

Marina Š. Budić

ETIČKE DILEME I STAVOVI PREMA PRIMENI VEŠTAČKE INTELIGENCIJE¹

SAŽETAK

Prvi deo rada sastoji se od teorijskog proučavanja etičkih izazova koji nastaju usled razvoja i primene veštačke inteligencije. Drugi deo rada čini empirijsko istraživanje stavova javnosti prema različitim aspektima upotrebe veštačke inteligencije. Učesnici u istraživanju bili su građani Republike Srbije (N=737) koji su dobrovoljno popunili onlajn upitnik. Upitnik se sastojao od opštih sociodemografskih pitanja, pitanja o upoznatosti sa veštačkom inteligencijom, kratkog uvoda u pojam veštačke inteligencije, i pitanja o stavovima učesnika prema primeni veštačke inteligencije. Rezultati su pokazali da na ove stavove utiču faktori poput godine ispitanika. Takođe, rezultati govore da javnost ispoljava zabrinutost u pogledu nestanka profesija usled razvoja veštačke inteligencije, kao i diskriminacije od strane veštački inteligentnih sistema. Kada je reč o poverenju građana da institucije da koriste veštačku inteligenciju u najboljem interesu javnosti, pokazalo se da javnost u Republici Srbiji najmanje poverenja ima u Vladu RS, a potom u MUP RS i Ministarstvo pravde RS, dok najviše veruje istraživačima sa Univerziteta i međunarodnim istraživačkim organizacijama. Rezultati ukazuju da je potrebno fokusirati se na obrazovanje javnosti o potencijalnim izazovima i načinima da se oni spreče. Pored toga, ovi rezultati informišu kako akademsku i širu javnost, tako i donosiocima odluka, čime se može ostvariti dijalog između navedenih strana. Tvrdim da razmatranje različitih aspekata stavova javnosti prema veštačkoj inteligenciji unapređuje ovu debatu. Stavovi filozofa, pravnika i naučnika koji se bave ovom temom prisutni su u debati koja je u toku, ali ne i stav javnosti, u čemu se ogleda potreba za ovom vrstom istraživanja.

KLJUČNE REČI

etika, veštačka inteligencija, stavovi javnosti, obrazovanje, privatnost, diskriminacija, transparentnost, demokratske vrednosti

¹ Ovaj rad je nastao u okviru projekta Etika i VI: Etika i stavovi javnosti o upotrebi veštačke inteligencije, u okviru programa Srbija i globalni izazovi: Ka pravednijim i demokratskim javnim politikama, u organizaciji Fondacije za otvoreno društvo Srbija u partnerstvu sa Institutom za filozofiju i društvenu teoriju. Projekat je sadržao teorijski i praktični deo. Prvi, teorijski deo projekta, bavio se ispitivanjem etičkih dilema i izazova povezanih sa primenom veštačke inteligencije. Drugi, praktični deo projekta, činilo je empirijsko proučavanje stavova javnosti prema različitim aspektima upotrebe veštačke inteligencije. Projekat se bavio istraživanjem etičkih pitanja, promovisanjem učešća građana u odlukama od javnog značaja, u cilju podržanja demokratskog društva u

Uvod

S obzirom na vrlo brz razvoj društva i tehnologije, diskurs o etičkim pitanjima postaje sve relevantniji. Sa jedne strane, postoji diskusija u naučnoj zajednici o primeni veštačke inteligencije, dok sa druge zjapi praznina kada je reč o stavovima javnosti o ovoj temi. Kako bi se veštačka inteligencija implementirala na što uspešniji način, pored relevantnih stavova istraživača, etičara i pravnika, neophodno je uzeti u obzir i znanje i stavove ljudi na tu temu. U Republici Srbiji istraživanja ove vrste nisu sprovedena, te je cilj ovog rada i empirijskog istraživanja da pomenutu prazninu popuni. Očekivani uticaj na naučnu i akademsku zajednicu, na donosiocel politika i na društvo u celini, jeste doprinos boljem razumevanju etičkih pitanja koja postavlja razvoj tehnologije, digitalizacije i veštačke inteligencije, kao i podizanje svesti o znanju i stavovima građana u Republici Srbiji o upotrebi veštačke inteligencije.

Pre nego što pređemo na samo istraživanje etičkih pitanja koja se javljaju u vezi sa veštačkom inteligencijom i na empirijsko istraživanje stavova javnosti, potrebno je ukratko razmotriti sam pojam veštačke inteligencije.

Veštačka inteligencija (VI)

Postoje mnoga određena veštačke inteligencije (VI). Literatura je ponudila različite definicije veštačke inteligencije, od kojih svaka obuhvata ključne koncepte ne-ljudske inteligencije programirane za obavljanje specifičnih zadataka (Dwivedi et al. 2021). Rasel i Norvig (Russell et al. 2010) definisali su termin VI da opišu sisteme koji oponašaju kognitivne funkcije generalno povezane sa ljudskim atributima kao što su učenje, govor i rešavanje problema. (Dwivedi et al., 2021) Detaljnija karakterizacija opisuje veštačku inteligenciju u kontekstu njene sposobnosti da nezavisno tumači i uči iz spoljnih podataka kako bi se postigli specifični ishodi putem fleksibilnog prilagođavanja (Kaplan & Haenlein 2019). Zajednička nit ovih definicija je sve veća sposobnost mašina da obavljaju specifične uloge i zadatke koje trenutno obavljaju ljudi na radnom mestu i u društvu uopšte (Dwivedi et al. 2021).

Jedno od široko prihvaćenih određenja pojma VI pružila je i Evropska komisija u izveštaju o definisanju veštačke inteligencije²: „Veštačka

Republici Srbiji. Razvoj i primena veštačke inteligencije jeste jedan od globalnih i nacionalnih izazova sa kojima se država suočava. Projekat je imao za cilj podizanje svesti o znanju i stavovima javnosti u Republici Srbiji, i kao takav, bio je usmeren na unapređenje znanja o izazovima i strahovima u zajednici koje donosi upotreba veštačke inteligencije.

2 A definition of AI: Main capabilities and scientific disciplines, Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission, 2018.

inteligencija (VI) odnosi se na sisteme koji pokazuju razumno, inteligentno, ponašanje na osnovu analize svog okruženja i donose odluke – sa određenim stepenom autonomije – da ostvare konkretne ciljeve. Sistemi zasnovani na veštačkoj inteligenciji mogu biti bazirani isključivo na softveru i delovati u virtuelnom svetu (na primer, virtuelni asistenti, softveri za analizu fotografija, internet pretraživači, sistemi za prepoznavanje govora i lica) ili mogu biti ugrađeni u uređaje – hardver (na primer, napredni roboti, autonomna vozila, dronovi i slično).“

VI se odnosi na računarske sisteme koji izvršavaju zadatke, donose odluke ili ostvaruju ciljeve u kompleksnim situacijama, bez eksplicitnih instrukcija ljudi. Termin veštačka inteligencija odnosi se na tehnologiju u kojoj su softver i/ili mašine u stanju da oponašaju (određene aspekte) ljudske inteligencije³ (Sindermann et al. 2021).

Primeri upotrebe VI jesu: roboti, industrijski koji se koriste u proizvodnji, humanoidni koji interreaguju sa ljudima, autonomna vozila, virtuelni asistenti i agenti za rezervaciju putovanja, alati za obradu prirodnog jezika, prevodenje, klasifikaciju slika i identifikaciju lica, dronovi kojima nije potreban kontrolor, vojni dronovi. U medicinskom sektoru koriste se roboti za asistenciju starim ljudima i hirurzima, kao recepcionisti, za dijagnostiku i mapiranje bolesti.

Primena veštačke inteligencije u oblasti računarske nauke ima za cilj da imitira kognitivne funkcije ljudi pomoću kompjuterskih algoritama (Jiang et al. 2017; Lawrence et al. 2016). Postoji i shvatanje VI kao institucionalnog hibrida, odnosno pojma koji je prilično problematičan i koji treba da se reinterpreтира u skladu sa potrebama različitih aktera (pravnik razume veštačku inteligenciju na način različit od kreatora politike, na način drugačiji od novinara, različit od stručnjaka za bioinformatiku, drugačiji od ljubitelja naučne fantastike, itd.) (Dwivedi et al. 2021).

Još u vreme industrijske revolucije, značajan razvoj tehničkih inovacija uspeo je da transformiše brojne manuelne zadatke i procese koji su postojali decenijama gde su ljudi dostigli granice fizičkog kapaciteta. Veštačka inteligencija nudi isti transformativni potencijal za povećanje i potencijalnu zamenu ljudskih zadataka i aktivnosti u okviru širokog spektra industrijskih, intelektualnih i društvenih primena. Tempo promena za ovo novo tehnološko doba veštačke inteligencije je zapanjujući, sa novim otkrićima u algoritamskom mašinskom učenju i autonomnom donošenju odluka, stvarajući nove mogućnosti za nastavak inovacija. Uticaj veštačke inteligencije mogao bi biti značajan, sa industrijama koje se kreću od: finansija, zdravstva,

3 Za detaljniju diskusiju o ovome videti Fetzer 1990.

proizvodnje, maloprodaje, lanca snabdevanja, logistike i komunalnih usluga, a sve to potencijalno je poremećeno pojavom VI tehnologija. (Dwivedi et al 2021)

VI tehnologija više nije oblast futurologa, već integralna komponenta poslovnog modela mnogih organizacija i ključni strateški element u planovima za mnoge sektore poslovanja, medicine i vlada na globalnom nivou (Dwivedi et al. 2021). Kao rezultat velikog napretka, sve veći broj poslova autonomno obavljaju VI sistemi bez ljudske kontrole i nadzora (Złotowski et al. 2017). VI prožima svaki aspekt društva, i zastupljena je u svim sferama našeg života. Koristi se u različitim sektorima – zdravstvu, ekonomiji, obrazovanju i učenju, bezbednosti, i drugim sferama života. Napredak u oblasti VI utiče na sve aspekte društva, uključujući tržište rada, transport, zdravstvenu zaštitu, obrazovanje i nacionalnu bezbednost. Ona drastično može ubrzati proste administrativne zadatke i pružiti najnaprednije sisteme analize koji su izvan ljudskih mogućnosti. VI se primenjuje u borbi protiv kriminala, terorizma, u reagovanju na prirodne katastrofe, u regulisanju saobraćaja i kontroli mase. Drastično ubrzava proste administrativne zadatke i pruža napredne sisteme analize, što podrazumeva uštedu vremena i sredstava iz budžeta.

Izazovi

VI donosi nove mogućnosti, ali i izazove.

Jedan od glavnih izazova u vezi sa usvajanjem i implementacijom VI su trenutne konotacije i percepcije predmeta. Većini ljudi VI je misteriozan koncept koji ne samo da je teško definisati, već je i teško razumeti kako se ona manifestuje u njihovom svakodnevnom životu. Iako očigledno postoje brojne pozitivne upotrebe veštačke inteligencije, mnogi ljudi je povezuju sa negativnim štampama i medijskim kampanjama koje prikazuju veštačku inteligenciju kao uzrok svega, od masovne nezaposlenosti do podataka, uklanjanja slobode, pa čak i potpunog globalnog ratovanja. Nedefinisan koncept i loša medijska pokrivenost dali su veštačkoj inteligenciji negativnu sliku. (Dwivedi et al. 2021)

Ono što se takođe obično predstavlja kao izazov u vezi sa veštačkom inteligencijom je što ona čini da se naizgled odstupa od problema iz stvarnog života, kao i dovođenje u pitanje trenutne mogućnosti veštačke inteligencije. To je otuda što je sam termin VI u isto vreme loše definisan, ali i višeznačan, u smislu izazivanja nade i strahova u rasponu od volje za njenim razumevanjem inteligencije do konzumacije od strane nečijeg potomstva

(Szollosy, 2017). Ovo nas dovodi do izazova VI koji se odnosi na problematične aspekte termina. Kao što se tvrdi (Galanos 2018), oba termina, „veštački“ i „inteligencija“, ako se razmatraju odvojeno, osporavani su zbog svoje rigidnosti i korisnosti. Razlikovanje između prirodnog i veštačkog (ili prirode/nege, prirode/kulture i drugih sličnih podela) nemoguće je definisati, jer je ili sva stvarnost prirodna (rezultat iste prirode; prirodnjačka perspektiva) ili je sve konstruisano (tumačeno unutar ljudskog mozga, labava konstruktivistička perspektiva). Takve razlike obično se postavljaju sa svesnom ili nesvesnom namerom da se ili „prirodizuje“ (dakle opravda) određeno ponašanje („ovo je neprirodno“) ili implicira ljudska nadmoć nad prirodom. Oba slučaja povezana su sa scenarijima „više od čoveka“ ili „manje od čoveka“, koji se nalaze u horor pričama, kao i u naučnoj fantastici. Dakle: „veštačko“ u VI je samo po sebi terminološki izazov (Dwivedi et al. 2021). Slično tome, nekoliko kognitivnih naučnika, kibernetičara, teoretičara sistema, stručnjaka za veštačku inteligenciju/robotiku i sociologa tvrde da je veštačka inteligencija jednostavno neodrediva jer nemamo jasno razumevanje reči inteligencija (Dwivedi et al. 2021).

Osim problema definisanja i razumevanja pojma, postoje važniji problemi sa kojima se suočava društvo kada je o upotrebi veštačke inteligencije reč. Iako VI poboljšava i olakšava naše živote, njen brz razvoj i upotreba otvaraju pitanja u oblasti bezbednosti i etike. Razvoj i primena veštačke inteligencije otvaraju brojna etička pitanja, među kojima su i pitanje interakcije veštačke inteligencije sa ljudskim pravima, poput prava na privatnost i problema diskriminacije (Barabas et al. 2017; Chancellor et al. 2019). Etički izazov predstavlja poverenje u VI, poverenje u one koji njome upravljaju, odgovornost za odluke i postupke VI. Pojedinci i organizacije mogu pokazati nedostatak poverenja i zabrinutosti u vezi sa etičkim dimenzijama VI sistema i njihovom upotrebom zajedničkih podataka (Sun & Medaglia 2019).

Jedan od izazova su i društveni i organizacioni stavovi prema tehnologiji (Vasiljeva et al. 2021). Dalje, VI može narušiti privatnost i slobodu građana, naročito kada je u pitanju javna bezbednost. Organizacije poput policije i avio-kompanija mogu prikupljati osetljive podatke građana. Građani svakodnevno ustupaju kontrolu nad privatnim informacijama. Vlade mogu koristiti te informacije u najrazličitije svrhe. Kada je reč o privatnosti, postoji velika i opravdana zabrinutost u pogledu ovog izazova. VI prikuplja ogromne količine informacija, uključujući podatke sa mobilnih uređaja i druge elektronike, i ekstrapolira ih, tako da profesionalci mogu donositi odluke zasnovane na tim podacima. Ugrožavanje privatnosti predstavlja, dakle, prvi izazov. Povećana upotreba automatizovanih sistema i veštačke inteligencije izlaže klijente riziku gubitka privatnosti. Kompanije su u mogućnosti da uporede kupovno ponašanje svojih kupaca, zbog čega su

kupci u opasnosti da pronađu ove informacije u javnom domenu ili da budu „prodati“ drugim zainteresovanim stranama. Takođe, prodavci mogu otkriti da su njihovi prethodno personalizovani podaci koji čine osnovu njihovih odnosa sa kupcima sada lako dostupni drugim delovima kompanije, smanjujući njihovu efikasnost (Dwivedi et al. 2021). Pored toga, postoji promena prirode interakcije prodavaca sa kupcima: opasnost od sve veće upotrebe veštačke inteligencije omogućava klijentima da sami izaberu svoje prodajne ugovore, uključujući cene, popuste i posebne karakteristike, uklanjajući potrebu za ljudskom interakcijom. Povećana upotreba veštačke inteligencije eliminiše ljudski dodir i emocionalnu povezanost koju su kupci razvili sa prodavcem. Ovaj gubitak autentičnosti u procesu prodaje može dovesti do toga da kupci nisu sigurni kome da veruju (Hurlei 2017). Takođe, zasnivanje predviđanja prodaje na osnovu prošlih ponašanja i aktivnosti na mreži može produžiti pristrasnost, npr. predstavljaju prošlu ili slučajnu zabrinutost koja više nije relevantna za trenutne potrebe kupca za kupovinu. Iako se putem VI mogu stvoriti novi tipovi uvida o kupcima, neki od njih mogu biti previše detaljni ili pogrešno protumačeni nauštrb profitabilnosti. Konačno, postoji opasnost od gubitka kontrole nad procesom prodaje, jer integrisani, automatizovani sistemi daju prioritet mejlovima, prate nove kontakte i kreiraju sastanke ili dnevni red kojih prodavac nije svestan (Dwivedi et al. 2021).

VI može negativno uticati na ljudska prava, poput prava na privatnost i na ravnopravnost. Privatnost je osnovno ljudsko pravo, neophodno za dostojanstven i bezbedan život. U digitalnom okruženju, kada koristimo aplikacije i platforme društvenih medija, prikupljaju se velike količine ličnih podataka – sa ili bez našeg znanja – i mogu se koristiti za profilisanje i predviđanje našeg ponašanja. Kada je reč o ravnopravnosti, ukoliko se u dizajn sistema VI uključe predrasude, on će ispoljavati iste predrasude i diskriminaciju kao i ljudi. Sistemi zasnovani na veštačkoj inteligenciji mogu da ispoljavaju različite nivoe diskriminacije (Bostrom & Yudkowsky 2019). Takođe, VI može dovesti do diskriminatornih ishoda kada se algoritmi primenjuju na odluke o zapošljavanju i otpuštanju, posebno ako su podaci o obuci zasnovani na homogenoj grupi. Neki VI programi za zaštitu od krađe koji koriste prepoznavanje lica mogu da diskriminišu kupce određenih rasa ili etničkih grupa. Brojni nedavni radovi pokazali su (a) međusobnu povezanost i istorijsku povezanost između automatizacije, podataka i VI tehnologija, i (b) kako jedan od glavnih problema koji se pojavljuju ima veze sa stvaranjem novih nejednakosti i održavanjem starije predrasude koja se odnosi na sve probleme interseksionalnosti sa rasom, rodom, seksualnom orijentacijom, vrstom, godinama i drugim problematičnim ljudskim kategorijama (npr. Eubanks 2017; Prainsack 2019).

Jedan od izazova predstavlja i zabrinutost u pogledu radnih mesta. VI može preuzeti neke od manuelnih i lakših zadataka, a čak i složenije zadatke, poput vožnje. VI uzrokuje gubitke poslova: automatizacija i automatizovane usluge zamenjuju ljude u mnogim industrijskim i uslužnim sektorima, kako bi smanjili troškove i poboljšali efikasnost. Efikasnost automatizacije u prodaji već je dokazana u situacijama teleprodaje i predviđanja, što je rezultiralo višestrukim gubicima poslova vezanih za prodaju (Dwivedi et al. 2021). Pretpostavka je da bi povećana upotreba VI mogla smanjiti ulaganja u obuku i razvoj prodaje, povećati broj situacija monopola/oligopola i smanjiti prodajni posao dostupan u toj industriji (Dwivedi et al. 2021). Upotreba VI mogla bi dovesti do masovnog gubitka radnih mesta i većeg dispariteta prihoda. Međutim, može dovesti i do ukupnog pozitivnog efekta na radnu snagu, stvaranjem ekonomskog rasta i smanjenjem cena. Poslovni lideri tada mogu da prebace svoju radnu snagu na nove poslove koji zahtevaju razmišljanje višeg nivoa. Prema izveštaju Svetskog ekonomskog foruma⁴, procenjuje se da će VI ukinuti 75 miliona radnih mesta, ali će zato stvoriti čak 133 miliona novih. Međutim, javnost često nije informisana o ovim činjenicama, te je zabrinuta zbog automatizacije postojećih poslova i nestanka profesija.

Jedan od glavnih izazova u pogledu usvajanja i primene VI jeste znanje i percepcija javnosti o njoj. Nedovoljno znanje o funkcionisanju VI može stvoriti strah ili nepoverenje. Stoga je važno ispitati da li je to slučaj sa javnošću u Republici Srbiji, kako bi se moglo raditi na ovom problemu, i kako bi se ostvarila edukacija šire javnosti.

Prethodne studije

Postojeće studije u ovom domenu malobrojne su i sprovedene prvenstveno u visokorazvijenim zemljama. Iako je bilo nekih istraživanja percepcije javnosti o VI, na primer, istraživanja u obliku ankete (Cave et al. 2019; Zhang & Dafoe 2019), istraživanje stavova studenata medicine prema upotrebi VI (Pinto dos Santos et al. 2019), analiza sentimenata (Garvey & Maskal 2020), ovakva istraživanja rađena su u zapadnom kontekstu engleskog govornog područja. Čak i u ovim bolje proučenim kontekstima, ostaje još mnogo toga da se nauči, jer se i tehnologija i javna diskusija brzo razvijaju (Sindermann et al. 2021). Istraživanje sprovedeno u okviru ovog projekta jeste prvo istraživanje o stavovima javnosti u Republici Srbiji o upotrebi VI.

Studija koju su sprovedeli istraživači sa Instituta Oksford Internet, „Globalni stavovi o VI i automatizovano donošenje odluka“ (Neudert et al. 2020) pokazuje da je percepcija javnosti o upotrebi VI podeljena, pri čemu je

4 World Economic Forum <https://www.weforum.org/>

populacija na Zapadu više zabrinuta od populacije na Istoku. Istraživanje je bilo o tome kako ljudi percipiraju rizike upotrebe VI i mašinskog učenja prilikom donošenja odluka. U upitniku se postavilo pitanje da li će VI uglavnom pomoći ili uglavnom naneti štetu ljudima u toku narednih 20 godina. Rezultati su pokazali sledeće: (1) Zabrinutost da će VI biti štetna najviša je u Severnoj Americi (47%) i Latinskoj Americi (49%), a znatno manja kod ispitanika jugoistočne Azije (25%) i istočne Azije (11%). U Evropi 43% ljudi smatra da će VI biti štetna i 38% veruje da će biti korisna. Ljudi iz Severne Amerike i zapadne Evrope smatraju da je razvoj VI i robotike više štetan nego koristan, dok ispitanici iz južne i istočne Azije ovaj razvoj vide kao koristan i pozitivan. (2) Optimizam u pogledu VI i donošenja odluka najviši je u Kini, gde samo 9% ispitanika veruje da je VI štetna. Stanovništvo Azije ima najpozitivnije stavove. (3) U pogledu profesija, poslovni i državni rukovodioci (47%) i drugi profesionalci (44%) su među najvećim entuzijastima za donošenje odluka u vezi sa VI, dok su radnici u proizvodnji (35%) i radnici u uslužnom sektoru (35%) manje sigurni u to da će VI pomoći društvu. Rezultati su pokazali da muškarci, mlađi odrasli, oni sa višim stepenom obrazovanja, vide VI kao pozitivnu za društvo. U istraživanju je postojalo samo jedno pitanje, te je sugestija istraživača za dalja istraživanja bila da se ispituju detaljniji aspekti ovog pitanja. Pitanje iz ovog istraživanja bilo je uključeno u moju studiju.

Rezultati istraživanja „U. S. Javno mišljenje o upravljanju VI“ (Zhang & Dafoe 2020) pokazali su da značajna većina američkih građana (82%) veruje da država i javne agencije treba pažljivo da nadgledaju uvođenje VI. Američki građani pokazali su nizak nivo poverenja u vladine organizacije, korporativne institucije i donosiocice odluka koje bi trebalo da razvijaju i primenjuju VI u javnom interesu. Javnost najviše veruje univerzitetskim istraživačima (50%) i američkoj vojsci (49%). Istraživanje je otkrilo da 23,6% ispitanika ne zna ništa o VI. U poređenju sa Amerikancima koji pohađaju verske ustanove barem nekoliko puta godišnje, oni koji to nikada ne čine u većoj meri podržavaju razvoj VI. Ispitanici iz Amerike koji nikada ne pohađaju verske ustanove imaju pozitivnije stavove o efektima VI na društvo.

Istraživanje *Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries* (Kelley et al. 2021) bilo je fokusirano na pitanja o očekivanom uticaju VI, o sentimentima prema VI, i varijacijama u odgovorima u okviru nekoliko zemalja. Rezultati su pokazali široko rasprostranjenu percepciju da će VI imati značajan uticaj na društvo, ali priroda ovih efekata nije utvrđena. Istraživači su identifikovali četiri grupe sentimentata prema VI: uzbudljivo, korisno, zabrinjavajuće, i futuristički.

Cilj studije *Public Perception of Artificial Intelligence in Medical Care: Content Analysis of Social Media* (Gao et al. 2020) bio je da istraži percepciju

javnosti o veštačkoj inteligenciji u medicinskoj nezi kroz analizu sadržaja podataka društvenih medija, uključujući specifične teme zbog kojih je javnost zabrinuta; i javno mnjenje o tome da li veštačka inteligencija može da zameni lekare. U studiji je analizirano 2315 postova povezanih sa veštačkom inteligencijom u medicinskoj nezi i klasifikovano analizom sadržaja. Rezultati su pronašli tri tipa tema o veštačkoj inteligenciji o kojima se raspravljalo na platformi: (1) tehnologija i primena, (2) razvoj industrije i (3) uticaj na društvo. Od 956 postova u kojima su izraženi stavovi javnosti, 59, 4%, 34, 4% i 6, 2% postova izrazilo je pozitivne, neutralne i negativne stavove. U 200 postova u kojima se pominju stavovi javnosti o zameni lekara veštačkom inteligencijom, 47, 5% i 32, 5% postova izrazilo je da bi veštačka inteligencija u potpunosti ili delimično zamenila lekare, respektivno, dok je u 20% postova prisutno uverenje da veštačka inteligencija ne bi mogla da zameni lekare. Nalazi su pokazali da su ljudi najviše zabrinuti za VI tehnologiju i primene. Generalno, većina ljudi imala je pozitivne stavove i verovala da će VI lekari u potpunosti ili delimično zameniti ljudske. U poređenju sa prethodnim studijama o lekarima, opšta javnost ima pozitivniji stav prema medicinskoj veštačkoj inteligenciji. Nedostatak poverenja u veštačku inteligenciju i odsustvo faktora humanističke nege, suštinski su razlozi zašto neki ljudi i dalje imaju negativan stav prema medicinskoj veštačkoj inteligenciji.

Istraživači su očekivali da pozitivniji stavovi prema VI budu povezani sa većom spremnošću da se koriste produkti veštačke inteligencije. Negativniji stavovi bi tako trebalo da budu povezani sa manjom spremnošću da se takvi produkti VI koriste (Sindermann et al., 2021). Opšti stavovi ljudi prema VI verovatno će igrati veliku ulogu u njenom prihvatanju (Schepman & Rodway 2020).

Istraživanje

Metod

Cilj istraživanja bio je ispitati stavove javnosti u Republici Srbiji prema upotrebi VI, utvrditi koji faktori utiču na ove stavove i ispitati da li javnost deli zabrinutosti identifikovane kod filozofa (etičara) u debatama o VI. Takođe, cilj je bio ispitati prethodno znanje javnosti o VI, kao i poverenje građana u korišćenje VI u nekoliko institucija u najboljem interesu javnosti.

Ciljana grupa bila je javnost u najširem smislu (građani Republike Srbije). Činili su je ispitanici različitog starosnog doba, pola, stepena obrazovanja, profesije. Kada je reč o etičkim aspektima, istraživanje je sprovedeno sa ljudima koji su dali informisanu saglasnost i pristanak da u njemu učestvuju. Ispitanici su bili informisani o tome da je njihovo učešće dobrovoljno,

da njihovo odbijanje da učestvuju ne podrazumeva nikakvu kaznu, i da u bilo kom trenutku mogu prekinuti svoje učešće. Istraživanje je bilo anonimno. Odobrio ga je Etički odbor Instituta društvenih nauka u Beogradu.

Postupak

Istraživanje u obliku onlajn upitnika postavljeno je putem platforme Lime Survey (resursa Instituta društvenih nauka, Internet panela koji je baziran na softveru Lime Survey). Ispitanici su bili ljudi u Republici Srbiji (737) koji su dobrovoljno popunili upitnik. Donja starosna granica za učešće bila je 15 godina (prva godina srednje škole u Srbiji), jer je koncept upitnika bio previše napredan za učenike osnovnih škola. Upitnik se sastojao od nekoliko kategorija pitanja. Prvu su činila ona koja se odnose na opšte podatke: pol, godinu rođenja, nivo stečenog obrazovanja, profesiju kojom se ispitanici bave. Drugu grupu činila su pitanja koja se odnose na upotrebu tehnologije. Treću grupu – pitanja koja se odnose na prethodnu informisanost i znanje o VI. Drugi deo upitnika sastojao se od kratkog objašnjenja pojma VI i konkretnih primera njene upotrebe. Poslednju grupu činila su pitanja koja su se odnosila na stavove prema VI, odnosno različitim aspektima primene VI. Neki od primera ponuđenih odgovora su: *VI će svakodnevni život učiniti lakšim; Plašim se da će VI diskriminisati ljude na osnovu njihove rase, pola, seksualne orijentacije i materijalnog statusa; Postoje mnogi benefiti primene VI.* Učesnici su na pitanja odgovarali zaokruživanjem podeoka koji pokazuje u kojoj se meri slažu ili ne slažu sa određenom tvrdnjom. Podeoci su činili Likert skalu od 1 do 5.

Statistička analiza

Podaci su obrađeni i analizirani u R (R Core Team 2014) programskom jeziku za statističko računarstvo i grafiku. Korišćeni su paketi za proširenje funkcija R jezika: psych (Revelle 2017), rstatix (Kassambara 2020), i epitools (Aragon 2020). Opšti stav prema upotrebi VI procenjen je kreiranjem kompozitnog skora na osnovu pitanja u kojima su ispitanici izražavali svoje stavove prema različitim aspektima upotrebe VI. Odgovori nisu bili normalno distribuirani (što je pokazao Kolmogorov–Smirnov test i Q-Q grafikoni), zbog čega sam se opredelila za neparametarske statističke metode u testiranju hipoteza.

Rezultati i diskusija

H1: Javnost u Republici Srbiji imaće dominantno negativne stavove prema primeni VI. H1 nije potvrđena. Javnost u Srbiji nema dominantno negativne stavove prema upotrebi VI.

S obzirom na to da podaci nisu normalno distribuirani (što je potvrđeno Kolmogorov–Smirnovim testom i Q-Q vizualizacijom), hipoteza je testirana neparametarskim testom znaka (Sign test). Srednja vrednost stavova javnosti prema VI bila je 2.995 od 5. U procentima, 49, 39% ispitanika ima pozitivan stav prema VI, 3, 80% ima neutralan stav, dok 46, 81% ima negativan stav prema upotrebi VI. Ne postoji statistički značajna razlika između broja ispitanika koji imaju negativne stavove i ispitanika koji imaju pozitivne stavove prema upotrebi VI ($s = 364, p < 0.77$).

H2: Godine ispitanika imaju uticaj na stavove prema VI, tako da mlađi ispitanici imaju pozitivnije stavove. Ova hipoteza je potvrđena. Kendalov rang koeficijenta korelacije pokazao je da godine ispitanika imaju uticaj na stavove prema VI ($\tau = -0,076, p < 0,0002$), tako da postoji negativna korelacija između godina ispitanika i pozitivnog stava prema upotrebi VI. To znači da što je osoba mlađa, ima pozitivniji stav prema upotrebi VI.

H3: Javnost će pokazati zabrinutost u pogledu nestanka profesija usled napretka VI. H3 je potvrđena. Rezultat je pokazao da je 49, 93% ispitanika pokazalo zabrinutost u pogledu gubitka posla usled napretka VI, 23, 47% ima neutralan stav, ne pokazuju ni zabrinutost, niti nezabrinutost, dok 26, 59% nije zabrinuto.

H4: Javnost u Republici Srbiji nije informisana o VI u velikoj meri. H4 je delimično potvrđena. Ogromna većina ispitanika čula je za VI. Iako upoznatost javnosti o VI nije slaba, 3, 9% ispitanika ne zna ništa o VI, većina (40,16%) je donekle upoznata sa VI, dok 30, 39% i 8, 27% zna prilično o VI.

H5: Javnost će ispoljiti zabrinutost zbog mogućnosti diskriminacije od strane VI. Ova hipoteza je potvrđena. Javnost pokazuje zabrinutost za potencijalnu diskriminaciju VI na osnovu pola, rase, seksualne orijentacije i materijalnog statusa. Samo 10% stanovnika ne pokazuje ovu zabrinutost, dok je 28, 76%, 18, 58% i 22, 66% zabrinuto donekle, prilično i veoma.

H6: Kada je reč o poverenju građana u različite institucije da razvijaju i koriste veštačku inteligenciju u najboljem interesu javnosti, bili su ispitivani sledeći resori: Vlada Republike Srbije, vojska RS, MUP RS, Ministarstvo zdravlja RS, Ministarstvo pravde RS, javna preduzeća i lokalne samouprave, istraživači sa Univerziteta, međunarodne istraživačke organizacije (poput CERN-a), tehnološke kompanije (npr. Gugl, Fejsbuk, Epl, Majkrosoft) i neprofitne organizacije za istraživanje VI (poput Open AI). Rezultati su pokazali da javnost u Republici Srbiji najmanje poverenja ima u Vladu RS, a potom MUP RS i Ministarstvo pravde RS, dok najviše veruje istraživačima sa Univerziteta i međunarodnim istraživačkim organizacijama (poput CERN-a). Pokazalo se da čak 70% ispitanika nema nimalo poverenja u Vladu RS, u MUP RS 66, 4%. Hijerarhijski posmatrano, ispitanici najviše poverenja imaju u istraživače sa Univerziteta, međunarodne istraživačke

organizacije, neprofitne organizacije za istraživanje VI, tehnološke kompanije, potom, znatno manje, u Ministarstvo zdravlja, vojsku RS, javna preduzeća i lokalne samouprave, Ministarstvo pravde RS, MUP RS i Vladu RS.

Zaključak

Kada je reč o postojećem stanju u ključnim sektorima u Republici Srbiji, rad je usmeren naročito na obrazovanje i nauku, kao i etičke aspekte primene veštačke inteligencije i smernice za budućnost. Kako se navodi u *Strategiji za razvoj veštačke inteligencije u Republici Srbiji*⁵, obrazovanje je ključni faktor za razvoj VI u Srbiji. U ovoj oblasti, kao najrelevantniji preduslovi za razvoj veštačke inteligencije, prepoznate su i istaknute veštine radne snage, veštine diplomaca i veštine opšte populacije. Poslednje navedeno podrazumeva sposobnost građana da pročitaju digitalne sadržaje ili da koriste računar na elementarnom nivou. U ovom segmentu postoji značajan prostor za unapređenje i edukaciju stanovništva, jer i samo korišćenje veštačke inteligencije podrazumeva da građani znaju da prepoznaju kada su u interakciji sa njom (na primer, kada im asistenciju pruža virtuelni asistent, a ne čovek). Istraživanje *Stavovi javnosti o upotrebi veštačke inteligencije* pokazalo je da je javnost u Republici Srbiji uglavnom čula za pojam *veštačke inteligencije*, ali da 4% ispitanika ne zna ništa o veštačkoj inteligenciji, dok 8% zna prilično, a da je 40% donekle upoznato sa ovim pojmom i njegovom primenom.

Kada je reč o etičkim i pravnim aspektima primene veštačke inteligencije, potrebno je da regulatorni okviri postignu adekvatan balans između zaštite građana i omogućavanja razvoja inovacija. Razvoj veštačke inteligencije donosi mnogo prednosti, ali i izazova, kako za pojedinca tako i za društvo u celini. Kako se ističe i u *Strategiji*, da bi se osigurala dostupnost, bezbednost i jednaka i pravedna primena VI, kao i izgradilo poverenje društva prema VI, važno je da tehnološki napredak u oblasti VI prati odgovorno i svesno promišljanje i odgovor i rešavanje izazova koji se javljaju usled razvoja i primene veštačke inteligencije. Ključni zahtevi koji se navode su sledeći: ljudsko delovanje i nadzor, tehnička robusnost i sigurnost, privatnost i upravljanje podacima, transparentnost, raznolikost, nediskriminacija i pravednost, dobrobit za društvo i životnu sredinu i odgovornost. Otvoreni podaci predstavljaju još jedan neophodan resurs za razvoj veštačke inteligencije.

Rezultati sprovedenog istraživanja u okviru projekta pružili su podatke o stavovima javnosti u Republici Srbiji prema različitim aspektima upotrebe

5 Strategija za razvoj veštačke inteligencije u Republici Srbiji za period 2020–2025.

veštačke inteligencije. Pokazalo se da ispitanici imaju podeljeno mišljenje o upotrebi veštačke inteligencije, pri čemu skoro polovina ispitanika ima pozitivne, druga polovina negativne stavove, dok nekolicina ima neutralne stavove prema ovoj upotrebi. Takođe, pokazalo se da je javnost donekle upoznata sa VI. Javnost je pokazala zabrinutost u pogledu nestanka profesija usled napretka u oblasti veštačke inteligencije, kao i zabrinutost zbog potencijalne diskriminacije od strane veštačke inteligencije prema polu, rasi, seksualnoj orijentaciji i materijalnom statusu. Rezultati su pokazali da javnost u Republici Srbiji najmanje poverenja ima u Vladu RS, a potom MUP RS i Ministarstvo pravde RS, dok najviše veruje istraživačima sa Univerziteta i međunarodnim istraživačkim organizacijama. Rezultati ukazuju na to da je potrebno fokusirati se na obrazovanje javnosti o potencijalnim izazovima i načinima da se oni spreče. Potrebno je više transparentnosti u procesima donošenja odluka vezanim za primenu VI, kao i informisanje javnosti o tome. Država treba da ulaže više u inicijative za podizanje svesti javnosti i obrazovanje kako bi se razvile kompetencije svih građana, a posebno mladih generacija, da bolje razumeju implikacije VI. Rad je imao za cilj informisati debatu o upotrebi VI donošenjem rezultata mišljenja javnosti. Stavove javnosti potrebno je uzeti u obzir ukoliko se VI želi implementirati na što efikasniji način. Prihvatanje, usvajanje i upotreba VI zavisi od opšte podrške građana (ili bar izostanka aktivnog otpora).

Uprkos značajnim koristima koje društvo može da ima od razvoja VI, postoje potencijalni i aktuelni izazovi, koje je potrebno sagledati i ugraditi u planove vezane za dalji razvoj. To su, pre svega, zaštita podataka o ličnosti, potreba prilagođavanja obrazovnog sistema i kvalifikacija, etički izazovi, kao što je prevencija diskriminacije, transparentnost i poverenje. Primarni cilj i želja nam je da javnost u Srbiji bude informisana o veštačkoj inteligenciji i njenim primenama u svetu, ali naročito u Republici Srbiji. Pored toga, ono što je takođe važno jeste poverenje građana u institucije (i državne i privatne) da koriste veštačku inteligenciju u najboljem interesu javnosti. Potrebno je da država u narednom periodu izgradi ovo poverenje, povećanjem edukacije i transparentnosti.

Rezultati istraživanja projekta vredan su resurs za buduće bavljenje ovom temom i zbog toga što do sada slična ispitivanja javnog mnjenja nisu rađena. U Republici Srbiji ne postoje istraživanja koja se bave ovom temom. Javno mnjenje nije oblikovalo razgovore o upotrebi veštačke inteligencije. Stoga je ovu prazninu u informacijama i stavovima u pogledu upotrebe VI bilo potrebno popuniti jednim dijalogom između države, privatnog sektora i šire javnosti, odnosno svih građana RS. Demokratsko i otvoreno društvo uključeno je ne samo u proces donošenja odluka već i u javnu raspravu, izražavanje mišljenja, ideja, debate, a time i donošenje odluka. Potrebno

je da donosioci odluka budu upoznati sa mišljenjem i stavovima javnosti (stručnjaka, ali i drugih ljudi) kako bi donošenje odluka uskladili sa etičkim i pravnim principima, znanjem i stavovima javnosti i društva. Rezultati istraživanja pružili su konkretne odgovore na pitanja kako javnost u Srbiji percipira upotrebu veštačke inteligencije, da li je zabrinuta, da li ima poverenja (i koliko) u konkretne institucije da koriste veštačku inteligenciju. Važno je bilo ispitati stavove stručnjaka u ovoj oblasti, ali isto tako i građana, budući da se određena praksa u društvu može primeniti samo ukoliko je to društvo otvoreno za ovo prihvatanje.

Ono što želimo postići kada je reč o ovoj temi jeste edukacija javnosti o veštačkoj inteligenciji, povećanje transparentnosti i dijalog između javnosti, stručnjaka, privatnog sektora, i države. Na koji način se ovo može ostvariti? Prvi korak jeste obrazovanje. Potrebno je javnosti pružiti informacije i edukaciju o ovoj temi. Takođe je potrebno više transparentnosti u procesima donošenja odluka vezanih za primenu VI. Država treba da ulaže više u inicijative za podizanje svesti javnosti i obrazovanje. U *Strategiji razvoja veštačke inteligencije u Republici Srbiji za period 2020–2025. godine* preporučuje se obrazovanje eksperata i poboljšanje situacije u Republici Srbiji kada je reč o stručnom kadru za razvoj i primenu veštačke inteligencije. Međutim, važan aspekt spremnosti celog društva za korišćenje i prihvatanje veštačke inteligencije jesu i stavovi i znanje javnosti o njenoj upotrebi. Stoga prepoznavamo da je važno fokusirati se i na obrazovanje javnosti o veštačkoj inteligenciji. Kako bi se sve ovo ostvarilo, neophodna je jača saradnja između državnih aktera – vlade, parlamenta, sudstva, privatnih kompanija, akademske zajednice, nevladinih organizacija, međunarodnih organizacija, kao i šire javnosti.

Dakle, potrebno je delati u dva pravca: ka edukaciji šire javnosti u Republici Srbiji i ka povećanju transparentnosti i dijaloga između javnosti sa jedne, i države i privatnog sektora, sa druge strane. Drugi fokus treba da bude na daljim istraživanjima stavova javnosti i eksperata o različitim aspektima primene veštačke inteligencije, kao i daljih etičkih izazova koji se javljaju u ovom kontekstu.

Literatura

- Aragon, Tomas (2020), *Epitools: Epidemiology Tools*, <https://CRAN.R-project.org/package=epitools>
- Barabas, Chelsea, Madars Virza, Karthik Dinakar et al. (2017), Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment, u *Proceedings of Machine Learning Research* 81: 1–15, Conference on Fairness, Accountability, and Transparency, <https://arxiv.org/abs/1712.08238> (17. 3. 2022).

- Bostrom, Nick, Eliezer Yudkowsky (2019), *Artificial Intelligence Safety and Security*, New York: Chapman and Hall/CRC.
- Cave, Stephen, Kate Coughlan, Kanta Dihal (2019), 'Scary Robots': Examining Public Responses to AI: 331–337, u *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, Honolulu HI USA: ACM, <https://dl.acm.org/doi/10.1145/3306618.3314232> (17. 3. 2022).
- Chancellor, Stevie, Birnbaum, Michael L., Caine, Eric D., Silenzio, Vincent M. B. and De Choudhury, Munmun (2019), A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media: 79–88, u *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Atlanta GA USA: ACM, <https://dl.acm.org/doi/10.1145/3287560.3287587> (17. 3. 2022).
- Dwivedi, Yogesh K., Laurie Hughes, Elvira Ismagilova et al. (2021), „Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy”, *International Journal of Information Management* 57(7): 101994.
- Eubanks, Virginia (2017), *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York: St. Martin's Press.
- Fetzer, James H. (1990), *Artificial Intelligence – Its Scope and Limits*, Dordrecht: Kluwer.
- Galanos, Vassilis (2018), Artificial Intelligence Does Not Exist: Lessons from Shared Cognition and the Opposition to the Nature/Nurture Divide, 359–373, u Kreps, D., Ess, C., Leenen, L., and Kimppa, K. (ur.), *This Changes Everything – ICT and Climate Change: What Can We Do?* IFIP Advances in Information and Communication Technology, Cham: Springer International Publishing, http://link.springer.com/10.1007/978-3-319-99605-9_27 (21. 3. 2022).
- Gao, Shuqing, Lingnan He, Yue Chen et al. (2020), „Public Perception of Artificial Intelligence in Medical Care: Content Analysis of Social Media”, *Journal of Medical Internet Research* 22(7): e16649.
- Garvey, Colin, Chandler Maskal, (2020), „Sentiment Analysis of the News Media on Artificial Intelligence Does Not Support Claims of Negative Bias Against Artificial Intelligence”, *OMICS: A Journal of Integrative Biology* 24(5): 286–299.
- Jiang, Fei, Yong Jiang, Hui Zhi et al. (2017), „Artificial Intelligence in Healthcare: Past, Present and Future”, *Stroke and Vascular Neurology* 2(4): 230–243.
- Kaplan, Andreas, Michael Haenlein (2019), „Siri, Siri in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence”, *Business Horizons* 62(1): 15–25.
- Kassambara, Alboukadel (2020), *Rstatis: Pipe-Friendly Framework for Basic Statistical Tests*, <https://CRAN.R-project.org/package=rstatis>
- Kelley, Patrick Gage, Yongwei Yang, Courtney Heldreth et al. (2021), „Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries”: 627–637, u *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. Virtual Event USA: ACM, <https://dl.acm.org/doi/10.1145/3461702.3462605> (27. 8. 2021).
- Lawrence, David R., César Palacios-González, John Harris (2016), „Artificial Intelligence: The Shylock Syndrome”, *Cambridge Quarterly of Healthcare Ethics* 25(2): 250–261.
- Neudert, Lisa-Maria, Aleksu Knuutila, Philip N. Howard, (2020), „Global Attitudes Towards AI, Machine Learning & Automated Decision Making”, Working paper 2020.10, Oxford, UK: Oxford Commission on AI & Good Governance.

- Pinto dos Santos, D., Daniel Giese, Sebastian Brodehl et al. (2019), „Medical Student’s Attitude towards Artificial Intelligence: A Multicentre Survey”, *European Radiology* 29 (4): 1640–1646.
- Prainsack, Barbara (2019), „Logged out: Ownership, Exclusion and Public Value in the Digital Data and Information Commons”, *Big Data & Society* 6 (1). <https://doi.org/10.1177/2053951719829773>
- R Core Team (2014), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, <http://www.R-project.org/>
- Revelle, William R (2017), *Psych: Procedures for Personality and Psychological Research*, R package version 1.9.12, <http://personality-project.org/r>, <https://personality-project.org/r/psych-manual.pdf>
- Russell, Stuart J., Peter Norvig, Ernest Davis (2010), *Artificial Intelligence: A Modern Approach*, Upper Saddle River: Prentice Hall.
- Schepman, Astrid, Paul Rodway (2020), „Initial Validation of the General Attitudes towards Artificial Intelligence Scale”, *Computers in Human Behavior Reports* 1: 100014.
- Sindermann, Cornelia, Peng Sha, Min Zhou et al. (2021), „Assessing the Attitude Towards Artificial Intelligence: Introduction of a Short Measure in German, Chinese, and English Language”, *KI – Künstliche Intelligenz* 35 (1): 109–118.
- Sun, Tara Qian, Rony Medaglia (2019), „Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare”, *Government Information Quarterly* 36 (2): 368–383.
- Szollosy, Michael (2017), „Freud, Frankenstein and Our Fear of Robots: Projection in Our Cultural Perception of Technology”, *AI & SOCIETY* 32(3): 433–439.
- Vasiljeva, Tatjana, Ilmars Kreituss, Ilze Lulle (2021), „Artificial Intelligence: The Attitude of the Public and Representatives of Various Industries”, *Journal of Risk and Financial Management* 14(8): 339.
- Zhang, Baobao and Dafoe, Allan (2019), „Artificial Intelligence: American Attitudes and Trends”. *SSRN Electronic Journal*, <https://www.ssrn.com/abstract=3312874> (17. 3. 2022).
- Zhang, Baobao, Allan Dafoe, (2020), U. S. Public Opinion on the Governance of Artificial Intelligence: 187–193, u *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. New York: Association for Computing Machinery. Available, <https://doi.org/10.1145/3375627.3375827> (24. 3. 2022).
- Złotowski, Jakub, Kumar Yogeeswaran, Christoph Bartneck (2017), „Can We Control It? Autonomous Robots Threaten Human Identity, Uniqueness, Safety, and Resources”, *International Journal of Human-Computer Studies* 100: 48–54.

Marina Š. Budić

ETHICAL DILEMMAS AND ATTITUDES TOWARDS THE USE OF ARTIFICIAL INTELLIGENCE

Summary

This paper was created within the project *Ethics and AI: Ethics and Public Attitudes on the Use of Artificial Intelligence*, within the program *Serbia and Global Challenges: Towards Fairer and More Democratic Public Policies*, organized by the Open Society

Foundation Serbia in partnership with the Institute of Philosophy and Social theory. The first part of the paper consists of a theoretical study of ethical challenges arising from the development and application of artificial intelligence. The second part of the paper is an empirical study of public attitudes towards various aspects of the use of artificial intelligence.

The term *artificial intelligence (AI)* describes systems that mimic cognitive functions generally associated with human attributes such as learning, speech, and problem solving. A more detailed characterization describes artificial intelligence in the context of its ability to independently interpret and learn from external data in order to achieve specific outcomes through flexible adaptation. AI is designed to achieve specific goals, and refers to computer systems that execute tasks, make decisions or achieve goals in complex situations, without explicit instructions from people. Examples of the use of AI are: robots, industrial used in production, humanoid, interacting with humans, autonomous vehicles, virtual assistants and travel reservation agents, natural language processing tools, translation, image classification and face identification, drones that do not require a controller, military drones. In the medical sector, robots are used to assist the elderly and surgeons, as receptionists, to diagnose and map diseases.

AI technology is no longer a field of futurology, but an integral part of everyday life. AI permeates every aspect of society. It is used in various sectors - health, economics, education and learning, security, and other spheres of life. Progress in area AI affects all aspects of society, including the labor market, transport, health care, education and national security.

AI can greatly facilitate people's daily and business lives, but it also brings with it certain challenges, such as privacy, transparency, discrimination, job loss, responsibility, knowledge, familiarity, perception and attitudes of people towards the use of AI. I have conducted an empirical study of public attitudes towards various aspects of the use of artificial intelligence. The participants in the research were the public in the Republic of Serbia (N = 737) who voluntarily filled out an online questionnaire. The questionnaire consisted of general socio-demographic questions, questions about familiarity with artificial intelligence, a brief introduction to the concept of artificial intelligence, and questions about the attitudes of participants towards the application of artificial intelligence. The results showed that these attitudes are influenced by factors such as the age of the respondents. Also, the results showed that the public is concerned about the disappearance of professions due to the development of artificial intelligence, as well as discrimination by artificially intelligent systems. When it comes to citizens' trust in institutions to use artificial intelligence in the best interest of the public, it has been shown that the public in the Republic of Serbia has the least trust in the RS Government, the RS Ministry of the Interior and the RS Ministry of Justice, while trusts the most researchers from universities and international research organizations. The results suggest a need to focus on educating the public about potential challenges and ways to prevent them. In addition, these results inform both the academic and general public, as well as decision-makers, which can lead to a dialogue between these parties. I argue that considering different aspects of public attitudes toward artificial intelligence enhances this debate. The views of philosophers, lawyers and scientists dealing with this topic are present in the ongoing debate, but not the attitude of the public, which reflects the need for this type of research.

Keywords

ethics, artificial intelligence, public attitudes, education, privacy, discrimination, transparency, democratic values