

Janez Bregant

Filozofska fakulteta Univerze v Mariboru

Umetna inteligenca v praksi (2. del): nekaj etičnih pomislekov

Nagel razvoj umetne inteligence (UI) v zadnjem obdobju in njena samoumevna uporaba v našem vsakodnevnem življenju sta povečala zanimanje za vse vrste umetno izdelanih avtonomnih sistemov tako na strani javnosti kot stroke. Ker UI ni več razumljena zgolj kot priročno orodje, ampak tudi kot samostojni akter, ne moremo mimo vprašanja, kakšen vpliv ima njena uporaba na naše življenje v moralnem smislu. Odgovori na vprašanja, kot so »Ali je UI odgovorna za svoja dejanja?«, »Kaj so družbene, pravne in kulturne posledice njenih odločitev?«, »Ali naj bi bili avtonomni sistemi prosto dostopni na trgu?« itn., so tisto, kar v veliki meri določa, ali bomo UI v prihodnosti zaupali ali ne. V članku predstavimo pet nevarnosti, ki nam ob nebrzdani uporabi UI pretijo in s pomočjo analize trenutnega dogajanja na področjih, s katerimi so povezane, zaključimo, da družba vse bolj postaja talec industrije inteligentnih sistemov, ki v nobeni fazi svojega razvoja (oblikovanje, izgradnja, namestitve in ocenjevanje) niso narejeni tako, da bi v interakciji z nami spoštovali človekove pravice in ravnali skladno s sprejetimi družbenimi vrednotami.

Ključne besede: umetna inteligenca, strojno učenje, obdelava naravnega jezika, etika, človekove pravice.

0. Uvod

Cilj ustvarjalcev modelov UI je bil že od nekdaj narediti takšen umetni sistem, ki bo sposoben samostojno prepoznati, obdelati in rešiti naloge ter probleme, na katere bo naletel v svojem okolju, ne da bi pri tem človek moral natančno določiti vsak njegov korak. V idealnem primeru bi se morali biti modeli UI sposobni prilagajati spremembam v okolju, kar bi od njih zahtevalo tudi sprejemanje odločitev. Kot smo videli v prejšnjem članku (Bregant, 2019), se prednost umetnih sistemov kaže tam, kjer je potrebno iz velike in neurejene količine podatkov razbrati ustrezne oziroma zahtevane vzorce. V korist strojnega odločanja se običajno navaja argument, da se s tem iz postopka sprejemanja odločitev izloči predsodke, ki pri človeškem odločanju pomembno vplivajo na rezultat. To pa ne drži, saj tehnologija niti slučajno ni nevtralna. Algoritem je dober le toliko, koliko so dobri podatki, ki jih obdeluje, pri čemer lahko predsodke, ki se zrcalijo v zbranih in-

formacijah posvoji do te mere, da postane obstoječi problem še večji. Še svež je primer rasistične UI, ki ni sposobna prepoznati obrazov temnopoltnih ljudi, zaradi česar jih označuje kot gorile. Ti v naši družbi globoko zakoreninjeni predsodki, se v slovarju UI imenujejo *predsodki strojnega učenja*,¹ naloga razvijalcev umetnih sistemov pa je, da jih prepoznajo in njihov vpliv čim bolj omejijo (Ogola, 2019).

Tako je očitno, da so odločitve modelov UI zgolj navidezno objektivne, iz česar izhaja kup praktičnih moralnih problemov. Kdo je odgovoren za odločitve, ki jih sprejmejo avtomatizirani umetni sistemi?, Ali se dajo moralne odločitve, ki jih človek sprejema intuitivno, sprogramirati?, Ali sme UI, razvita v azijskem okolju, sprejemati moralne odločitve v evropskem okolju? itn. Vprašanje je, katere moralne kriterije mora UI izpolniti, da bo pri sprejemanju svojih odločitev, ki se dotikajo naših življenj (npr. prijava za službo, odobritev kredita, odločitev o krivdi toženega v sodnem rocesu itd.), pravična? Politične smernice glede tega, kako uskladiti strojno odločanje s pravom in moralo, je na ravni Evropske unije pripravila tako imenovana *Strokovna skupina za UI*,² ki jo je leta 2018 imenovala Evropska komisija. Njen cilj je bil določitev etičnih standardov, ki naj bi pomagali pri tem, da bi odslej na UI gledali z zaupanjem. Poročilo vključuje naslednje zahteve:

- (i) nadzor (umetni sistemi morajo človeku pomagati pri sprejemanje argumentiranih odločitev, skladnih z njegovimi pravicami in svoboščinami; poleg tega morajo vključevati tudi primerne nadzorne mehanizme, ki se pri svojem delu zgledujejo po človeškem ravnanju);
- (ii) varnost (umetni sistemi morajo biti robustni in varni, vključevati morajo rezervni načrt v primeru, da gre kaj narobe, pa tudi natančni in zanesljivi);
- (iii) zaščita podatkov (umetni sistemi morajo zagotavljati in spoštovati pravico do zasebnosti; poleg tega morajo vključevati tudi mehanizme upravljanja s podatki, ki zagotavljajo kakovost njihove obdelave in omogočajo dostop do njih vsem, ki imajo do tega pravico);
- (iv) transparentnost (umetni sistemi morajo vključevati mehanizme za sledljivost sprejetih odločitev, človek pa mora vedeti, kdaj je v interakciji s takšnim sistemom in česa je ta zmožen);
- (v) nediskriminacija (umetni sistemi se morajo izogibati nepravilnim odločitvam na osnovi predsodkov, ker to povečuje marginalizacijo že tako odrinjenih skupin, dostopni pa morajo biti vsem skozi svoje celotno življenjsko obdobje);

¹ Angl. Machine-Learning-Biases.

² Angl. European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG).

(vi) trajnostna orientacija (umetni sistemi morajo biti prijazni do okolja, od njih naj bi v končni fazi imelo korist več generacij, njihov vpliv na vsa, v naravi živeča bitja, pa mora biti skrbno preverjen in preiščlen);

(vii) odgovornost (razviti je treba mehanizme, s katerimi se zagotavlja in preverja moralna in/ali pravna odgovornost umetnih sistemov ter njihovih dejanj; poleg tega je treba priskrbeti ustrezna pravna sredstva, ki se uporabljajo v primeru morebitnih kršitev) (Ogola, 2019).

V članku predstavimo pet področij človekovega življenja, na katerih se UI nepremišljeno in nebrzdano uporablja, pri čemer ni o spoštovanju omenjenih smernic za to, da bi dočakali zaupanja vredno UI, ne duha ne sluha. Prvo področje je (če z njihovimi opisi malo dramatiziramo) *zbiranje podatkov*, kjer umetni sistemi vedo o nas več, kot mi sami, drugo *odpoved zasebnosti*, kjer nekdo vedno ve, kaj počnemo, tretje *razvoj orožja*, kjer bodo bodoče (vedno bolj pogoste) vojne potekale med stroji, četrto *slovo od služb*, kjer smo vsi nadomestljivi z UI in peto *uničevanje okolja*, kjer pri proizvodnji elektrike, nujne za potešitev naših spletnih potreb, uporabi fosilnih goriv ni videti konca. V članku zaključimo, da trenutna nepreudarna uporaba umetnih sistemov na omenjenih področjih, predstavlja resno grožnjo obstoječim moralnim načelom, na katerih sloni družba.

1. Kopičenje in koncentracija podatkov

Očitno je, da si danes življenja brez družbenih omrežij, spletnega nakupovanja in elektronske pošte nihče več ne predstavlja. Preko platform, kot so *Facebook*, *YouTube*, *Twitter*, *Instagram* itn. uporabniki delijo svoje izkušnje z drugimi. Na različne načine, to je z besedilom, glasbo, sliko ali vsehkom, izražajo svoja mnenja o na primer znamkah, filmih, izdelkih, vladah itn. Ni pomembno, ali kritizirajo, zavračajo, sprejemajo, polemizirajo itn., s tem odkrito kažejo, kaj so njihove preference, pričakovanja in prioritete, s čimer vplivajo na svoje zasebno in javno življenje. Tukaj nam je UI v pomoč, saj lahko z njeno uporabo mnenja ljudi, izražena na takšne načine, odkrivamo, povezujemo in interpretiramo, kar v splošnem imenujemo *rudarjenje podatkov*.³

Rudarjenje podatkov v osnovi ne pomeni, da iz količine, ki je na voljo, vzamemo tiste, ki so za nas iz različnih razlogov uporabni, ampak predvsem odkrivanje skritih vzorcev in odstopanj v njih ter ocenjevanje tega, ali so med seboj povezani: na primer združevanje kupcev z istim okusom, opozarjanje na anomalije v proizvodnem postopku, ki kažejo na napake v delovanju sistema ali iskanje zveze med količino pridelka in vremenom. Kot vir se uporabljajo besedila, slike, filmi, glasba itn., iz katerih nevronske mreže s pomočjo strojnega učenja izluščijo želje-

³ Angl. *Data-Mining*.

ne informacije, največkrat to, kako se ljudje pri posredovanju različnih vsebin počutimo, kar omogoča UI tvorjenje emocionalnih razredov, v katere se potem vsebina razvršča (Dengel, 2019c).

Problem nastane, ko se rudarjenje podatkov izrodi v njihovo zbiranje, uporabnik sodobnih UI modelov pa postane izdelek. Pri tem prednjači pet velikih tehnoloških podjetij: *Google, Apple, Facebook, Amazon* in *Microsoft* (GAFAM). *Google* je s svojimi brezplačnimi storitvami počasi, a zanesljivo, postal ljubljenelec občinstva št. 1, ki je v zameno za svoje izdelke želel le eno, predajo naših osebnih podatkov, glede česar smo mu rade volje ustregli in mu tako omogočili neprestano črpanje informacij. *Gmail, Chrome, YouTube, Maps, Nest* itd., mu omogočajo, da ve, kaj zanima večino uporabnikov spleta. Njegov iskalnik je najbolj obiskana stran na spletu, opravi 90% vseh iskanj na spletu, operacijski sistem Android, ki je nameščen na 75% vseh mobilnih telefonov, pa trenutno največji sovražnik zasebnosti na sploh. Aplikacije so narejene tako, da naše podatke posredujejo naprej, na primer o lokaciji, razvoju nerojenega otroka, tlorisu stanovanja, nakupovalnih navadah itn., z nakupom *Google Asistenta* pa dobimo upravljalca naprav, ki ves čas posluša, spremlja in beleži, kaj počnemo ter tako pridno polni *Google*-ve podatkovno zbirke. Ker *Mastercard* še ni njegov, se je z njim dogovoril, da bo zbiral podatke tudi o nakupih v klasičnih trgovinah, opravljenih s to kreditno kartico, ki vsebujejo ime in priimek, čas, lokacijo, kupljeno blago, količino itn. Skratka, *Google* ve, kje smo, kdaj vstanemo, kdaj gremo v službo, kaj in kje smo nakupovali, kako dolgo smo bili na malici, kdaj smo kolesarili, kakšen je naš srčni utrip in drugo⁴ (Masten, 2019a).

Za boljši občutek glede tega, o kakšni količini podatkov, ki je na voljo omenjenim družbam, govorimo, si oglejmo, koliko informacij je na svetu znotraj različnih omrežij v obtoku v eni minuti: 18 milijonov poslanih SMS sporočil, 4.3 milijona ogledov video vsebin na *YouTubu*, 481.000 poslanih čivkov, 187 milijonov poslanih elektronskih sporočil, 3.7 milijona iskanj z *Googlom*, 973.000 vpisov v *Facebook*, za 862.823 ameriških dolarjev opravljenih nakupov, 375.000 naloženih aplikacij, 67 nameščenih virtualnih asistentov itn. (Dengel, 2019c). Povečana računska moč ter lahek in poceni dostop do zmogljivih računalnikov, pa niso omogočili zgolj pridobivanja, obdelave in uporabe podatkov, ampak tudi možnost zlorabe, še posebej, ker so informacije razporejene izredno neenakomerno. Vprašanje je, ali lahko kdo zagotovi, da se zbrani podatki, ki so bili tako ali drugače (zakonito ali nezakonito) pridobljeni, ne bodo uporabljali za izsiljevanje?

⁴ Tudi ostali tehnološki velikani niso nobena izjema: *Microsoft* se ukvarja s ciljnimi oglasi, tehnologijo prepoznavanja obraza, virtualno resničnostjo itn., *Facebook* preko oglasov in všečkov zbira informacije o imenih in naslovih ljudi, odnosih med njimi, njihovih družinah, lokaciji in pogovorih itn. ter je tako lastnik največje podatkovne baze o nas, *Amazon* pa preko spletne trgovine, v kateri je mogoče kupiti tako rekoč vse in *Alexi*, ki govori iz zvočnika *Echo*, podatke pa pošilja v oblak, pozna naše želje, potrebe in interese, da o naslovu dostave, kreditnih karticah in mestu nakupa niti ne govorimo (Masten, 2019a).

Skoncentriranost podatkov, ki so, kot se rado reče, nova nafta, v rokah posameznikov, saj na primer pri *Googlu*, *Amazonu* in *Facebooku*, drugje pa ni dosti drugače, odloča en sam človek, omenjenim podjetjem omogoča prevlado na informacijskem trgu, saj upravljajo, če malo pretiravamo, s skoraj celotno količino podatkov, ki je na voljo. To pa lahko ima negativne posledice na družbo, ker peščica ljudi s filtriranjem podatkov o svetu, vpliva na njegovo dojemanje in dogajanje v njem, zaradi česar že lahko govorimo o upravljanju družbe s pomočjo UI. Družba postane potem takšna kot avto, podobno kot avto upravljamo z vzvodi, z vzvodi upravljamo tudi družbo. Tako kot lahko avto zavije levo ali desno, lahko »levo« ali »desno« zavijejo tudi družbe, odvisno od želje, potreb ali interesov upravljalca, to je *Googla*, *Facebooka* ali *Amazona* (Grobelnik, 2018). V praksi se je to pokazalo pri zlorabi osebnih podatkov podjetja *Cambridge Analytica*, ki je z aplikacijo, ponujeno *Facebooku*, s prošnjo sodelovanja v akademski raziskavi, nezakonito pridobilo osebne podatke 50 milijonov uporabnikov Facebooka (njegov dizajn je omogočal tudi to, ne samo zbiranje odgovorov) in iz njih izdelalo njihove psihografične profile, s katerimi so ugotavljali, kakšno propagando je treba uporabiti, da bodo glasovali po naročnikovih željah.

Nekateri raziskovalci menijo, da smo ljudje v dobi digitalnih tehnologij, ki jih s pridom uporabljamo, postali surovina, iz katere tehnološka podjetja naredijo končni izdelek, to je napovedi o nas: kaj bomo kupili, kam bomo šli, s kom se bomo družili, kaj bomo delali, koga bomo volili itn. Ti podatki se prodajajo naprej tudi z namenom spreminjanja naših navad, običajev in želja, s čimer izgubimo pravico do prihodnosti. S tem GAFAM glede življenjskih navad ustvarjajo monopol ne zgolj nad sedanjostjo tretjine človeštva, ampak tudi nad njihovo prihodnostjo, saj lahko s ciljnim izdelki poljubno vplivajo na njihovo obnašanje in ga usmerjajo skladno z lastnimi interesi, kar lahko imenujemo *nadzorovalni kapitalizem*. (Zuboff, 2019)

Vprašanje je, kako to preprečiti? Ena možnost je podružabljanje korporacij, kar pa seveda ni realno, druga možnost pa je regulacija, ki bi razbila monopol tehnoloških gigantov in družbo tako zaščitila pred njimi.⁵

2. Izguba zasebnosti

Mediji so pred kratkim poročali, da se bo državam, ki pod pretvezo večje varnosti, uporabljajo nadzorne kamere z vgrajenim sistemom za prepoznavanje obraza, pridružila tudi Indija, katere glavno mesto New Delhi že brez tega velja za eno izmed najbolj nadzorovanih mest na svetu. UI, ki represivnim organom na ta način pomaga pri identifikaciji ljudi, ni novost, že lep čas je v uporabi v številnih

⁵ Kitajska je regulacijo izvedla tako, da je pred časom postavila velik požarni zid, ki je omenjenim podjetjem preprečil zbiranje podatkov s področja njihove industrije (Grobelnik, 2018).

evropskih državah: Veliki Britaniji, Nemčiji, Franciji, Danski, Švedski, Nizozemski, Italiji, Srbiji in še kje. Očitno postajajo algoritmi, ki pomagajo policiji pri preprečevanju kriminala in iskanju krivcev ter tako povečujejo njeno učinkovitost, del našega vsakdanjika ne glede na to, da namestitev nadzornih kamer, prepoznavanje obrazov s slik, ki jih naredijo, in hitra identifikacija oseb, ki so na njih, mobilni telefoni, ki stalno sporočajo lokacijo imetnika itn., bistveno zmanjšujejo našo zasebnost in odpirajo neskončne možnosti zlorab. Na primer iz vsega tega je mogoče rekonstruirati dnevno rutino neke osebe in odstopanje od nje: kje je doma, kje je v službi, kaj so njeni hobiji, kaj so njeni interesi, kaj je njena priljubljena hrana itd. Zdi se, da bi se lahko v bližnji prihodnosti uresničila nočna mora vseh tistih, ki že dolgo opozarjajo na to, da bo razvoj novih tehnologij v slogu 1984, kjer te veliki brat opazuje, ljudi povsem prikrajšal za svobodo.

Opozoriti velja vsaj na tri probleme, ki se pojavljajo pri uporabi sistemov za prepoznavanje obraza. Prvič, v praksi so se izkazali za neučinkovite, saj je odstotek napačnih identifikacij še vedno visok: 0.3 % zveni malo, vendar na mestu, kjer je pretok ljudi velik, na primer letališča, to lahko pomeni več sto krivičnih zaslišanj ali aretacij (Rajšek, 2019). Drugič, takšni algoritmi imajo težave s prepoznavanjem ljudi, ki so temnopolti, pripadajo etničnim manjšinam ali pa so spremenili spol, saj so v fazi urjenja v svojo podatkovno bazo shranili največ obrazov belopoltnih ljudi,⁶ pri čemer gre krivdo pripisati njihovim učiteljem, to je programerjem. Tretjič, kako je mogoče, da so se po Evropi tako razpasli, ko pa je zbiranje biometričnih podatkov, med katere poleg značilnosti obrazov sodijo tudi druge fiziološke in vedenjske značilnosti posameznika, v Evropski uniji prepovedano s Splošno uredbo o varstvu osebnih podatkov (SUV)⁷ (Rajšek, 2019).

Težava je v tem, da država, ki želi namestiti takšno tehnologijo v svojih mestih, s ponudnikom takšnih storitev, sklene pogodbo, ki ostane tajna. Oba seveda trdita, da biometrične podatke izbrišeta, ne vemo pa, kaj se z njimi resnično dogaja, saj niti enim niti drugim ne gre zaupati. Razen njune besede nimamo nobenega drugega zagotovila, da jih ne shranjujeta v le njima znanih podatkovnih bazah. Če to sedaj povežemo z UI, ki je do te mere sposobna posnemati različne glasove (tudi slavnih osebnosti), da jih ne ločimo od pravih in ki je pri obdelavi naravnega jezika že tako napredovala, da se ne razlikuje več od človeškega osebnega asistenta, potem imamo resen problem. Ko pokliče mamó in ji čestita za rojstni dan, čeprav ga nima, je to še najmanj, če pa ji ukažemo, da naj z glasom Primoža Rogliča pokliče sto oseb in jih prosi za donacijo 50 evrov v dobrotelne namene, nakaže pa naj jih na naš transakcijski račun, stvar ni več tako nedolžna (Kremp, 2018). Ali se ne bi morali takšni sistemi UI, ko nekoga pokličejo, najprej predstaviti, da ljudje ne bi bili zavedeni? Nekateri avtorji (Welsh, 2016) so to predlagali že pred leti,

⁶ To aktivisti za človekove pravice imenujejo *tehnološki rasizem*.

⁷ Angl. General Data Protection Regulation (GDPR).

tako bi se namreč zlorabam v primeru, ko virtualni pomočnik pri telefoniranju posnema glas nekoga, ki ga poznamo, izognili. Samo upamo lahko, da bo nova zakonodaja za etično uporabo UI, ki jo pripravlja Evropska komisija tako dovršena, da države, ki bi rade postavile nadzorne kamere z vgrajenim sistemom za prepoznavanje obraza, ne bodo našle lukenj, ki jim jo bodo omogočale zaobiti.

Vse to pa ni nič proti temu, kar pod krinko zagotavljanja varnosti državljanov z različnimi pilotnimi projekti nadzora, katerih bistvo je nenehno zbiranje vseh dostopnih informacij o posameznikih, počne Kitajska. Najstrožji je tisti v provinci Sinkiang,⁸ gre pa za eno izmed različic prihajajočega sistema spremljanja in ocenjevanja državljanov, ki ljudi glede na vse zbrane podatke, to je, kje delajo, kaj govorijo, kam hodijo, s kom se družijo, kaj objavljajo itn., razvršča v tri skupine: zaupanja vredne, povprečne in zaupanja nevredne. Skladno s tem so eni nagrajeni, drugi pa kaznovani, to pa je mogoče le, če informacijska tehnologija razdrobljene baze podatkov, pridobljene s sistemi za prepoznavanje obraza, zgradbe telesa in načina hoje (iz tega dobi oblast tudi podatke o spolu, starosti, rasi itn.), poveže v enovite profile fizičnih in pravnih oseb. Na primer, če uporabnik mobilnega telefona prečka cesto poleg prehoda za pešce, ga aplikacija, ki jo mora imeti naloženo na telefonu, da država vidi, kaj dela, o tem prekršku obvesti in mu s transakcijskega računa odtegne denar za plačilo kazni: algoritem zazna, obsodi, razsodi in kaznuje. Ljudje s povezavo na splet v bistvu državi sami izdajajo vse skrivnosti, ki so jih včasih morale mukoma odkrivati agenti obveščevalnih služb. Seveda pa ni cilj takšnega množičnega zbiranja informacij zgolj nadzor početja, ampak tudi njegovo napovedovanje. UI namreč omogoča prepoznavanje in razvrščanje vzorcev, iz njih pa je mogoče predvideti kup stvari: posameznikovo mnenje, s kom se bo družil, koga bo volil, kaj bo kupil ali kje bo ob določeni uri dneva. Očitno je, da gre v državi z več kot 800 milijonov uporabnikov spleta, kar je daleč največ na svetu, splet je sicer v državni lasti, saj daje pasovno širino v najem ona, s čimer lahko gleda, kaj se po žicah in valovih pretaka, za državni projekt, njegov cilj pa ustvariti enotno mrežo različnih povezanih baz podatkov o ljudeh in podjetjih, da bi lahko nadzorovala in usmerjala njihovo obnašanje. Omenjen sistem bo po letu 2020 obvezen za vse Kitajce, trenutno jih je 1.4 milijarde.⁹ Kitajska je tako dober primer države, ki je UI zlorabila za postavitev vseprisotnega, največjega in najučinkovitejšega nadzorovalnega aparata v človeški zgodovini (Masten, 2019b).

⁸ Gre za največjo kitajsko provinco, ki meri kar 1.6 milijona km² in leži na severozahodu države, znana pa je po zatiranju muslimanske manjšine Ujgurov.

⁹ Kitajska pa uvaja tudi tako imenovane bralnike možganskih valov. Gre za opremo, ki meri človekovo možgansko aktivnost, iz nje pa je mogoče razbrati, v kakšnem čustvenem stanju se človek nahaja. Nositi jo morajo številni pripadniki vojske, nekateri delavci v proizvodnih enotah in v transportnem sektorju ter še kdo, omogoča pa, da je tudi naša notranjost pod neprestanim nadzorom delodajalca in s tem države (Masten, 2019a, 2019b).

3. Inteligentno orožje

Velika nevarnost, ki zahteva moralni preudarek, je izdelava UI v vojaške oziroma vojne namene. Medtem ko so deloma avtonomni sistemi orožja že dolgo v uporabi, na primer, troti, s katerimi je mogoče na daljavo in brez lastnih žrtev uničevati sovražnikove cilje, pa povsem avtomatizirano orožje šele prihaja.¹⁰ V sicer zgodnji fazi izdelave so robotski sistemi za samostojno delovanje na bojišču, nekaj takega, kar smo lahko do sedaj opazovali zgolj v takšnih in drugačnih znanstveno-fantastičnih filmih. Razvoj UI gre v smeri ustvarjanja sistemov, ki bodo sposobni brez neposrednega človekovega ukaza ali nadzora ubijati ljudi. Njihova prednost v primerjavi z na primer jedrskim orožjem je v tem, da zanje ne rabimo nobenih redkih elementov, izdelati jih je možno zelo enostavno in to iz komponent, ki jih lahko deloma že danes kupimo preko spleta (Stöcker, 2015). To je kot naročeno tudi za teroriste, ki lahko takšno UI uporabljajo za podporo pri svojih napadih, v splošnem pa se lahko tako razvijejo modeli, ki majhnim skupinam pomagajo, da z nizkimi stroški povzročijo ogromno škodo. Ker bodo hitro zreli za masovno proizvodnjo, je samo vprašanje časa, kdaj se bodo pojavili tudi na črnem trgu in v rokah diktatorjev, oligarhov in generalov, z namenom boljšega nadziranja prebivalstva, zatiranja miroljubnih protestov ali etničnega čiščenja. Dejstvo je, da UI, če se bo še dalje razvijala brez organizacijskega, varnostnega in tehničnega nadzora, predstavlja grožnjo.

Zato je leta 2015 skupina vodilnih znanstvenikov s področja UI in robotike napisala odprto pismo z naslovom *Autonomous weapons: an open letter from ai & robotics researchers*, v katerem opozarjajo na nevarnost razvoja avtonomnega orožja, s svojimi podpisi pa so se izrekli tudi za njegovo prepoved izdelave po vzoru prepovedi uporabe kemičnega in biološkega orožja s strani kemikov ter v veselju postavljenega jedrskega in laserskega orožja s strani fizikov. Poudarjajo, da je nadomeščanje vojakov z roboti sicer dobro z vidika zmanjševanja žrtev, a slabo z vidika določanja praga, kdaj se za vojno odločimo, ker se ta zniža. S tem, ko avtonomno orožje imenujejo »jutrišnje kalašnikovke«, slikovito ponazarjajo, kam nas lahko globalno povpraševanje po njem pripelje. Opozarjajo, da je idealno za atentate, destabilizacijo držav, podrejanje prebivalstva in izbrisane posameznih etničnih skupin. Zaradi tega menijo, da z razvojem sistemov UI v vojaške namene človeštvo ne bi ničesar pridobilo, ampak izgubilo.

Resnici na ljubo to ni prvi poskus vplivati na regulacijo področja razvoja inteligentnega orožja, poročilo *Human Rights Watch*-a je že leta 2012 opozarjalo na njegovo nevarnost, res pa je, da je seznam podpisnikov nove peticije dolg, imena pa ugledna, zaradi česar lahko samo upamo, da bodo njihovi pozivi tokrat padli na plodna tla.

¹⁰ Govori se o tretji revoluciji v vojskovanju, takoj za iznajdbo smodnika in jedrskega orožja.

4. Izguba služb

C. Mims, reporter *Tech-Blogs-Quartz*, je imel svoje prvo srečanje z (delovnim) robotom na zabavi MIT-ja. Stal je v kotu in baje izgledal osamljeno, dokler se mu Mims ni približal, vzel njegovo roko in jo potopil v škatlo z majhnimi predmeti. *Baxter*, kot mu je bilo ime, je enega izmed njih zagrabil, potem pa ga je Mims odpeljal do mize, kjer ga je robot odložil. Nadaljeval je *Baxter* sam, vneto je hodil do škatle in iz nje vlekel predmete ter jih nosil na mizo. Ni ga motilo, da so bili različnih oblik in velikosti, niti da so bili gibi, ki mu jih je pokazal Mims in iz katerih se je v nekaj sekundah naučil, kako to narediti, vse prej kot nazorni. Zbrano je opravljal svoje delo, delo, ki ga sicer opravlja Janez Novak v bližnjem Amazonovem paketno distribucijskem centru (Schulz, S., 2016).

Pred nečim podobnim je svaril že Weizenbaum (1976), to je pred nadomeščanjem ljudi z UI, kar na dolgi rok vpliva na spremembo področja zaposlovanja. Tehnološki giganti kot *Apple*, *Microsoft*, *Google* itd., si prizadevajo za to, da bi v prihodnosti čim več delovnih postopkov postalo avtomatiziranih, to je, da bi opravljanje dela, ki je sedaj v domeni ljudi, postopoma v celoti in na vseh področjih gospodarstva prevzeli stroji. Ker bo vedno več opravil samodejnih, ljudje ne bodo več imeli služb, zaradi česar bo za njihovo preživetje potrebno poskrbeti drugače (prekvalifikacija, UTD, nadomestila ipd.). Na primer, avtonomna vozila bodo odvezla službe taksistom, šoferjem avtobusov in tovornjakov, strojevodjem itn. S tem bo človek, vsaj tako se zdi, postal odvečna oziroma presežna delovna sila. Povečalo se bo število ljudi, ki živijo pod pragom revščine, kar bo vodilo v socialne nemire po vzoru, recimo, protesta »rumenih jopičev«, ki so se leta 2018 začeli v Parizu. Skratka, družba bo postala še bolj nezadovoljna in razslojena, kot je sedaj. Ali je ta strah upravičen?

Dejstvo je, da bo naslednji val avtomatizacije in digitalizacije trg dela pretresel do temeljev in to prej, kot si mislimo. Ni sicer prvič, da se bo kaj takega zgodilo, res pa je, da so vse spremembe do sedaj sledile nekemu naravnemu zakonu: stroji so prevzeli umazane in zdravju škodljive službe, s tem pa ustvarili nove, bolj zdrave in kreativne za ljudi, ki so na začetku običajno pomenile nadzor, vzdrževanje in popravilo avtomatiziranih sistemov. Tokrat naj bi bilo drugače, ne bo šlo zgolj za zamenjavo fizičnega z intelektualnim delom, ampak tudi za nadomeščanje ljudi s stroji v službah, ki zahtevajo ustvarjalnost in načrtovanje. Zdi se, da so na tapeti skoraj vsi poklici, študija Univerze v Oxfordu pa kaže, da je samo v ZDA v nevarnosti 47% služb in to na vseh področjih: od kmetijstva, preko industrije do storitev (Benedikt, Osborne, 2013). Deutsche Bank, na primer, je v okviru svojega varčevalnega programa napovedala, da naj bi do leta 2022 službo izgubilo 18.000 zaposlenih. Zasluga za to gre uporabi UI, ki je v določenih sektorjih poslovanja izjemno povečala njeno produktivnost. Deutsche Bank sploh ne skriva, da bo v prihodnosti še odpuščala, njen cilj je zmanjšati stroške, povečati prihodke in, zanimivo, s pomočjo uporabe UI celo izboljšati uporabniško izkušnjo strank.

Kakorkoli, vsi strokovnjaki ne mislijo, da je prihodnost področja zaposlovanja zaradi uporabe UI tako črna. Nastali naj bi novi poklici, pri tistih, ki se omenjajo kot primeri, pa fantazija očitno ne pozna meja: strokovnjak proti staranju, pisatelj v virtualni realnosti, urbani kmet, psiholog domačih živali, da o vseh vrstah računalničarjev in programerjev ter vzdrževalcev in čistilcev robotov niti ne govorimo. Cilj naj bi bil ljudi osvoboditi rutine in jim tako pomagati, da se lahko posvetijo bolj kreativnim in produktivnim nalogam (Schulz, S., 2016).

Kljub temu vse ni tako rožnato. Število zaposlitev za nedoločen čas naj bi se zmanjšalo, skoraj nihče več ne bo istega dela opravljal celo življenje, služba pa bo odvisna od projektov, na katerih bodo delale vedno različne skupine ljudi. Pospešena avtomatizacija za delavca pomeni, da se mora spremeniti, postati mora bolj ustvarjalen, prilagodljiv in poln idej, z drugimi besedami, neprestano mora dokazovati, da še ni prišel čas za njegovo zamenjavo z UI (Schulz, T., 2016). Drugi izvedenci so še bolj pesimistični, menijo, da bo število ustvarjenih delovnih mest manjše od števila izgubljenih, prekvalifikacije pa drage in da se sčasoma delavcev ne bo več izplačalo izobraževati, da bi lahko držali korak s stroji, ampak jih bo smotrnejše odpustiti.

Ima pa seveda revolucija trga zaposlovanja, ki je pred nami, če ironiziramo, eno prednost: ker bodo sčasoma nevronske mreže postale tako močne, da sploh ne bomo več rabili delati, bomo imeli več časa zase. Država pa bo z univerzalnim temeljnim dohodkom (UTD) poskrbela za to, da bo vsak izmed nas imel dostojno življenje in se lahko posvetil razmišljanju o tem, zakaj je sploh na svetu, ali pa si bo za to, da se bo dokopal do tega odgovora, sestavil računalnik (Schulz, S., 2016).

Kakorkoli, že Joseph Weizenbaum je verjetno zaradi opisane groze, ki jo je doživel z *Elizo*, leta 1976 navedel nekaj poklicev, v katerih ljudi v nobenem primeru ne bi smeli nadomestiti računalniki: policaj, sodnik, vojak, socialni delavec, terapevt in delavec v podporni službi strankam (Weizenbaum, 1976). Še več, takšna uporaba UI naj bi ogrozila človeško dostojanstvo, saj omenjene službe od zaposlenih zahtevajo izkazovanje in posedovanje verodostojnega občutka empatije, to je sočutja. Njegovo opozorilo je bilo žal preslišano, hiter razvoj UI z vedno bolj sofisticiranimi proizvodi, je na moralna vprašanja glede njihove izdelave in uporabe gladko pozabil.

Kljub temu se zdi, da smo še vedno v fazi, ko stroj človeka sicer lahko preseže pri marsikaterem opravilu, ga tudi pretenta ali prevzame del njegovega posla, nikakor pa ga še ne more nadomestiti. Upamo lahko samo, da bo tudi nadaljnji razvoj UI potekal postopoma, in sicer preko vedno tesnejšega sodelovanja človeka in stroja. To nam bi pomagalo, da postanemo učinkovitejši, bolj prilagodljivi in iznajdljivi ter se tako bolje pripravimo na to, kaj nam bo prihodnost z vidika služb resnično prinesla.

5. Obremenitev okolja

Za svoj skokovit razvoj v zadnjih letih se ima UI zahvaliti predvsem eni stvari: tako imenovanemu globokemu učenju, ki smo ga že omenili, ne pomeni pa nič drugega kot to, da se stroji sami učijo, kar jim omogoča, da postajajo vse pametnejši in zmogljivejši. Ideja je, poenostavljeno rečeno, ustvarjanje večplastnih mrež iz umetnih nevronov in kopiranje delovanja človeških možganov z njimi. Problem je, da živimo v času, ko je skrb za okolje postala osrednja politična tema, zaščita podnebja pa že zdaj naša glavna naloga, saj trenutni hiter razvoj UI niti slučajno ni skladen z zeleno okoljsko politiko.

Zadnja študija Univerze v Amherst Masschusetssu je pokazala, da je urjenje nevronske mreže drago z dveh vidikov: prvič, finančno, zaradi cene elektrike, ki je za to potrebna, in drugič, okoljsko, zaradi ogljičnega odtisa, ki ga proizvede delovanje osrednje procesne enote (Strubell, Ganesh, McCallum, 2019). V njej so avtorji preučevali štiri modele UI, ki temeljijo na nevronske mrežah in so sposobni obdelovati naravni jezik. Gre za sisteme, ki na takšen ali drugačen način analizirajo jezik, natančneje, algoritme, ki se uporabljajo kot virtualni asistenti, prevajalniki, pisci zgodb itn. Da bi izmerili porabo energije njihovih procesorjev in grafike, so štiri takšne modele izmenično urili en dan. Iz količine energije, ki so jo modeli porabili za to, da so opravili posamezne dele naloge, so izračunali, koliko elektrike bodo porabili do konca treninga, potem pa to pretvorili v količino CO₂, ki ga bodo pri tem izpustili v ozračje. Rezultati so pomenljivi: urjenje zgolj enega takšnega sistema proizvede 313 ton CO₂, kar je približno petkrat toliko kot avto v svoji celotni življenjski dobi. Potreba po ogromni količini energije je posledica velike računske moči stroja, brez katere ne bi mogel obdelati ogromne količine podatkov, kar pa je za uspešen trening, kot smo že omenili, ključnega pomena.

Da so računalniki energetsko potratni, ni nič novega, že dolgo je znano, da eno iskanje na na primer Googlu proizvede toliko CO₂ kot 2 grelnika vode za čaj, to je 15 gramov. Kriva je elektrika, ki jo rabimo za napajanje računalnika in za pošiljanje zahteve na strežnike po svetu. Pred leti je IBM-ov superračunalnik v ugankarskem kvizu sicer premagal človeška tekmeca, vendar je pri tem porabil 85 kW energije; za primerjavo, človeški možgani pri sodelovanju v takšnem šovu porabijo zgolj 20 W. In še en podatek, že danes porabijo računski/podatkovni centri 2 % vse elektrike na svetu (cel IT sektor pa 7 %) in proizvedejo toliko CO₂ kot cela letalska industrija skupaj. Vprašanje je torej, kako trajnostno sploh je strojno učenje? (Lobe, 2019)

Obstaja pa še en pomemben vidik, ki je povezan z ogromno porabo energije: od kod sploh izvira elektrika, ki jo spletni giganti porabijo za svoje delovanje? Glede na zadnje poročilo Greenpeaca se pri tem med znanimi najslabše odrežejo *Amazon*, *Microsoft* in *Netflix*, ki več kot tretjino svoje energije pridobijo iz premoga,

najbolje pa *Apple*, *Facebook* in *Google*, ki je več kot polovico pridobijo iz obnovljivih virov¹¹ (Cook et al, 2017). Slednji so tudi na vrhu pri vključevanju zavez o uporabi obnovljivih virov energije v svojo politiko, sledljivosti njene nabave in energetske učinkovitosti. O kitajskih ponudnikih spletnih storitev, kot so *Alibaba*, *Baidu*, *Tecent* in ostali, nima smisla razpravljati, bolj ali manj vsi pokrijejo dve tretjini svojih potreb po energiji z elektriko pridobljeno iz premoga, prav tako pa capljajo daleč zadaj glede posredovanja podatkov o svojem energetskem odtisu, sledljivosti uporabljene zelene energije in vključevanju trajnostnih zavez v politiko svojega delovanja.

Glede na to, da »je internet centralni živčni sistem sodobnega globalnega gospodarstva« (Cook et al, 2017: 5), ni čudno, da je tudi izredno velik porabnik energije. Izdelava in napajanje naših domačih naprav, podatkovnih centrov in strežnikov vseh vrst, če naštejemo zgolj najbolj očitne, terjata svoj davek. In prihodnost ni svetla: promet na spletu naj bi se do leta 2020 povečal za trikrat, število njegovih uporabnikov pa naj bi naraslo na 4 milijarde. Očitno je, kaj to pomeni. Zato je ključno naslednje vprašanje: ali bomo pri pokrivanju energetskih potreb IT sektorja (in na sploh) sposobni narediti prehod iz fosilnih goriv na obnovljive vire in se tako izogniti podnebnim spremembam?

V nekem smislu bi lahko naša odvisnost od spleta to celo pospešila, ampak zgolj v primeru, (i) če bi kot njegovi uporabniki tehnološka podjetja znali prisiliti v to, na primer z izogibanjem nakupovanju njihovih »umazanih« izdelkov, da bi njihovo digitalno infrastrukturo poganjala zgolj čista elektrika, (ii) če bi tudi ponudniki energije to prepoznali kot priložnost in zagotovili ugodne fiksne cene za elektriko, pridobljeno iz obnovljivih virov in (iii) če bi od tehnoloških družb zahtevali, da gradijo svoje ime na zavezanosti k uporabi obnovljivih virov energije in transparentnem posredovanju informacij o njeni nabavi ter zavedanju, da bodo tako prispevali k zaščiti podnebja (Cook et al, 2019). Omenjena študija Univerze Amherst v Massachusetsu pa vidi rešitev glede izjemne potratnosti digitalnih tehnologij v optimizaciji modelov UI, to je v razvoju energetsko bolj učinkovitega hardvera (računalnikov/strojev) kot tudi softvera (algoritmov/programov) (Strubell, Ganesh, McCallum, 2019). To naj bi prispevalo k bistveno varčnejšemu urjenju nevronske mreže, ki bo zaradi kasnejše masovne uporabe takšnih modelov UI, na primer različni asistenčni programi v avtomobilih s sposobnostjo preprečevanja nesreč, v resnici še cenejše.

¹¹ *Apple* pridobi iz obnovljivih virov celo 83 % vse potrebne energije.

6. Sklep

Ali se razvijalci, ponudniki in uporabniki UI danes držijo vsaj minimalnih etičnih standardov, ki jih je kot smernice za zaupanja vredno UI predlagala Evropska komisija? Res je, da gre pri tem za političen projekt, pa vendarle. Pri njihovi pripravi so sodelovali posamezniki iz akademske sfere, industrije in civilne družbe, zaradi česar imajo kljub vsemu določeno težo. V članku s pomočjo analize petih področij človekovega življenja ugotovimo, da je nebrzdana in nepremišljena uporaba UI v konfliktu s človekovimi pravicami in družbenimi vrednotami.

Prvo nevarnost predstavlja skoncentriranost podatkov, ki največjim petim tehnološkim podjetjem na svetu, *Googlu*, *Applu*, *Facebooku*, *Amazonu* in *Microsoftu* (GAFAM), omogoča prevlado na informacijskem trgu, zaradi česar je strah pred negativnimi posledicami, ki bi jih njihovo filtriranje podatkov o svetu lahko imelo na družbo, kot kaže primer *Cambridge Analytica*, povsem na mestu. Druga nevarnost je nameščanje kamer s sposobnostjo takšnega in drugačnega prepoznavanja osebnih značilnosti človeka, ki bistveno zmanjšujejo našo zasebnost in svobodo, saj je mogoče iz tako zbranih podatkov do potankosti spoznati rutino človeka in odstopanja od nje ter skladno s tem prilagoditi ciljno ponudbo zanj na spletu celo do te mere, da se bo spremenil glede na naše želje, potrebe in interese. Tretja nevarnost je razvoj inteligentnih vojaških sistemov, ki so sposobni brez neposrednega človekovega ukaza ali nadzora ubijati ljudi brez lastnih žrtev, kar vstop v vojno samo še olajša. Njihova izdelava ne predstavlja praktično nobenega finančnega bremena in je v rokah avtokratov kot naročena za discipliniranje državljanov, zatiranje demonstracij ali etnično čiščenje. Četrto nevarnost predstavlja izguba služb, kar je posledica tega, da stroji že zdaj v vedno večji meri opravljajo delo, ki je bilo še pred kratkim povsem v domeni ljudi. Človek vse bolj postaja presežna delovna sila, ki lahko preživi le, če se neprestano prilagaja, dela pod prekarimi pogoji in dokazuje, da še ni zrel za zamenjavo. Vprašanje je, kako dolgo še? Peta nevarnost pa je izjemno velika poraba energija za delovanje UI na globalni ravni in temu primeren slab ogljični odtis, kar ni skladno z današnjimi trendi na področju varovanja okolja. Zaskrbljujoče je predvsem to, da v večini primerov še vedno preveč elektrike za izdelavo in pogon naših pametih naprav pridobimo iz umazanih fosilnih goriv, namesto da bi pri tem prevladovala energija proizvedena iz obnovljivih virov.

Kaj lahko glede tega storimo? Glede skoncentriranosti podatkov smo v težkem položaju, podružabljanje korporacij ni realna možnost, regulacija digitalnih gigantov pa se lahko kot v primeru Kitajske hitro sprevrže v digitalno cenzuro in diktaturo države. Pri izgubi zasebnosti bi lahko pomagala domišljena in dovršena zakonodaja brez pravnih praznin, ki bi onemogočale, da bi jo države zaobšle. Seveda bi morala veljati za vse, kar pa je slejkoprej iluzija, če jo bomo dobili v EU, bo to že velik dosežek. Tudi v zvezi z izdelavo avtomatiziranega orožja so naši izgledi slabi. Težava je, da države financirajo njegov razvoj, podjetja, ki se s tem

ukvarjajo, pa denar velikodušno in brez zadržkov sprejemajo. Ko je izdelek narejen, so financerji tudi njegovi kupci, zaradi česar je uspešnost prizadevanj za prepoved izdelave in uporabe samodejnega orožja pod velikim vprašajem. Glede revolucije na trgu delovne sile med nekaterimi strokovnjaki prevladuje prepričanje, da bo nastalo več služb, kot pa jih bo izginilo. Pa ne samo to, novi poklici bodo ljudem omogočili večjo ustvarjalnost in učinkovitost ter jih tako osvobodili rutine in zasičenosti. Ker so stroji sposobni prevzeti tudi takšne službe, je opisan idealen scenarij po našem mnenju možen zgolj v enem primeru: nevronske mreže so tako močne, da ljudem ni več treba delati, država pa vsakemu izmed nas nakažuje UTD kot plačilo za dostojno življenje. Ali je to uresničljivo, bo pokazal čas. Pri manjši porabi energije pa bi lahko pomagal pritisk na tehnološka podjetja, da mora njihovo digitalno infrastrukturo poganjati zgolj energija, pridobljena iz obnovljivih virov. To pa zahteva dovolj veliko število ozaveščenih uporabnikov, ki si lahko na takšen način narejene izdelke, običajno dražje od konkurence, ki jaha na premogu, tudi privoščijo. Verjetno je bolj realna rešitev za zelen IT sektor na globalni ravni v tem trenutku optimizacija urjenja nevronske mreže s pomočjo varčnejšega hardvera in softvera.

Kakorkoli, svet pametnih strojev ni več literarna fikcija. UI naše življenje spreminja, mi pa temu, tako se zdi, v moralnem smislu, ne posvečamo dovolj pozornosti. Razvoj inteligentnih strojev je hiter, glede tega ne moremo storiti nič. Mogoče ga lahko usmerjamo, ampak potem moramo s tem začeti zdaj.

Artificial Intelligence at Work: Some Ethical Concerns

A recent development of artificial intelligence (AI) and its self-evident use in our everyday life have increased interest of the public and experts for all kinds of artificially made autonomous systems. Since AI is not anymore understood merely as a handy tool but also as an independent agent, the question of its impact on our life in terms of morality emerges naturally. Answers to questions such as »Is AI responsible for its actions?«, »What are social, legal and cultural consequences of its decisions?«, »Should its autonomous models be freely accessible in the market?« etc., determine to a great extent whether AI will be trustworthy in the future or not. The paper introduces five threats associated with our unbridled use of AI and concludes through the analyses of current developments in the fields related to those threats that the society is increasingly becoming a hostage of the AI industry: intelligent systems are not made in a way that they would, while interacting with people, in any of their evolution stages (design, construction, installation and evaluation) respect human rights nor act according to accepted social values.

Keywords: artificial intelligence, machine learning, natural language processing, ethics, human rights.

Literatura

Bregant, J. (2019). »Umetna inteligenca v praksi (1. del): razvoj, obnašanje in učenje strojev«. *Analiza*, 2, str. 39–55.

Autonomous weapons: an open letter from ai & robotics researchers (2015). Dostopno na: <https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1> [29.11.2019].

Benedikt, C., Osborne, M. A. (2013). »The future of employment: how susceptible are jobs to computerisation?«. Dostopno na: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf [30.11.2019].

Cook et al (2017). »CliCking Clean: Who is Winning the RaCe to Build a gReen inteRnet?«.

Dengel, A. (2019c). »Multimedia-Data-Mining: Trends und Emotionen in Big Data erkennen«. V Dengel, A., Socher, R., Kirchner E. A., Ogolla, S., Künstliche Intelligenz: Die Zukunft von Mensch und Maschine. Hamburg: ZEIT Akademie GmbH, str. 74–84.

European Commission's High-Level Expert Group on Artificial Intelligence (2018). *Ethics guidelines for trustworthy AI*.

Grobelnik, M. (2018). »Podatki so nova nafta. Kdor ima dostop do podatkov, lahko rešuje probleme«. *Mladina*, 39.

Human Rights Watch (2012). Losing humanity: The Case against Killer Robots. Dostopno na: https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf [29.11.2019].

Kremp, M. (2018). »Google Duplex ist gruselig gut«. Spiegel Online. Dostopno na: <https://www.spiegel.de/netzwelt/web/google-duplex-auf-der-i-o-gruselig-gute-kuenstliche-intelligenz-a-1206938.html> [28.11.2019].

Lobe, A. (2019). »KI ist alles andere als grün«. Spektrum.de. Dostopno na: <https://www.spektrum.de/news/kuenstliche-intelligenz-verbraucht-fuer-den-lernprozess-unvorstellbar-viel-energie/1660246> [01.12.2019].

Masten, A. (2019a). »Kaj vse pomeni klik na 'Strinjam se': O ekonomiji podatkov in monopolih«. MMC RTV SLO. Dostopno na: <https://www.rtvsllo.si/mmc-podrobno/na-pragu-digitalne-diktature-brez-zasebnosti-in-brez-svobode/489378> [02.12.2019].

Masten, A. (2019b). »Ko nadzor prevlada nad svobodo: Veliki kitajski brat: prva digitalna diktatura«. MMC RTV SLO. Dostopno na: <https://www.rtv slo.si/mmc-podrobno/na-pragu-digitalne-diktature-brez-zasebnosti-in-brez-svobode/489378> [02.12.2019].

Ogola, S. (2019). »Verantwortung, Erklärbarkeit und Transparenzalgorithmischer Entscheidungen«. V Dengel, A., Socher, R., Kirchner E. A., Ogolla, S., *Künstliche Intelligenz: Die Zukunft von Mensch und Maschine*. Hamburg: ZEIT Akademie GmbH, str. 93–101.

Rajšek, J. (2019). »Veliki evropski brat se prebuja – 'Obraza ne moremo zamenjati'«. MMC RTV SLO. Dostopno na: <https://www.rtv slo.si/evropska-unija/veliki-evropski-brat-se-prebuja-obraza-ne-moremo-zamenjati/502913> [02.12.2019].

Schulz, S. (2016). »Die Jobfresser kommen«. Spiegel Online. Dostopno na: <https://www.spiegel.de/wirtschaft/soziales/arbeitsmarkt-der-zukunft-die-jobfresser-kommen-a-1105032.html> [30.11.2019].

Schulz, T. (2016). »Wie Kollege Computer Ihren Job überflüssig macht«. Spiegel Online. Dostopno na: <https://www.spiegel.de/wirtschaft/kuenstliche-intelligenz-wie-kollege-computer-ihren-job-ueberfluessig-macht-a-1110661.html> [30.11.2019].

Stöcker, C. (2015). »Künstliche-Intelligenz-Forscher warnen vor künstlicher Intelligenz«. Spiegel Online. Dostopno na: <https://www.spiegel.de/netzwelt/netzpolitik/elon-musk-und-stephen-hawking-warnen-vor-autonomen-waffen-a-1045615.html> [28.11.2019].

Strubell, E., Ganesh, A., McCallum, A. (2019). »Energy and Policy Considerations for Deep Learning in NLP«. *Študija Univerze v Massachusettsu Amherst*. Dostopno na: <https://arxiv.org/pdf/1906.02243.pdf> [01.12.2019].

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. New York: PublicAffairs.

Welsh, T. (2016). »Turing's Red Flag«. *Communications of the ACM*, 7, str. 34–37. Dostopno na: <https://www.cse.unsw.edu.au/~tw/wcacm15.pdf> [28.11.2019].

Weizenbaum, J. (1976). *Computer Power and Human Reason*. San Francisco: W. H. Freeman.