Tomislav Bracanović, *Etika umjetne inteligencije* (Zagreb: Institut za filozofiju, 2022), 223 str.

U monografiji *Etika umjetne inteligencije* Tomislav Bracanović odgovara na izazove vremena te nudi argumentiranu raspravu koja će zainteresiranom čitatelju biti od velike pomoći pri formuliranju vlastitih stajališta glede izazova koje budućnost pred nas stavlja "[...] uslijed sve bržeg prodora umjetne inteligencije u sve brojnije slojeve naše individualne i društvene egzistencije" (198). Koliko je taj "prodor" ozbiljan najbolje ilustrira otvoreno pismo Pause Giant AI Experiments[8] objavljeno 22. ožujka ove godine kojega je supotpisnik, između ostalih lidera u industriji, Elon Musk. U pismu se poziva na obustavljanje ulaganja u moćne sustave umjetne inteligencije na šest mjeseci. Potpisnici se zalažu da "[s]nažne AI sustave treba razvijati tek kada budemo sigurni da će njihovi učinci biti pozitivni a njihovi [...] rizici [...] podložni kontroli."[9]

Bracanovićeva monografija, vjerojatno jer je pisana prije nego što je ChatGPT pušten u javnost i prije nego što je upućen poziv na zaustavljanje ulaganja u razvoj moćne umjetne inteligencije na šest mjeseci, ima manje dramatičan ton. To nije nužno nedostatak knjige, no zasigurno upućuje na nevjerojatnu brzinu kojom se umjetna inteligencija razvija te na potrebu stalnog ažuriranja s obzirom na protok novih informacija.

Bracanovićev pristup je, dakle, konzervativniji. To ima i svojih prednosti jer izbjegava zamku u koju mnogi upadaju kada pišu o umjetnoj inteligenciji, a to je maštanje i špekuliranje o tome u kakvim bi se sve fantastičnim oblicima umjetna inteligencija u budućnosti mogla pojaviti i koliko bi ljudska vrsta bila nepripremljena na egzistencijalne rizike koje bi takav tehnološki razvitak stavio pred nas. Takve rasprave uglavnom završavaju s nekakvim apokaliptičnim epilozima koji nam time čitavu tematiku čine stranom i dalekom.

U *Etici umjetne inteligencije* takvog senzacionalizma nema, već autor postupno uvodi čitatelja u problematiku počevši s jasnim definiranjem ključnih pojmova, prije svega pojma "umjetne inteligencije". On može označavati tri različite stvari: "[1] jedno široko istraživačko područje […] [2] konkretne strojeve i računalne programe koji su u tom području

[8] https://futureoflife.org/wp-content/uploads/2023/05/FLI_Pause-Giant-AI-Experiments_An-Open-Letter.pdf

[9] *Ibid.*

osmišljeni […] [i] [3] neke funkcije uređaja kojima se koristimo“ (2). Dakle, pod umjetnu inteligenciju potpadaju i istraživački programi na sveučilištima i institutima, dijagnostički sustavi u medicini, Roomba roboti za usisavanje, virtualni asistenti poput Amazonove Alexe te chatbotovi, među kojima je napoznatiji OpenAI-ev ChatGPT.

Umjetna inteligencija u prvom smislu dijeli se na teorijsku i pragmatičnu. Bracanović prvu definira kao "[...] ono područje umjetne inteligencije koje je usredotočeno prvenstveno [...] na konceptualna ili apstraktna pitanja: Što uopće jest inteligencija? U kakvom odnosu stoje mišljenje i mozak? [...] Je li inteligencija moguća isključivo u biološkim organizmima ili je moguća i u strojevima kao što su računala" (3), dok je zadaća pragmatične inteligencije "[...] konstruiranje strojeva ili komponenti strojeva koji će biti u stanju izvoditi određene zadatke i operacije za koje je inače neophodna ljudska inteligencija" (7) a koje se svode na: obradu prirodnog jezika, reprezentaciju znanja, automatizirano zaključivanje, strojno učenje, računalni vid te robotiku (8).

Nadalje, autor podsjeća na Searlovu (1980) podjelu na "slabu" umjetnu inteligenciju, "[...] koja je u stanju na temelju svog programa simulirati ljudsku inteligenciju i učinkovito obavljati neki poseban zadatak" (4) te "jaku" umjetnu inteligenciju, koja stroj koji je oprimjeruje čini *doista* inteligentnim (5).

Imajući u vidu da je "jaka" umjetna inteligencija još uvijek sporan projekt u čiju realizaciju dobar dio stručnjaka i ne vjeruje da je dostižna, ili barem ne u tako skoro vrijeme, Bracanović se fokusira na obradu pitanja i izazova koji dolaze s razvojem pragmatične inteligencije kroz prizmu različitih etičkih teorija. Dakle, u fokusu knjige su, kao što ćemo kasnije vidjeti, pitanja sasvim praktične naravi: kakve etičke postavke bi naša autonomna vozila trebala imati? Na koji način razvoj socijalnih robota mijenja naše poimanje dobrog života? Ugrožavaju li strojno učenje i njemu svojstveni algoritmi našu privatnost te kako bi stvaranje u svakom smislu nadprosječnih pojedinaca ugrozilo naše poimanje ljudskih prava i posebnosti *homo sapiensa*? To su itekako aktualna pitanja na koja moramo biti spremni dati odgovor prije nego stvari izmaknu kontroli.

No, prije obrade stvarnih problema, autor obrazlaže što filozofska disciplina etike obuhvaća i na koja pitanja nastoji dati odgovor. Tako autor vrlo rano u knjizi pojašnjava da etika može biti općenitije, teoretske naravi te nastojati definirati pojmove poput morala ili vrline, ali i da postoji druga grana etike, primjena koje je u fokusu ove knjige, a to je normativna etika i niz teorija koje ona obuhvaća. Najpoznatiji primjeri

normativne etike, a kojima se autor služi u suočavanju s izazovima razvoja umjetne inteligencije su: deontološka etika, utilitarizam, etički egoizam, etika vrlina i etički relativizam (10). U ovom prikazu neću se osvrtati na prednosti i nedostatke svake od tih teorija,[10] već ću se usredotočiti na vrijednost filozofskog pristupa izazovima umjetne inteligencije koji ova monografija nudi, kao i na nekoliko točaka gdje smatram primjerenim uputiti prigovor te eventualnu dobronamjernu kritiku.

Prije izlaganja specifičnih izazova umjetne inteligencije autor kratko izlaže ono što naziva generičkim etičkim izazovima umjetne inteligencije, a tako ih definira jer "[...] su svojstveni većini njezinih inovacija i primjena, ali jednako tako i nekim drugim tehnologijama koje se temelje na umjetnoj inteligenciji" (16). Prvi generički izazov jest sigurnost, odnosno strah od štete koju bi nekontrolirani razvoj umjetne inteligencije mogao prouzročiti. Oni više oprezni, među kojima su i potpisnici otvorenog pisma Pause Giant AI Experiments, željeli bi da se nastoji unaprijed predvidjeti kako će razvoj umjetne inteligencije utjecati na društvo i da bismo u skladu s takvim predviđanjima trebali upravljati svoje napore u daljnjem tehnološkom razvitku.

Sljedeća problematika tiče se privatnosti, konkretnije, zaštite osobnih podataka, jer je strah mnogih da tehnologija ima preširok i preslobodan pristup podacima koje bi radije zadržali za sebe te koji bi mogli biti korišteni u svrhe koje ne bismo odobrili. Primjerice, "dijagnostički sustavi mogu znati našu krvnu grupu, specifične bolesti koje smo preboljeli i sl." (20), a takvim bi se saznanjima dalje mogle koristiti osiguravajuće kuće u izradi svojih polica osiguranja i odgovarajućih premija ili poslodavci u procesu zapošljavanja potencijalnih kandidata. Jasno je, dakle, kako bi takva saznanja mogla biti zloporabljena.

Bracanović potom razlaže načelo predostrožnosti koje "[...] u osnovi kaže da u primjenu neke inovacije ne treba ulaziti ako se ne može dokazati da ona neće imati nepopravljive štetne posljedice" (22). No, kao što autor ističe, naša predviđanja su ograničena i preveliki oprez i odvagivanje mogu kočiti tehnološki napredak te sve pogodnosti koje s time dolaze. Upravo u ovom kratkom poglavlju autor prvi puta definira pojam singularnosti (iako ne izravno) kada parafrazira Hawkinga:

> Umjetna inteligencija sve je prisutnija u ljudskim životima te, stvorimo li njezinu "jaku" verziju, postoji opasnost da će ona stvoriti još jaču, a ova još jaču i tako sve

[10] Zainteresiranog čitatelja uputio bih na Bracanovićev podrobniji prikaz tih teorija u njegovoj knjizi *Normativna etika* (Zagreb: Institut za filozofiju, 2018).

> do trenutka kada će umjetna inteligencija postati toliko nadmoćna onoj ljudskoj da njezine odluke i postupke – moguće kobne po ljude i čitavo čovječanstvo – više nećemo moći ni razumjeti ni zaustaviti. (25)

Bracanović izražava skepticizam prema takvim strepnjama – jer "[...] nije jasno koja bi bila osnova za takvu pretpostavku o zlonamjernosti 'jake' ili 'široke' umjetne inteligencije" (26) – te naglašava da, unatoč popularnim filmovima čija radnja sadrži slične scenarije, razloga za takva predviđanja nema. No, smatram prikladnim dodati, ne moramo zalaziti u kino dvorane da bismo podastrli dokaze za takve, po ljudsku vrstu nesretne, scenarije. Dovoljno je obratiti pažnju na aktualno stanje, odnosno na vlastito postupanje s manje inteligentnim, ne-ljudskim, životinjama s kojima dijelimo planet. Uzgajamo ih za hranu, prenamjenjujemo njihova prirodna staništa u industrijska postrojenja ili u plantaže za uzgoj raznih kultura koja ostvaruju profit, zatvaramo ih u zoološke vrtove te se njima koristimo u znanstvenim i medicinskim eksperimentima. Ukratko, podčinjavamo ih sebi i svojim interesima. Nije li, analogno tome, razložno očekivati da bi u scenariju nalik Hawkingovu *homo sapiens* igrao vrlo sličnu ulogu kao što je danas igraju ne-ljudske životinje?

Autor je upravo na tragu takvih razmišljanja kada u poglavlju o strojno poboljšanim ljudima (182) izlaže Agarova razmišljanja koji argumentira kako imamo snažne razloge protiv stvaranja poboljšanih ljudskih bića ili "postosoba" – ljudi koji "[...] zahvaljujući genetičkom inženjeringu ili umjetnoj inteligenciji [...] imaju natprosječnu inteligenciju, tjelesnu konstituciju, psihičko stanje i sposobnosti moralnog rasuđivanja" (186) te shodno tome, argumentira Agar, i viši moralni status. Posljedično takvoj moralnoj hijerarhiji slijedilo bi da "[k]ao što ljudske osobe imaju moralno pravo žrtvovati osjećajuće neosobe u potrazi za boljim lijekovima za ozbiljne ljudske bolesti, tako bi i postosobe mogle imati pravo da žrtvuju obične osobe kako bi stekle bolje razumijevanje svojih bolesti".[11] Dakle, imajući gornje scenarije u vidu, čini se da je teret dokazivanja na onima manje zabrinutima i ležernijima oko daljnjeg razvoja umjetne inteligencije. Barem kada se radi o tako velikim ulozima.

Sljedeći izazov tiče se dvojne upotrebe, odnosno mogućnosti da ista tehnologija bude uporabljena u dobre i loše svrhe. Školski primjeri tiču se razvoja "[...] nuklearne, kemijske, digitalne, nanotehnologije, biotehnologije i sl." (28). Dakle, dilema je sljedeća: trebamo li raditi na razvoju opasnih oblika gore navedenih tehnologija kako bismo u budućnosti bili

[11] N. Agar, "Why is it possible to enhance moral status and why doing so is wrong?", *Journal of Medical Ethics* 39(2) (2013), 67–74, navedeno prema Bracanović, *Etika umjetne inteligencije*, 193.

spremni odgovoriti na one koji bi takve tehnologije namjeravali uporabiti u manje plemenite svrhe? I kako kontrolirati razvoj takvih tehnologija da ne izmakne kontroli? Hoćemo li poštivati znanstvenu slobodu ili staviti jaka ograničenja i regulacije oko takvih znanstvenih napora? Ili ćemo pak nastojati naći srednji put između tih dviju strategija?

Kao što je do sada postalo jasno, razvoj umjetne inteligencije i tehnoloških inačica koje se njome služe nije projekt jedne profesije ili industrije, već obuhvaća širok spektar ljudskog djelovanja. Upravo u činjenici da doprinosi mnogih profesija imaju udjela u konačnoj verziji proizvoda obilježenog umjetnom inteligencijom leži sljedeći izazov: kome pripisati odgovornost ako nešto pođe po zlu? Takvi su scenariji brojni i vrlo realni, a Bracanović i navodi neke od njih. "Robot za pomoć nepokretnim osobama može ozlijediti svoga korisnika, bespilotna naoružana letjelica može usmrtiti nedužne civile, a sustav za regulaciju gradskog prometa može prouzročiti prometnu nesreću" (35). Ustanoviti koji je točno dio uređaja, ili njihova kombinacija, prouzročio nesreću i u kojoj mjeri, te tko za to snosi odgovornost nevjerojatno je kompleksno pitanje kada promotrimo cijelu genezu relevantnog tehnološkog uređaja – od znanstvenika i inženjera koji su osmislili sam uređaj, tehničara koji su ga testirali, etičkih povjerenstava i vladinih agencija koje su uređaj pustile u uporabu itd.

Kao posljednji generički etički izazov umjetne inteligencije Bracanović navodi problematiku "vrijednosno osjetljivog dizajna". Kao što autor jasno sažima: "[r]adi se o ideji da bi svi koji sudjeluju u stvaranju tehnologija utemeljenih na umjetnoj inteligenciji [...] trebali u što većoj mjeri anticipirati i nastojati spriječiti njihove etički nepoželjne upotrebe i posljedice" (39). A takve mjere opreza su nužne jer: 1) zakonski propisi će, kao što je uvijek slučaj, kasniti za razvojem tehnologije i 2) regulative će, za razliku od samih tehnologija koje su najčešće u uporabi na globalnoj razini, biti ograničene teritorijalnom jurisdikcijom države ili političke zajednice koja ih donosi.

Nakon iscrpnog uvoda, Bracanović daje ocrt ostatka knjige koji se bavi specifičnim etičkim izazovima a koje je čitatelj do sada mogao i naslutiti. Počevši od problematike etičkih postavki autonomnih vozila, pa kroz propitivanje naših pretpostavki što čini dobar život i kako socijalni roboti utječu na to poimanje, kako strojno učenje potkopava prakticiranje autonomije i pravednosti u društvenom kontekstu te, konačno, kako razvoj umjetne inteligencije stavlja pod upitnik naša sveprisutna uvjerenja o posebnosti ljudske vrste kao neprikosnovenog nositelja posebnog i najvišeg, drugim vrstama nedostižnog, moralnog statusa te pripadajućih

moralnih prava. Dakle, od drugog poglavlja pa sve do kraja same knjige autor nas vodi kroz specifične etičke izazove izazvane razvojem umjetne inteligencije i nastoji na svaki od danih izazova odgovoriti koristeći se prije navedenima normativnim etičkim teorijama.

Tako se u srži drugog poglavlja, "Etičke postavke autonomnih vozila", nalazi široj javnosti već otprije poznat "problem tramvaja" (*trolley problem*) i "dilema pješačkog mosta" (*footbridge dilemma*). Ta dva misaona eksperimenta na kušnju stavljaju naše intuicije o tome "[...] postoji li moralno značajna razlika između toga kada nekoga ubijete i toga kada ga samo pustite da umre" (41). Dakle, budući da je trenutak kada će potpuno autonomna vozila vladati cestama samo pitanje vremena, moramo već sada biti spremni ponuditi odgovore na pitanja kao što su: kako bi se naša vozila trebala ponašati u scenarijima u kojima spašavanje jednog života znači žrtvovanje drugog? Koje kriterije uzeti u obzir? Što čini moralno značajan razlog koji bismo kao zajednica trebali uzeti u obzir prilikom implementacije etičkih postavki u naša autonomna vozila? i sl.

Na ovom mjestu upoznajemo se s etičkim egoizmom, utilitarizmom, deontološkom etikom, ugovornim stajalištem te etičkim relativizmom. I upravo je razrada tih teorija na pitanjima etičkih postavki autonomnih vozila problematika na koju bih se želio više osvrnuti.

Iako je suzdržan i neutralan ton za znanstvenu monografiju sasvim prikladan pristup, na trenutke se čini svrhom samom sebi. Naime, u rješavanju pojedinog konkretno etičkog ili općenito filozofskog problema neke teorije doista jesu bolje od drugih te na sveobuhvatniji i intuitivniji način rješavaju danu dilemu ili izazov.

Pogledajmo kao primjer ugovorno stajalište izloženo u poglavlju o etičkim postavkama autonomnih vozila. Ugovorno stajalište temelji se na Hobbesovu razmatranju po kojemu ljudi, da bi osigurali mirnu koegzistenciju sa svojim susjedima, prebacuju autoritet i legitimnu uporabu sredstava prisile na državu kao arbitra koji osigurava "[...] da se pojedinci doista drže sklopljenih ugovora i danih obećanja". Analogno tome, u kontekstu implementacije etičkih načela u autonomna vozila zastupnici ugovornog stajališta predlažu kao obvezan etički naputak sljedeće: "Minimiziraj ukupnu štetu" (74). Na taj način izbjegavaju se sukobi koje bi autonomna vozila vođena načelima etičkog egoizma zasigurno izazvala jer bi neki pojedinci drage volje žrtvovali druge iz sebičnih interesa, dok bi drugi najvjerojatnije stradali jer su imali altruistične pobude ili su iz nekog drugog razloga imali manje egoistično podešeno autonomno vozilo. Dakle, vlada, kao središnje tijelo koje jamči da su sva autonomna

vozila vođena istim etičkim načelima, osigurava jednakost u prometu a time i u društvu.

Kao verziju *knock-down* argumenta ugovornom stajalištu u poglavlju o etičkim postavkama autonomnih vozila autor navodi nepostojanje stvarnog ugovora koji svi sudionici prometa potpisuju, činjenicu da se moralne intuicije sudionika u prometu često ne bi podudarale s etičkim postavkama njihovih vozila, kao i da bi mnoga ljudska bića, zbog svoje intelektualne nezrelosti (djeca i mentalni bolesnici), bili isključeni iz datog ugovora jer ga nisu razumjeli te, prema tome, niti sklopili.

No, je li to valjan prigovor? Nije li tako da ugovorno stajalište već uređuje mnoga područja našeg svakodnevnog života? Osvrnimo se, primjerice, na prometna pravila. Prometni znakovi i pravila vrijede bez iznimke za svakog sudionika u prometu i nikakav eksplicitan ili implicitan pristanak nije potreban. Niti silom njihova nerazumijevanja prometnih pravila djecu i mentalne bolesnike ne smatramo neravnopravnim sudionicima prometa. Pojedinosti prometnih pravila i sveukupni zakoni koji uređuju ponašanje u prometu delegirani su višim instancama koje kodificiraju data pravila te se od svih očekuje da ih slijede. Analogno tome, mogli bismo kazati, implementacija etičkih pravila u autonomna vozila temeljenih na ugovornom stajalištu zakonska je obveza, vrijedi za sve jednako, ne podrazumijeva eksplicitan pristanak a niti njihovo nerazumijevanje ne lišava dotičnog pojedinca odgovornosti u slučaju njihova nepridržavanja.

U trećem poglavlju, "Socijalni roboti i dobar život", upoznajemo se s raznovrsnim ulogama koje socijalni roboti već igraju, a u budućnosti će vjerojatno odigravati, u našim svakodnevnim životima te s implikacijama koje bi to moglo imati po naše poimanje dobrog života. Također, u ovom se poglavlju susrećemo s problematikom "jezive doline", pojmom koji se koristi za opisivanje nelagode koja se javlja kada humanoidni roboti ili računalno generirani likovi snažno podsjećaju na stvarne ljude, ali još uvijek ne uspijevaju biti potpuno uvjerljivi. "Jeziva dolina" se ustvari odnosi na sinusoidni graf koji prikazuje odnos između sličnosti s ljudima i emocionalne reakcije koju ta sličnost izaziva (usput, možda bi engleski termin *uncanny valley* bilo bolje prevesti kao "dol jeze"). Kako roboti ili virtualni likovi postaju sve sličniji ljudima, povećava se osjećaj poznatosti i, shodno tome, pozitivna reakcija. Međutim, na koncu dolazimo do točke u kojoj male nesavršenosti ili odstupanja od stvarnosti izazivaju snažnu negativnu reakciju, uzrokujući gađenje, nelagodu ili jezivost kod promatrača (usp. 92–93). "No ova neočekivana odbojnost," naglašava

Bracanović, "iščezava kada sličnost robota ljudskome izgledu nastavi dalje rasti prema 100%" (93).

Pitanje je, nastavlja Bracanović, u kojoj je mjeri to stvarni fenomen a u kojoj mjeri tek nedokazana hipoteza koja počiva na anegdotalnim dokazima. No, zanimljivije je pitanje što leži u osnovi "jezive doline". Prema jednoj tezi "[...] robotske replike izgledom podsjećaju na ljude oboljele od zaraznih bolesti i aktiviraju naš evolucijski mehanizam za njihovo izbjegavanje" (94), dok prema hipotezi "ukazivanja na smrtnost" antropomorfni roboti u nama ljudima izazivaju strah jer nas "podsjećaju na vlastitu smrtnost" (94). Dakle, poanta je u tome da tek s razvojem novih tehnologija postajemo svjesni vlastite ljudskosti i dugom evolucijskom poviješću usađenih mehanizama čija adaptivnost u budućem okruženju nije toliko očita.

Socijalni roboti potiču još mnoge druge dileme i izazove od kojih ovdje vrijedi spomenuti barem dva. Bracanović u poglavlju o "Skrivenim opasnostima" dobro ilustrira "obmanjujuće psihološke utjecaje koje roboti, zahvaljujući svojim sve naprednijim sposobnostima, mogu imati na ljude" (101). Među jedan od takvih utjecaja svakako možemo ubrojiti slučaj novijeg datuma, kada se OpenAI udružio s Alignment Research Centrom u svrhu testiranja sposobnosti GPT-4. Slijedeći njihove upute, GPT-4 je, pretvarajući se da je osoba s oštećenjem vida, prevario djelatnika TaskRabbit-a da riješi CAPTCHA – sigurnosni mehanizam koji se koristi na web stranicama kako bi se razlikovali ljudi od automatiziranih računalnih programa – kako bi dobio pristup željenoj web stranici.[12]

Drugi izazov tiče se možebitnog negativnog utjecaja robota za seks i mogućnosti da njihova uporaba dovede do instrumentaliziranja drugih ljudi i opasnosti da ljudi općenito odustanu od potrage za pravim, ljudskim partnerom (usp. 110–111). Ta rasprava, čini se, uvelike nalikuje onoj o učincima pornografije na društvo. Drugim riječima, do kojeg god zaključka došli, seks roboti će vjerojatno biti u optjecaju a štete i blagodati će se tek naknadno zbrajati. Takva se dinamika čini izglednom za gotovo sav tehnološki razvitak inspiriran umjetnom inteligencijom jer svaki novi izum mnogo obećava (u vidu potencijalnog kapitala i podizanja životnog standarda i komfora) a svaki pokušaj regulacije doima se kao puka formalnost čija je jedina zadaća usporiti tehnološki napredak.

Četvrto poglavlje, "Algoritmi, činjenice i vrijednosti", najaktualnije je od svih obrađenih tema, jer algoritmi strojnog učenja u dobroj mjeri već čine našu svakidašnjicu te na njih nailazimo u brojnim sferama ljudskih

[12] https://gizmodo.com/gpt4-open-ai-chatbot-task-rabbit-chatgpt-1850227471

djelatnosti, primjerice u "[...] raznim vrstama poslovanja i trgovine, u *online* sustavima za prevođenje, u osobnim asistentima na pametnim telefonima [...]" (126) i sl. U ovom kontekstu najčešći problem jest zaštita privatnosti i s tom svrhom je Europska unija donijela *Opću uredbu o zaštiti podataka*. U fokusu te uredbe je pravo pojedinca na kontrolu vlastitih osobnih podataka (GDPR 2016, nav. prema 135). No, Bracanović ističe, privatnost shvaćena kao "kontrola nad osobnim podacima" nije načelo do kojeg držimo u brojnim drugim sferama osobnog života. Primjerice, uzmimo u obzir koliko o sebi otkrivamo svojim ponašanjem, odjevnim predmetima, prijevoznim sredstvom, društvenim događajima koje pohađamo (usp. 136) itd. Dakle, problem je u dosljednosti primjene toga načela i, kao što Bracanović zapaža, u činjenici da se ljudi općenito olako odriču kontrole nad vlastitim podacima kada im je, na primjer, ponuđeno da ispune anketu s osobnim podacima u zamjenu za popust na neki proizvod (136).

Četvrto poglavlje obrađuje još mnoge zanimljive teme, od problema transparentnosti dijagnostičkih sustava – pitanja je li etički prihvatiti i postupiti u skladu s dijagnozom koju bi nam dijagnostički sustav dao bez adekvatnog objašnjenja procesa koji je doveo do dane dijagnoze i kome pripisati odgovornost ako nešto pođe po zlu – pa sve do problema pristranosti u rasuđivanju algoritma za koju se vrijedi pitati: "[k]ada neka statistička pristranost zaista jest dokaz moralne pristranosti ili diskriminacije, a kada je ona tek zrcalni odraz stvarnosti koja možda, bez ičije krivice, neugodno odstupa od naših očekivanja i želja" (155)?

Konačno, u petom poglavlju naslovljenom "Moralni status: roboti, ljudi i transljudi", o kojem je već bilo riječi kad sam se dotaknuo problema singularnosti, Bracanović razlaže brojne teorije koje za cilj imaju ponuditi nužne i dovoljne uvjete za pripisivanja moralnog statusa s ciljem odgovaranja na pitanje jesu li i roboti prihvatljivi kandidati za moralni status. Ovdje su moguće dvije strategije. Prva se oslanja na racionalnost kao ključnu komponentu a druga na pitanje je li dani entitet uopće živ. Čini se da kognitivni roboti zbog svoje racionalne komponente imaju svako pravo uživati moralni status jer smo moralni status skloni pripisati inteligentnijim životinjama, a svakako ga pripusujemo ljudima "[...] sa smanjenom (ili nepostojećom) racionalnošću [...]" (175). Slučaj kibernetičkih organizama – strojno poboljšanih ljudi ili biološki poboljšanih strojeva (usp. 177) – stavlja naše intuicije na kušnju jer zamagljuje razliku između organskih životnih oblika i hibrida koji u konačnici ispunjavaju iste kriterije za moralni status (npr. sposobnost osjećanja boli) te bi prema tome trebali uživati isti moralni status.

U vrijeme kada čelni ljudi u industriji umjetne inteligencije pozivaju na oprez, a pionir razvoja umjetne inteligencije Geoffrey Hinton napušta Google[13] kako bi mogao nesmetano govoriti o rizicima umjetne inteligencije, svaki argumentirani doprinos raspravi je dobrodošao, osobito kad dolazi od etičara. Ova će knjiga svakom zainteresiranom čitatelju pomoći u navigaciji kroz složeni etički krajolik umjetne inteligencije. Primjeri navedeni u knjizi, kao i ovi navedeni u prikazu knjige, podsjećaju nas na hitnu potrebu za snažnim etičkim okvirima i odgovornim donošenjem odluka kako bismo osigurali da se umjetna inteligencija razvija i primjenjuje na način koji će biti na dobrobit čovječanstva.

NINOSLAV KRIŽIĆ
Zagreb
ninoslavkrizic@gmail.com
doi: https://doi.org/10.26362/20230106

Boran Berčić, Aleksandra Golubović, Majda Trobok (ur.), *Human Rationality:* Festschrift *for Nenad Smokrović* (Rijeka: Filozofski fakultet Sveučilišta u Rijeci, 2022), 308 str.

Zbornik radova *Human Rationality:* Festschrift *for Nenad Smokrović* uredili su Boran Berčić, Aleksandra Golubović i Majda Trobok. Kako urednici kažu u predgovoru, profesor Smokrović zaslužuje ovaj *Festschrift* ne samo na temelju svog istaknutog profesionalnog rada, već i kao kolega i prijatelj. Postignuća tog riječkog filozofa su mnogobrojna, kako za filozofiju u njegovu gradu tako i za filozofiju u Hrvatskoj, ali i šire – zato ih ja ovdje neću ni pokušati sažeti, a kamoli nabrojati. Pozivam čitatelja odnosno čitateljicu da se o dijelu sveg onog što je Smokrović postigao uvjeri pročitavši već spomenuti predgovor. Reći ću samo ovo. Svi mi koji se u Hrvatskoj (profesionalno) bavimo filozofijom možemo se samo nadati da će naša imena jednog dana biti jednako zvučna kao njegovo. A ako ove retke čita netko čije je ime još i zvučnije, neka se osvrne. Smokrović mu/joj je za petama.

[13] https://mitsloan.mit.edu/ideas-made-to-matter/why-neural-net-pioneer-geoffrey-hinton-sounding-alarm-ai