligence/ai4socialgood.html> (2022-09-21)

rizici za krajnjeg korisnika i društvo općenito sveli na minimum ili čak i izbjegli. Naravno, u tom kontekstu potrebno je kontinuirano pratiti i evaluirati aktivnosti i projekte u području UI kako bi se utvrdilo pridržavaju li se propisanih smjernica za odgovorni razvoj. Važno je i stvoriti jasna načela i kriterije za stvaranje, razvoj i implementaciju sustava UI, istovremeno izbjegavajući rizik usporavanja i ograničavanja inovacije i tehnološkog napretka. Takav pristup dizajniranju, stvaranju i implementaciji inteligentnih sustava omogućuje siguran i pozitivan znanstveni i tehnološki razvoj industrije i društva što je važno ne samo u kontekstu poslovanja, gospodarstva i komercijalnog iskorištavanja inteligentnih sustava i strojeva, nego i u svrhu njihova integriranja u javni i državni sektor, odnosno šire društvo. Kada je u pitanju pravni okvir za odgovorni razvoj UI na području EU, razlikuju se pravni okvir javnog i državnog sektora, od kojih je trenutno najpoznatiji i najaktualniji onaj Europske komisije i pojedinih nacionalnih vlada, te pravni okvir privatnog sektora za kojeg je karakteristično da se uglavnom temelji na pravnom okviru javnog i državnog sektora, ali problematici odgovornog razvoja UI ne pristupa tako sveobuhvatno. Dijelom je razlog tomu komercijalni interes privatnog sektora i osiguravanje konkurentnosti u industriji. Dvojnost pravnog i regulatornog sustava te njegova sporost uspješno su riješeni primjenom modela procjene rizika pri procjenjivanju budućih modela i sustava UI i inteligentnih rješenja. Fleksibilnost i uniformnost sustava procjene i reguliranja inteligentnih sustava omogućuju neometan razvoj budućih tehnologija i inovacije u polju, a uz to pružaju ujednačeni etički model za razvoj UI koja ima veliku mogućnost integracije. Rješavanjem pitanja privatnosti i sigurnosti krajnjeg korisnika koji ulazi u interakciju s takvim sustavima i rješenjima, stvara osjećaj povjerenja i doprinosi uspješnijoj interakciji i učinkovitijem iskorištavanju inteligentnih sustava. Tijela koja će donositi odluke i kreirati pravni okvir za odgovorni razvoj UI trebaju biti utemeljena na stručnosti, ali i interdisciplinarnosti upravo zbog interdisciplinarnog karaktera znanstvene discipline UI. Pristup razvoju i primjeni UI usmjeren prema stvaranju otvorenijeg, ravnopravnijeg i inkluzivnijeg društva važan je aspekt odgovornog razvoja UI, ali i osiguravanja društvenog razvoja i napretka. Kako UI postaje sveprisutna tehnologija u čovjekovom životu, potrebno je odbaciti arhaični dualistički pristup koji se svodi na utopijsku i distopijsku percepciju i razumijevanje UI. Pojednostavljeno dualističko poimanje neće omogućiti ostvarivanje punog potencijala ove tehnologije. Potrebno je sagledati kako dobiti tako i rizike razvoja i primjene UI te na taj način osigurati odgovoran, učinkovit i siguran razvoj UI orijentiran prema čovjeku i dobiti društva.

5. Literatura

1. 2021 AI. 2021.AI's Responsible AI Principles. URL: <https://2021.ai/responsible-ai-principles/> (2022-09-21)
2. AlphaGo. URL: <https://www.deepmind.com/research/highlighted-research/alphago> (2022-09-21)
3. Bostrom, Nick. Superintelligence: Paths, Dangers, Strategies. Oxford: University Press, 2014.
4. Communication Artificial Intelligence for Europe. URL: <https://digital-strategy.ec.europa.eu/en/library/communication-artificial-intelligence-europe> (2022-09-21)
5. Dignum, Virginia. Artificial Intelligence: Foundations, Theory, and Algorithms. Umeå: Springer, 2019.
6. ECOPNET. AI Rules: what the European Parliament wants!. 2021. URL: <https://www.ecopnet.com/post/ai-rules-what-the-european-parliament-wants> (2022-09-21)
7. Floridi, Luciano. The 4th Revolution: How the Infosphere is Reshaping Human Reality. Oxford: University Press, 2014.
8. Garcia del Blanco, Iban. REPORT with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies. 8.10.2022. URL: https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_EN.html (2022-09-21)
9. Google, Responsible AI design. URL: <https://ai.google/responsibilities/responsible-ai-practices/> (2022-09-21)
10. Hrvatska udruga za umjetnu inteligenciju CroAI. URL: <https://www.croai.org/regulation?lang=hr> (2022-09-21)
11. Intel. AI for good. URL: <https://www.intel.com/content/www/us/en/artificial-intelligence/ai4socialgood.html> (2022-09-21)

12. McCarthy, John; Minsky, Marvin; Rochester, Nathaniel; Shanon, Claude. A proposal for the Dartmouth summer research project on artificial intelligence. 1955.
URL: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> (2022-09-21)
13. McCorduck, Pamela. Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence. Massachusetts: A K Peters/CRC Press, 2004.
14. Microsoft. Responsible AI principles. URL:
<https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6>
(2022-09-21)
15. Partnership of AI. URL: <https://partnershiponai.org/partners/> (2022-09-21)
16. Roselli, Drew; Jeanna Matthews; Nisha Talagala. Managing bias in AI. // Companion Proceedings of The 2019 World Wide Web Conference. San Francisco, CA USA, May 2019. Str. 539-544. URL: <https://dl.acm.org/doi/pdf/10.1145/3308560.3317590>
(2022-09-21)
17. UK Research and Innovation. £8.5 million programme to transform AI ethics and regulation, 15.6.2022. URL:<https://www.ukri.org/news/8-5-million-programme-to-transform-ai-ethics-and-regulation/> (2022-09-21)
18. HM Government. National AI Strategy: Command Paper 525. London: Office for Artificial Intelligence, 2021. URL: [national-ai-strategy.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/97222/national-ai-strategy.pdf) (enterprise.gov.ie) (2022-09-21)