Etika, robotika i problem moralnog statusa

Anđelić, Antea

Master's thesis / Diplomski rad

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Department of Croatian Studies / Sveučilište u Zagrebu, Hrvatski studiji**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:111:654973

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI
UNIVERSITY OF ZAGREB
CENTER FOR CROATIAN STUDIES

Repository / Repozitorij:

Repository of University of Zagreb, Centre for Croatian Studies

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI

Antea Anđelić

ETIKA, ROBOTIKA I PROBLEM MORALNOG STATUSA

DIPLOMSKI RAD

Zagreb, 2018.

SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI
ODSJEK ZA FILOZOFIJU

ANTEA ANĐELIĆ

ETIKA, ROBOTIKA I PROBLEM MORALNOG STATUSA

DIPLOMSKI RAD

Mentor: doc. dr. sc. Tomislav Janović

Zagreb, 2018.

Sadržaj

1. Uvod 1

2. Robotika – prošlost, sadašnjost, budućnost 3

3. Etika robotike 9

4. Moralni status robota 14

4.2. Život kao uvjet za moralni status 17

4.2. Osjećajnost kao uvjet za moralni status 20

4.3. Osobnost kao uvjet za moralni status 25

4. 4. Odnosi kao uvjet za moralni status 30

5. Zaključak 35

6. Literatura 37

Sažetak 41

Summary 41

1. Uvod

U ovom radu prikazane su neke od osnovnih teorija moralnog statusa i mogućnost njihove primjene na jednu posebnu vrstu entiteta – robote.

Pojam „robot“ u upotrebi je gotovo cijelo stoljeće, a otkad je nastao, pisci znanstvene fantastike u svojim ga djelima vežu uz etička i filozofska pitanja o tome krije li se u njima nešto više od običnih strojeva stvorenih kako bi zamijenili ljude u obavljanju dosadnih ili opasnih poslova. Sve do nedavno spomenuta pitanja nisu izlazila iz takvih priča, međutim, s razvojem tehnologije i pojavom naprednijih vrsta robota, etika robotike postaje sve češća tema kojom se bave etičari. U kontekstu ovog rada posebno su zanimljiva istraživanja koja uključuju rad na „biorobotima“ djelomice izrađenih od bioloških materijala ili „kognitivnim robotima“ koje pokreće umjetna inteligencija koja im omogućuje da uče iz iskustva i prilagođavaju se neočekivanim okolnostima u vlastitoj okolini.

Teme kojima se etika robotike bavi se mogu podijeliti na tri vrste. Jedna se tiče samih robotičara koji kreiraju robote i njihove uloge u reguliranju etičnog ponašanja robota za vrijeme i prije njihove izrade. Druga se tema odnosi na ugrađivanje etičkih načela u robote što bi im omogućilo kontroliranje vlastitog ponašanja bez ljudske intervencije. Posljednja tema je vezana uz ljude koji koriste robote i njihov međusoban odnos. U kontekstu ovog rada najviše su obrađene posljednje dvije teme zbog svoje uske povezanosti s moralnim statusom.

U prvom dijelu rada dan je kratak pregled nastanka upotrebe pojma robota koji se u početku najviše koristi u već spomenutoj znanstvenoj fantastici. Potom slijedi kratak prikaz razvoja robota od najjednostavnijih robotskih ruku koje se koriste u industriji pa sve do onih najnaprednijih koji se danas koriste u medicini, ratovanju ili u istraživanja svemira. Podjelom robota s obzirom na funkcije koje obavljaju doći ćemo do odgovora na pitanje koji od navedenih robota mogu biti kandidati za pripisivanje moralnog statusa.

U drugom dijelu, naslovljenom „Etika robotike“, izložene su neke od osnovnih tema kojima se ovo područje etike bavi. Nakon kratkog definiranja predmeta etike robotike, problematizira se već spomenuto ugrađivanje etičkih načela u robote kojim bi se stvorili AMA (*artificial moral agents*) – umjetni moralni agenti. Posebna pozornost posvećena je ulozi

moralnog agenta koja će kasnije biti važna za pripisivanje moralnog statusa robotima na temelju uvjeta osobnosti.

Glavni i najveći dio rada posvećen je osnovnim teorijama moralnog statusa – konkretno: teoriji utemeljenoj na životu, teoriji utemeljenoj na osjećajnosti, teoriji utemeljenoj na osobnosti i teoriji utemeljenoj na odnosima kao uvjetima za moralni status – i njihovoj potencijalnoj primjeni na moralni status robota. Najviše prostora posvećeno je utilitarističkoj i deontološkoj teoriji koje za uvjete moralnog statusa uzimaju osjećajnost odnosno osobnost. Iako se na prvi pogled čini da je lako odbaciti moralni status robota na temelju tih karakteristika, neki primjeri izneseni u radu ukazuju na to da bi napredak u kreiranju robota mogao promijeniti taj dojam. Posebno se u tim kontekstima problematiziraju već spomenuti „kognitivni roboti“ i „bioroboti“ koji se prema karakteristikama koje posjeduju mogu mjeriti s kriterijima svake od navedenih teorija. Uz ove glavne teorije, jedan je dio posvećen životu, a drugi odnosima kao uvjetima za moralni status, kao također relevantnim za razmatranje pitanja o moralnom statusu robota.

2. Robotika – prošlost, sadašnjost, budućnost

Pojam robota već je skoro stoljeće jedna od središnjih znanstvene fantastike u kojoj i nastaje. Pojam se prvi put pojavljuje u drami češkog spisatelja Karela Čapeka *Rossumovi univerzalni roboti* (R.U.R.) iz 1921. godine. Glavni lik drame, izumitelj Rossum, stvara posebnu vrstu radnika koji su izrađeni od biološkog materijala i dovoljno pametni da zamijene čovjeka u bilo kojem poslu. Za radnike, Čapek rabi riječ „robot" koja potječe od češke riječi *robota* što u prijevodu označava fizički rad. U Čapekovoj drami, Rossum je imao namjeru osloboditi ljude od fizičkog posla ovakvim umjetnim zamjenama, ali se pokazuje da se strojevima ta uloga dugoročno ne sviđa. Oni se stoga okreću protiv svojih dizajnera i istrebljuju ih kako bi uspostavili vladavinu vlastite vrste.

Osim kao biološki „sluge", nešto kasnije se roboti pojavljuju kao mehanički strojevi i to u filmovima *Metropolis* (1926), *The Day the Earth Stood Still* (1951) i *Forbidden Planet* (1956). Unatoč razvoju tehnologije termin se zadržao i na samom početku označavao stroj koji je „bezuman i dobar samo za unaprijed definirane ponavljajuće tipove poslova" (Murphy, 2000: 3). Murphy također ističe kako je nešto antropomorfniju vrstu robota u svoja djela uveo poznati pisac znanstvene fantastike Isaac Asimov. Njegova tri zakona robotike, o kojima će biti nešto više riječi kasnije, pojavljuju se u kratkoj priči *Runaround* koja postaje dio poznate zbirke *Ja, robot* iz 1950. godine. Spomenuta priča, ali i ostale iz zbirke, sjećanja su doktorice Susan Calvin koja je radila kao „robopsiholog" u tvrtki Američki roboti. Svaka priča uključuje nekog od robota proizašlog iz spomenute tvrtke. Roboti su bili napravljeni tako da slijede takozvana Tri zakona robotike. Spomenimo zasad tek da su ta tri zakona bili etički propisi koji su u nekim situacijama stvarali etičke dileme koje roboti nisu mogli jasno razriješiti. Jedan od najpoznatijih pripadnika ove vrste autonomnih, inteligentnih robota u znanstvenoj fantastici je Terminator iz istoimenog filma iz 1984. godine koji, kao što navodi Murphy, unatoč uništenom sloju kože i gubitku jedne kamere koja se nalazila na mjestu oka, nastavlja progoniti Sarah Connor, svoju metu koju je programiran uništiti.

Naravno, roboti ne dolaze isključivo u antropomorfnom obliku kao već spomenuti – u mnogo široj upotrebi su upravo oni roboti koji to nisu. Neki od robota u općenitoj upotrebi koje spominje i Murphy su roboti usisavači ili primjerice robot koji se koristi u bolnicama za dostavu hrane pacijentima koji izgleda poput kolica, ne poput čovjeka. U znanstvenoj fantastici, poznati

primjer robota koji nema antropomorfan izgled je HAL 9000 (akronim od *Heuristically programmed ALgoritmic computer*) koji se pojavljuje u *2001: A Space Odyssey*, knjizi Arthura C. Clarka i filmu Stanleya Kubricka iz 1968. godine. HAL 9000 je računalo koje je zaduženo za nadzor svemirskog broda *Discovery One* koji se zaputio na misiju prema Jupiteru, odnosno Saturnu u romanu. Ono što HAL-a 9000 razlikuje od običnih računala je njegova sposobnost autonomnog djelovanja, koja ukazuje na to da robot može djelovati u održivim uvjetima bez pomoći čovjeka i uzimajući u obzir promjene u vlastitoj okolini i u sebi samome.

Postoje još mnogi primjeri različitih prikaza robota u znanstvenoj fantastici, čiji su autori obično nastojali izraziti strahove i entuzijazam vezan uz ovu vrstu strojeva, ali i znatiželju koja je potaknula tehnološki razvoj i stvaranje. Jedan od najnovijih prikaza robota kroz znanstvenu-fantastiku je serija *Westworld* iz 2016. godine koja se temelji na istoimenom filmu iz 1973. Priča je smještena u posebno dizajniranom zabavnom parku s temom divljeg zapada u kojem žive humanoidni roboti koje posjećuju takozvani „gosti“, pravi ljudi, kojima jedini cilj obitavanja u tom fikcijskom svijetu jest zabava. Serija na zanimljiv način obrađuje upravo ono o čemu će također biti više riječi u nastavku rada, a to je moralni status robota i čovjekovo ponašanje prema stroju koji je postao samosvjestan, ali i dalje predstavlja puk alat za razonodu.

Nakon kratkog prikaza nastanka pojma „robot“ i njegovog pojavljivanja u znanstvenoj fantastici, okrenut ćemo se razvoju robota od najjednostavnijih strojeva, poput prvih robotskih ruku u industriji, preko onih koje danas svakodnevno koristimo, te najnaprednijih strojeva koji služe u medicini, ratovanju ili svemirskim misijama.

Joseph Angelo, u svojoj knjizi *Robiotics: A Reference Guide to the New Technology* (2007: 5), kao začetak upotrebe robotske tehnologije navodi razdoblje nakon Drugog svjetskog rata kada se uslijed razvoja različitih vrsta tehnologija razvijaju i takozvani „misleći strojevi“. Angelo pak navodi (2007: 16) da je motiv za razvijanje industrijskih robota bilo uvođenje strožih pravila o zaštiti na radu. Industrijska revolucija je tu također igrala veliku ulogu jer je upravo jedan industrijski robot 1961. godine u tvornici General Motorsa u New Jerseyu bio među prvima koji se koristio u industriji. Međutim, kao što Angelo ističe, industrijski roboti su se ipak masovnije počeli više koristiti tek 70-ih godina prošlog stoljeća. Među njima je bio i PUMA, *Programmable Universal Machine for Assembly* koji se pojavljuje 1978. godine kao industrijski robot s raznolikim namjenama. Prvi pokušaj povezivanja robotike i umjetne inteligencije dogodio se 1960-ih godina, kako navodi Angelo (2007: 7). Robot *Shakey* imao je kamere kojima je skenirao okolinu i bazične sustave pomoću kojih je mogao prepoznati gdje se

sam nalazi, prepoznati objekte u svojoj okolini i izbjegavati prepreke. To je bio jedan od prvih mobilnih robota koji je mogao obavljati jednostavne zadatke. Prve takve projekte razvijalo je američko Ministarstvo obrane posredstvom agencije DARPA (*The Defense Advanced Research Projects Agency*).

S DARPA-om je surađivala i NASA koja je radila na nekoj vrsti humanoidnog robota koji bi sudjelovao u misijama istraživanja drugih planeta i svemira. Angelo navodi (2007: 12) kako je tadašnji razvoj računalne tehnologije i rad na tranzistorima uvelike utjecao na razvoj kompleksnijih robota koji postaju sve češći alat u svemirskim misijama. Tako su 1976. godine sonde *Viking 1* i *Viking 2* dostigle površinu Marsa i utabale put revolucionarnijim istraživanjima Crvenog Planeta. Dok su spomenuta vozila pokazala da je uz pomoć umjetne inteligencije moguće sigurno sletjeti na Mars bez ljudskog navođenja, uslijedio je novi pokušaj istraživanja s *Mars Pathfinderom* 1996. koji ne samo da je sletio na planet, već je rover *Sojourner*, koji se nalazio u letjelici, uspješno poslao mnogo korisnih podataka natrag na Zemlju. Nakon ovog uspješnog pionirskog istraživanja Marsa, njegovi danas najpoznatiji stanovnici su roboti *Curiosity* i *Opportunity*. Dok je *Curiosity* sletio tek 2012. godine i još uvijek uspješno obavlja svoje misije pokušavajući saznati je li na Marsu ikada bio moguć mikrobiološki život („Mars Curiosity“, 2018.), *Opportunity* je na Marsu još od 2004. godine, ali mu se 2018. gubi signal uslijed pješčane oluje (koja za vrijeme pisanja ovog rada i dalje traje, a nada da će se signal pojaviti ne jenjava) („Six Things About Opportunity's Recovery Efforts“, 2018.).

Osim u svemirskim misijama koje iziskuju strojeve zbog okolina koje su preopasne za čovjeka, roboti se koriste u sličnim uvjetima i na Zemlji. Rad s radioaktivnim materijalima, misije u svemir i potreba za izradom mikročipova postali su plodno tlo za razvoj robota kakve danas poznajemo. Prvenstveno su to bili strojevi koji su sudjelovali u industrijskim poslovima poput rukovanja s opasnim kemikalijama ili drugim po život opasnim poslovima. Ista briga ponukala je razvoj robota u vojne svrhe koji su imali ulogu u obavljanju opasnih operacija poput detektiranja bombi i mina ili u operacijama na opasnim područjima na kojima se sumnja u kriminalne ili terorističke aktivnosti.

Ipak, kao što to i sam Angelo ističe (2007: 17), obično je prva vrsta robota koja padne ljudima na pamet ona iz zabavne industrije, bilo iz već spomenute znanstvene fantastike ili možda iz zabavnih parkova. Još 1999. godine kompanija Sony predstavila je AIBO komplet robota namijenjenih za zabavu koji su izgledali poput kućnih ljubimaca. Najpoznatiji među njima AIBO ERS-7 je robotski pas koji je imao sposobnost odgovaranja na preko 100 govornih

zapovijedi. Tvrtka Honda otišla je korak dalje i 2000. predstavila humanoidnog robota ASIMO-a (*Advance Step in Innovative Mobility*). Cilj je bio napraviti robota koji će poput čovjeka stajati na dvjema nogama, hodati i obavljati određene korisne funkcije. Godine 2005. razvili su još naprednijeg humanoidnog robota koji je imao sposobnost obavljanja poslova u uredu ili kod kuće (primjerice, donošenje pića ukućanima ili zaposlenicima).

Svi navedeni roboti mogu se podijeliti s obzirom na funkcije koje obavljaju, a zbog raširenosti njihove upotrebe, jedan od načina kako ih možemo podijeliti je na uloge koje obavljaju u društvu. Takvu podjelu daje Patrick Lin (2012: 5), a prvu vrstu koju ističe je ona koja se koristi u fizičkom radu i uslugama. Najpoznatiji primjeri te vrste robota su Roomba usisavači koji se sve češće pojavljuju u kućanstvima diljem svijeta. Naravno, postoje još mnogi drugi roboti – za košnju trave, čišćenje bazena, glačanje odjeće i brojne druge fizičke poslove. Sljedeća vrsta je ona koja se koristi u vojne i sigurnosne svrhe. Osim uporabe u vojsci i drugim već spomenutim operacijama, roboti se koriste i u policiji te kao čuvari granica i zgrada, ali i u kućanstvima kao protuprovalni sustavi. Tu je također i upotreba robota prilikom znanstvenih istraživanja i laboratorijskih eksperimenata, a tu svakako spadaju i već spomenuti roboti koji služe za istraživanje svemira. Ne treba zaboraviti ni na robote koji se koriste u industriji zabave, a koji, tvrdi Lin, uvelike doprinose znanosti robotike i boljem razumijevanju kognitivnih procesa, kako umjetno stvorenih tako i postojećih. Važan dio robotike je i onaj vezan uz medicinsku skrb i zdravstvo gdje se strojevi koriste za obavljanje izrazito zahtjevnih operacija, ali i kao pomoć u pružanju sigurnosti i utjehe nekim pacijentima (primjerice, robot PARO u obliku tuljana koji za pacijente ima terapeutske učinke). Vrlo bliski takvoj vrsti robota su i oni koji služe u osobnoj njezi te suživotu, gdje imamo robote koji se koriste ne samo kod djece i starijih ljudi kao neka vrsta prijatelja i sudruga, već i u seksualnoj industriji. Posljednja vrsta koju Lin spominje (2012: 6) roboti su koji se koriste za rad u okolišu na poslovima poput prikupljanja smeća, rad s opasnim otpadima i rad na prikupljanju podataka za istraživanja vezana uz ekologiju.

Kod razmatranja robota u moralnom smislu, međutim, važno je odrediti koja od tih vrsta robota može biti ili moralni agent ili moralni subjekt (za koje se još koristi i naziv „moralni pacijent“). Prema *Cambridge Dictionary of Philosophy* (Johnson, 1999: 590) *moralni agenti* su oni moralni akteri čije je djelovanje podložno moralnom vrednovanju, dok su *moralni subjekti* oni koji su pod direktnim utjecajem moralnih agenata i čija patnja dopušta ili zahtjeva moralno vrednovanje. Podjelu koja može pomoći pri određivanju moralnih uloga kod robota dao je John

P. Sullins (2011: 154-155). On dijeli robote s obzirom na dvije vrste tehnologije koje ih pokreću, a to su teleroboti i autonomni roboti. Teleroboti, kao što im i samo ime kaže, roboti su kojima se upravlja na daljinu i koji donose minimalno autonomnih odluka u svom djelovanju. Primjer za to su, što i sam autor ističe, NASA-ini roveri. Druga vrsta robota koju Sullins spominje su autonomni roboti. Oni, da bi se zvali autonomnima moraju biti sposobni donositi barem neke autonomne odluke u svom djelovanju, koristeći isključivo vlastiti program, ne i vanjsku intervenciju. Neki od najjednostavnijih robota koji se služe upravo svojim programom su već spomenuti kućni čistaći Roomba roboti, a neki od najnaprednijih bi mogli biti autonomni medicinski skrbnici koji će se brinuti za pacijente i donositi odluke vezane uz njihovu dobrobit. Referirajući se na prethodno spomenute definicije moralnih agenata i subjekata, može se reći da bi u slučaju ove podjele autonomni roboti koji donose autonomne odluke o svom djelovanju mogli biti moralni agenti, a svi oni koji trpe posljedice njihovog djelovanja, moralni subjekti. Ipak, autonomija koja se spominje u kontekstu robota (autonomija u tehnološkom smislu) teško se može izjednačiti s autonomijom u filozofskom, odnosno, moralnom smislu. Autonomno ponašanje kod robota, kao što i sam Sullins kaže, jest korištenje vlastitog programa bez vanjskog utjecaja.

Pripisivanje robotima neke vrste slobodne volje objašnjiva je kao proizvod našeg antropomorfiziranja i učitavanja namjera u nežive objekte. Roboti čija je autonomija drugačija od spomenutih strojeva su kognitivni roboti. Za razliku od ostalih vrsta, kognitivni roboti su sposobni učiti na temelju iskustva i prilagoditi se uvjetima okoliša u kojemu se nalaze. Prema Odsjeku za kognitivne znanosti Politehničkog instituta Rensselaer iz SAD-a kognitivni roboti se mogu opisati kao pametni, kreativni i karizmatični („Cognitive Robotics: What is Cognitive Robotics?“, 2018.). Jedinstvene metode učenja ovih robota vezane su uz umjetnu inteligenciju, točnije područje umjetne inteligencije poput „strojnog učenja“, „dubokog učenja“ ili „neuralnih mreža“, a o nekim moralno relevantnim apsektima strojnog učenja bit će više riječi u dijelu koji se odnosi na osobnost kao kriterij moralnog statusa.

Osim autonomije, robote bi moralnim agentima ili subjektima mogla učiniti dispozicija za osjećanje boli ili ugode, o kojoj će također nešto više biti rečeno kasnije. Robote se uzima kao nežive objekte bez dispozicije za osjetilnost, a takvu perspektivu daje još Rene Descartes koji je čak i za životinje tvrdio da su strojevi ili automati koji ne osjećaju bol pa je vršenje eksperimenata nad njima moralno opravdano. Ipak, kao što je to slučaj kod autonomnih robota, postoji nova vrsta robota čiji je razvoj sve napredniji, a to su bioroboti koje još nazivamo

„kibernetičkim organizmima“ ili, popularno, „kiborzima“. Ovi hibridi između mehaničkog robota i biološkog sustava definiraju se kao „čovjek čije se tijelo u velikoj ili nekoj mjeri sastoji od robotskih elemenata, ali je još uvijek živ“ (Gibilisco, 2003: 59). No kibernetički organizmi ne moraju nužno biti ljudi jer se bilo koje hibridno biće sastavljeno od bioloških i ne-bioloških dijelova može smatrati kibernetičkim. O ovoj vrsti robota će također biti više riječi kasnije, u dijelu vezanom uz raspravu o osjetilnosti kao kriteriju za moralni status. No prije prelaska na raspravu o moralnom statusu, osvrnut ćemo se na neke od ideja o kojima se raspravlja u etici robotike.

3. Etika robotike

Iz dosadašnjih razmatranja je vidljivo kako je novo područje etike – etika robotike – itekako zaživjelo u razmišljanjima i djelima suvremenih etičara. Ubrzani razvoj tehnologije i inovacija uvelike otvaraju sve veću potrebu za razvijanjem ovog područja etike. Etika robotike je potrebna više nego ikada jer bi upravo ovaj razvoj mogao uvelike promijeniti sliku svijeta i naša dosadašnja moralna načela. Primjerice, još 2009. godine *The Royal Academy of Engineering* izdala je publikaciju *Autonomus Systems: Social, Legal and Ethical Issues*. U njoj se upozorava da tehnologija sve više ulazi u naš svakodnevni život, ali se uvelike razvija bez odgovarajućih etičkih smjernica. Zbog toga napominju kako je važno da se u potragu za odgovorima na neka od važnih etičkih pitanja u pogledu programiranja i rada strojeva uključe i druge strane, a ne da sve ostane isključivo na inženjerima i programerima (2009: 13).

Patrick Lin, Ryan Jenkins i Keith Abney u predgovoru zbornika *Robot Ethics 2.0: From Autonomus Cars to Artificial Intelligence* (2017) spominju neke od čestih argumenata koji se koriste prilikom obrane ubrzanog tehnološkog razvoja bez obraćanja pozornosti na etiku. Prva što spominju je isticanje brojnih prednosti strojeva koje se uzima zdravo za gotovo i bez previše razmišljanja o njihovim potencijalno negativnim posljedicama. U većini slučajeva, napominju autori, pitanja o pravima, dužnostima i drugim čimbenicima ne mogu se riješiti nekom jednostavnom jednadžbom, pomoću koje bi bilo moguće izračunati sve pozitivne i negativne ishode upotrebe stroja. Druga stvar koja se ističe jest da etika robotike može dovesti do usporavanja tehnološkog razvoja. Lin, Jenkins i Abney ne vide ništa loše u tome što se uzimanjem u obzir etike robotike mogu izbjeći problemi koji često prate prve generacije nekih izuma, što je posebno važno kada se izbjegavaju ozbiljni rizici poput izgubljenih ljudskih života. Primjer proizvoda koji imaju mnoge prednosti, ali se pazi i na njihove nedostatke, su lijekovi koji se ne puštaju u prodaju bez prethodno provedenih kontroliranih kliničkih istraživanja i testiranja na ljudima. Etika robotike, prema autorima, svodi se na neku vrstu sličnog preliminarnog ispitivanja, ali u ovom slučaju tehnologije. Treća stvar koja se prigovara etičkom razmatranju je postojanje zakonskih okvira koji već propisuju što se smije, a što ne smije, kod razvijanja novih tehnologija. Međutim, autori smatraju da nije dovoljno osloniti se na postojeće zakone jer oni nisu nužno etični, a u slučaju robotike su i manjkavi jer ne predviđaju mnogo toga što uključuje strojeve i nove tehnologije. U tom slučaju, složni su autori,

potrebno je krenuti iz početka s etičkim promišljanjem koje bi moglo poslužiti kao nit vodilja u kreiranju zakona.

Zbog svega navedenoga nužna je posebna grane etike – etike robotike. Prije daljnjeg razmatranja ove problematike potrebno je definirati što etika robotike jest i čime se ona bavi. Bit će riječi o putu koji je etika robotike prešla od tri Asimovljeva zakona robotike, koji su nekada pripadali samo znanstvenoj fantastici pa sve do modernih kodeksa za robotičare koji su učinili cijelu priču s etikom robotike stvarnom. U završnom dijelu poglavlja bit će razmotrena i problematika programiranja moralnih načela u same robote.

Prvo pitanje na koje valja odgovoriti jest točno definiranje ove vrste etike, odnosno predmeta kojim se ona bavi. Keith Abney, u članku „Robotics, Ethical Theory and Metaethics: A Guide for the Perplexed“ (2011: 35), kreće upravo od tog pitanja. Za njega se etika robotike odnosi, u dva slučaja, na nešto stvarno, a u trećem slučaju na nešto što možda niti ne postoji. Prvi slučaj se odnosi na etiku koje se trebaju držati oni koji kreiraju robote – robotičari. Drugi pristup odnosi se na moralni kodeks koji bi trebao biti dio programiranih robota. Treći se odnosi na samosvjesnu sposobnost etičkog razmišljanja od strane samih robota. Abney u ovom posljednjem dijelu radi skok koji robota stavlja u poziciju samosvjesnog bića koje ima vlastiti moralni kodeks prema kojem djeluje.

Michael i Susan Leigh Anderson u uvodu zbornika *Machine Ethics* (2011) dijele etiku robotike na sljedeći način: Jedna strana etike robotike tiče se davanja strojevima etičkih načela ili procedura za otkrivanje načina rješavanja etičkih dilema s kojima se mogu suočiti, a koje će im omogućiti da funkcioniraju na etički odgovoran način. Druga strana etike robotike uključuje nešto što Abney izostavlja, a to su ljudi kao korisnici strojeve. Time u etiku robotike dodaju i novu komponentu: odnos ili relcija između čovjeka i robota što može, kao što će se moći vidjeti kasnije u radu, igrati bitnu ulogu prilikom pripisivanja moralnog statusa robotima.

Ukratko: kombiniranjem ovih dvaju pogleda na etiku robotike dolazimo do tri komponente etike robotike. Jedna se odnosi na robotičare koji dizajniraju strojeve. Druga se tiče ugrađivanja etičkih načela u robote koja im omogućuju rješavanje etičkih dilema. Treća se odnosi na ljude koji koriste strojeve te njihov međusobni odnos. Posljednja komponenta bit će jednim dijelom pokrivena u dijelu vezanom uz odnose uvjete za moralni status stoga ovdje neće biti dalje problematizirana. U ovom dijelu najvažniji je drugi dio vezan uz ugrađivanje moralnih

načela u robote, dok je o prvom već bilo riječi na početku ovog poglavlja i doveo je do stvaranja različitih etičkih kodeksa za robotičare.

Wallach i Allen u uvodnom dijelu svoje knjige *Moral Machines* (2009: 4) također ističu već spomenuti ubrzani razvoj robota i njihovu sve veću kompleksnost i samim time potrebu za ugrađivanjem moralnih načela u robote. Roboti se već sada nalaze u ljudskim okruženjima te uključuju donošenje odluka koje utječu na ljude. Wallach i Allen stoga ističu da nije dovoljno pouzdati se u ljudsko nadziranje rada robota. Kao primjer ističu niz potencijalnih tragičnih događaja koji se već sada mogu dogoditi s obzirom na količinu softvera, strojeva i robota koji upravljaju sustavima poput burzi, sigurnosnih i zdravstvenih sustava ili sustava za opskrbu energijom.

Wallach i Allen (2009: 8) smatraju da je načelno moguće napraviti autonoman moralni stroja, što iziskuje pogled na postojeće moralne teorije. Jedna od mogućnosti bio bi etički sustav temeljen na pravilima. Primjer takvog sustava su već spomenuta Tri zakona robotike pisca znanstvene fantastike Isaaca Asimova. Tri pravila dio su kratke priče *Runaround* iz 1942. koja se kasnije pojavljuje u knjizi *Ja, robot* iz 1950. Asimov je naveo kako su navedeni zakoni pisani za (fikcijsko) 56. izdanje *Priručnika robotike* iz 2058. godine. Zakoni kažu sljedeće:

1. Robot ne smije naškoditi ljudskom biću ili nedjelovanjem dopustiti štetu ljudskom biću.
2. Robot mora slijediti zapovijedi koje mu izdaju ljudska bića, osim kada su takve zapovjedi u suprotnosti s Prvim zakonom.
3. Robot mora štititi vlastito postojanje dokle god to nije u sukobu s Prvim ili Drugim zakonom.

Međutim, kao što je već rečeno u prvom poglavlju, ovakav sustav jednostavnih pravila može dovesti do brojnih ozbiljnih etičkih dilema. Pravila se mogu zaobilaziti i lako se dospijeva u beskonačnu petlju. Druga mogućnost je neka vrsta Aristotelove etičke teorije koja se temelji na vrlinama. Prema njoj dobra djela proizlaze iz dobrog karaktera. Međutim, tu etičku teoriju teško je prevesti u etički sustav za strojeve. Teško je odabrati idealan etički sustav koji bi omogućio robotima da djeluju moralno u svakoj situaciji, uslijed čega Wallach i Allen preporučuju okretanje temeljima ljudske moralnosti i ostalim načinima na koje funkcionira ljudsko donošenje odluka, učenje i razvoj. Smatraju da se kroz istraživanje toga koje ljudske

sposobnosti mogu biti implementirane u strojeve može puno toga saznati i o samim ljudima, kao što nam i sami roboti mogu poslužiti u testiranju moralnih teorija. Pitanje o potencijalno idealnoj moralnoj teoriji za strojeve ćemo ovdje ostaviti po strani, a više se osvrnuti na ulogu moralnog agenta koju roboti već sada imaju.

Kada govorimo o robotima kao moralnim agentima, postoje različiti načini na koje se te uloge mogu manifestirati. Jednu podjelu daje James H. Moor (2006). U prvom smislu strojevi su „*ethical-impact agents*" (2006: 19-20), odnosno strojevi koji mogu imati tek etički utjecaj na nas i naše živote. Etika strojeva u tom širokom smislu može se shvatiti kao neka vrsta računalne etike prema kojoj nam računala mogu pomoći u obavljanju dosadnih i monotonih poslova te nam olakšati život, ali i učiniti ga gorim. U drugom smislu strojevi mogu biti „implicitni etički agenti" (*implicit ethical agents*). To su, primjerice, strojevi ili dijelovi softvera koji nam omogućuju da stignemo sigurno do neke lokacije ako uključimo autopilot u pametnom automobilu ili oni koji će nas upozoriti da nam je gorivo pri kraju ili da će nam auto udariti u zid ako se uskoro ne zaustavimo. Treća stavka u Moorovoj podjeli su roboti kao eksplicitni etički agenti (*explicit ethical agent*). Ta vrsta stroja bi u sebi imala ugrađenu neku vrstu etičkog rezoniranja koje bi mu omogućilo donošenje etički ispravnih odluka. Takvi roboti bili bi korisni, primjerice, za vrijeme neke globalne katastrofe u kojoj bi brzo reagiranje stroja prilikom odlučivanja o tome koga prvog spasiti od ugroženih ljudi mogla omogućiti spašavanje većeg broja života. Međutim, prijeporno je treba li takve odluke ostaviti isključivo čovjeku, a ne stroju. Moor nabraja i četvrtu vrstu moralnog agenta, a to su potpuni moralni agenti (*full moral agents*). Takav agent, prema Mooru, može „donositi eksplicitne etičke odluke i biti kompetentan da razumno ih opravdati" (2006: 18). Potpunim moralnom agentom obično smatramo čovjeka koji ima svijest, intencionalnost i slobodnu volju.

Mnogi se ne bi složili s time da je u stroju moguće realizirati spomenuta tri svojstva karakteristična za čovjeka, a argument kojim se ocrtava taj nemogući prijelaz se naziva *bright-line argument*. Postoje dvije vrste *bright-line* argumenata koje spominje Moor. Prva se odnosi na to da samo potpuni etički agenti mogu biti etički agenti. Iz tog gledišta slijedi da ostale uloge koje strojevi mogu imati (*ethical impact agents, implicit ethical agents* i *explicit moral agents*) ne uključuju bivanje moralnim agentom. Ono što se prvenstveno prigovara tim „agentima" je to što njihovo „etičko zaključivanje" potječe od njihovih ljudskih programera, a ne od njih samih. Naravno, to što takva vrsta strojeva nema tri karakteristike svojstvene ljudima (svijest, intencionalnost i slobodnu volju) znači samo da im ne bi trebali davati veće etičke odgovornosti,

ali ne znači da ne mogu biti „etični“ u onim situacijama za koje su napravljeni, tvrdi Moor (2006: 18). Drugi oblik te vrste argumenta kaže kako nijedan stroj ne može postati potpuni moralni agent, odnosno da niti jedan stroj ne može dostići tri karakteristike svojstvene čovjeku – svijest, intencionalnost i slobodnu volju. O tome će biti više riječi kasnije u raspravi o osobnosti kao kriteriju za moralni status.

Nakon ovog kratkog uvoda i pregleda etike robotike, odnosno upoznavanja s njezinim temeljnim teorijama, autorima i problemima, posvetit ćemo se jednom posebnom etičkom problemu povezanom s robotima: moralnom status samih robota.

4. Moralni status robota

Kada se govori o moralnom statusu općenito, on se najčešće vezuje uz bioetičke probleme kao što su moralni status fetusa, životinja ili ljudi s trajnim i neizlječivim invaliditetom. Prije razmatranja glavnih koncepcija moralnog statusa, poželjno je razmotriti pitanje zašto nam uopće treba pojam moralnog statusa. Slično polazište ima i američka filozofkinja Marry Anne Warren koja je veliki dio svog rada posvetila upravo moralnom statusu. Pojam moralnog statusa je važan zbog dviju ljudskih karakteristika: Jedna od njih je spoj ljudske inteligencije i oportunizma (1997: 11-12), a što ima veze s našom sposobnošću za nanošenje zla kako jedni drugima, tako i ostalim bićima koja nas okružuju. Druga, uvelike pozitivnija karakteristika, naša je prirodna sposobnost za brigu o drugim ljudima, ali i o ne-ljudskim životinjama i ne-živim objektima. Te dvije karakteristike ljudi posebno su važne prilikom programiranja robota ukoliko njihovo ponašanje oblikujemo prema ljudskom ponašanju. Ako smo svjesni ovih karakteristika, tada ćemo lakše raditi na ublažavanju negativnih odnosno pojačavanju njihovih pozitivnih aspekata. Također, istu stvar je potrebno uzeti u obzir u obrnutom slučaju, kada razmatramo posljedice ponašanja ljudi prema drugim bićima i entitetima (u budućnosti možda i robotima).

Možemo se dalje zapitati što moralni status znači u praktičnom smislu. Odgovor na to pitanje također nudi Warren, navodeći dvije funkcije moralnog statusa (1997: 13). Moralni nam status prvenstveno služi kao alat za snalaženje u „šumi“ konfliktnih tvrdnji o tome što smijemo, a što ne smijemo činiti. On nam također služi kako bismo barem u minimalnoj mjeri znali što smijemo ili ne smijemo činiti prema entitetima određene vrste. Naravno, sam pojam moralnog statusa ne može zamijeniti našu već spomenutu prirodnu sklonost za brigu o drugim bićima i entitetima, ali zasigurno služi njezinu osnaživanju. Primjerice, još je Hume (App. I, 1777: 239/1963: 289) smatrao da se moralnost temelji na sentimentu odnosno osjećajima. On će reći da je prirodno za nas da se osjećamo ugodno kada djelujemo ili se prema nama netko odnosi u skladu s nekim vrlinama. Prema tome, ako je Hume u pravu, sasvim je prirodno što brinemo i postavljamo sebi pitanje o moralnom statusu entiteta kod kojih možda trenutno ne postoji jasna potreba za moralnim postupanjem, ali bi, s obzirom na tehnološki razvoj, mogla postojati.

U nastavku slijedi razjašnjenje pojma moralnog statusa te na koga različiti teoretičari smatraju da se odnosi. Nakon upoznavanja s osnovnim teorijama moralnog statusa, rasprava će prijeći na glavne argumente za moralni status robota.

Marry Anne Warren u knjizi *Moral status: Obligations to Persons and Other Living Things* (1997: 9) daje koristan pregled glavnih teorija moralnog statusa. Prilikom odgovaranja na pitanje što je to moralni status ističe da je važno imati na umu kako on nije opipljiv predmet, nešto što se može proučavati empirijskim metodama, već sredstvo ukazivanja na one entitete prema kojima imamo moralne obveze, ali i samih obveza koje prema njima imamo. Warren ističe dvije važne karakteristike moralnog statusa. Jedna od njih je općenitost, odnosno to da je moralni status primjenjiv na cijelu skupinu, a ne samo na njezine pojedine pripadnike. Druga važna karakteristika su moralne obveze prema entitetima kojima je moralni status pripisan. Da bi to razjasnila, Warren (1997: 10) daje primjer čuvanja stana prijatelju u kojem čuvar, umjesto da se brine za stan, rasproda iz njega sve kućanske aparate. Warren kaže da je razumljivo kako nećemo reći da je moralna nepravda učinjena rasprodanim kućanskim aparatima, već osobi čiji je stan čuvan i čiji se prijatelj pokazao poprilično nepouzdanim.

Definiciju moralnog statusa možemo pronaći i u *Cambridge Dictionary of Philosophy* (Johnson, 1999: 590), prema kojem se moralni status odnosi na primjerenost nekog bića da bude smatrano prikladnim predmetom izravne moralne obveze. Ova definicija u prvi plan stavlja drugu karakteristiku moralnog statusa koju Warren ističe, a to je moralna obveza. Warren (1997: 3) sažima ono što obje ove definicije govore o moralnom statusu, a to je da ako neki entitet ima moralni status, onda s njime ne možemo postupati kako god želimo, već smo moralno dužni u našem planu djelovanja uzimati u obzir njegove potrebe, interese i dobrobit.

Spomenuta rječnička definicija ukazuje na još jedan važan dio moralnog statusa i zapravo bilo koje moralno relevantne situacije, a to su dvije vrste moralnih aktera koji su već ranije spomenuti. To su „moralni agenti“ i „moralni subjekti“ (koji se još nazivaju i „pacijenti“). „Moralni agenti“ odnose se na one čije je djelovanje podložno moralnom vrednovanju i procjeni, dok su „moralni subjekti“ oni čija patnja, budući da jesu pod utjecajem djelovanja „moralnih agenata“, dopušta ili zahtijeva moralno vrednovanje ili procjenu. Iz toga se može zaključiti da su „moralni agenti“ oni koji imaju moralnu odgovornost.

Autori poput Johna Martina Fischera i Marka Ravizze, S. J. (1998: 1) tvrde da je upravo moralna odgovornost ono što razlikuje ljude od ostalih bića. Oni smatraju da ako prihvatimo da netko jest moralno odgovoran agent, onda to uključuje i spremnost da prihvatimo određene stavove o toj osobi i određena ponašanja prema njoj. Prema njima, dakle, moralno odgovoran agent ujedno je agent koji posjeduje moralni status. Jedino što autori prihvaćaju po pitanju

odgovornosti nekih drugih ne-ljudskih životinja (poput kućne mačke koja je rabila vazu) je kauzalna odgovornost, ali ne i moralna. Također se pozivaju na Strawsonovo viđenje moralne odgovornosti prema kojoj je moralno odgovoran samo onaj agent prema kojemu možemo osjećati poštovanje, zahvalnost, ljubav pa čak i gnjev ili ljutnju (Fischer i Ravizza 1998: 6). Takve osjećaje, kažu autori, u nama ne može pobuditi, primjerice, laboratorijski štakor kojeg koristimo kao subjekt istraživanja.

Spomenuti autori, možemo reći, pripadaju onoj skupini teoretičara koja smatra da samo ljudska osoba može biti i „moralni subjekt“ i „moralni agent“. To pak otvara pitanje: Na koga se moralni status može odnositi? Jedan od načina na koji možemo odgovoriti na to pitanje je kroz razmatranje određenih svojstava koje autori smatraju nužnima da bi se određenom entitetu pripisao moralni status. U slučaju koji razmatraju Fischer i Ravizza riječ je o svojstvu osobnosti, odnosno „biti osoba“. Osobu je definirao još John Locke kao „misleće umno biće koja ima razbor te razmišlja i može smatrati sebe sobom, istom mislećom stvari, u različitim vremenima i mjestima“ (1690: knjiga I, pogl. 9, par. 29/1964: 211). Ukratko: osoba je samosvjesno i racionalno biće.

Stav suprotstavljen gornjemu zastupa sama Warren (1997: 90). Ona tvrdi da osobnost ne može biti jedini kriterij za posjedovanje moralnog statusa. Osobnost kao kriterij za moralni status odbacuju, primjerice, borci za prava životinja. Tako razmišlja Peter Singer (2003: 99), koji za ključni kriterij uvažavanja nečijih interesa uzima „osjećajnost“ (*sentience*), odnosno „interes doživljavanja što je više moguće ugode i što je manje moguće boli“. Prema njemu, sama pripadnost vrsti kojoj se pripisuje osobnost, u ovom slučaju *homo sapiensu*, ne možemo smatrati moralno relevantnom značajkom jer i bića drugih vrsta mogu osjetiti bol. Sličan stav ima i Steve Torrence koji u članku „A Robust View of Machine Ethics“ (2005) izlaže poziciju koju naziva „robustan odgovor“ (*Robust Response*), a prema kojoj moralni agenti mogu biti samo ona bića koja su sposobna osjećati, a takva su biološki organizmi. On se s tim ne slaže i vjeruje da će u budućnosti biti razvijeni roboti s naprednom umjetnom inteligencijom koja će zahtijevati prihvaćanje moralnog statusa robota.

U kontekstu ovoga rada, zanimljivo je kako se roboti mogu promatrati i kao jedno i kao drugo: kao „moralni agenti“ i kao „moralni subjekti“, ali o tome će biti više riječi nešto kasnije. Ono što je potrebno najprije razmotriti jesu osnovne teorije moralnog statusa koje će biti

relevantne prilikom razmatranju pitanja trebaju li roboti imati moralni status, te ako trebaju, kakav moralni status trebaju imati.

Dvije osnovne teorije koje se razmatraju kada se govori o moralnom statusu jesu deontološka i utilitaristička. Klasični utilitarist će reći da je za moralnog agenta važno da uzme u obzir patnju koju njegov ili njezin postupak može uzrokovati svakom živom biću koje će biti pod utjecajem tog postupka. Deontolog smatra da je prilikom djelovanja obveza moralnog agenta poštivati autonomiju svih drugih moralnih agenata i uzimati ih kao svrhu po sebi, a nikada kao sredstva. Warren (1997: 6) spominje i one filozofe koji u potpunosti odbacuju pojam moralnog statusa, smatrajući da on ne može poslužiti kao pomoć za rješavanje etičkih problema. Među njih ubraja etičke egoiste koji će reći da svaki moralni agent ima moralnu obvezu isključivo prema samome sebi te moralne nihiliste koji odbijaju bilo kakva moralna načela. Etički kulturni relativisti pak smatraju da ne postoje nikakve općenite ili međukulturalne vrijednosti – pa tako ni objektivni kriteriji moralnog statusa – jer smatraju da prevladavajuća vjerovanja ovise o pojedinoj kulturnoj skupini. Ovima Warren pridaje i moralne subjektiviste koji smatraju da moralne tvrdnje o moralnom statusu ovise isključivo o pojedinačnom mišljenju te se prema tome ne može doći do temeljnog dogovora oko moralnog statusa jer nemamo razloga prihvatiti nečije mišljenje, a nečije tuđe zanemariti. U nastavku ćemo se usredotočiti, oslanjajući se na Warrenin (1997), na četiri teorije moralnog statusa – život, osjećajnost, osobnost i odnosi kao uvjeti za moralni status – uz poseban naglasak na pitanju kako se roboti uklapaju u svaku od njih.

4.1. Život kao uvjet za moralni status

Teorija o životu kao uvjetu za moralni status pripada Albertu Schweitzeru. Schweitzer smatra da je život nužan i dovoljan kriterij za moralni status. Time ne samo da daje moralni status svim živim bićima, već izjednačava njihove pojedinačne statuse neovisno o vrsti života koju određeni organizam živi. U svojoj knjizi *Civilization and Ethics: The Philosophy of Civization Part II* (1929: 246-247) Schweitzer kaže da se etika „sastoji u tome da spoznajem nužnost imanja jednakog strahopoštovanja pred svakom voljom za životom kao i pred svojom vlastitom“ te da je „čovjek doista etičan samo onda kada se pokorava postavljenom uvjetu da pomaže svakom

životu kojemu može priteći u pomoć, te kada zazire ozlijediti bilo što živo“. Iako se prema njegovim riječima može zaključiti da misli samo na živa bića, u jednom dijelu tumačenja svoje pozicije spominje da u svetost života spada i „neuništavanje nijednog kristala leda što svjetluca na suncu“.

Uzimanje u obzir živog i neživog svijeta u svojoj poziciji jasnije ocrtavaju ekološki etičari. Oni smatraju da je život tek nužan uvjet za moralni status, ali ne i dovoljan, te uzimaju obzir i druga svojstva, poput osjetilnosti ili moralnog djelovanja koja moralnom statusu nekih bića daju veću težinu od ostalih. Warren to stajalište naziva „život plus“ (*Life Plus View*). Ekološki etičari zahvaljujući ovome dodatku mogu pripisati veći moralni status, primjerice, ugroženim vrstama koje su posebno važne za određeni ekosustav, neovisno o tome pripadaju li živom ili neživom svijetu. Schweitzerovo uzimanje života kao kriterija za moralni status je klimavo. On je inspiriran Schopenhauerovom voluntarističkom metafizikom prema kojoj *volja* predstavlja krajnju stvarnost (Martin, 1993: 205) i stoga argumentira kako sva živa bića imaju volju za životom (Warren, 1997: 34). Problem tog argumenta, kao što ističe njegov kritičar autor Mike Martin, jest što je jako blizu „panteizmu, ali bliže bioteizmu – gledištu prema kojem je bog manifestiran i konstituiran cjelokupnim životom“ (1993: 205). Oni koji ne dijele Schweitzerove poglede volju smatraju svjesnom mentalnom aktivnošću, a ne dijelom neke posebne metafizike i ontologije.

Drugi prigovor Schweitzerovoj poziciji (1997: 37) jest da bi takvo radikalno gledište na moralni status imalo značajne posljedice na obavljanje naših svakodnevnih aktivnosti. Čak i sam Schweitzer kaže kako obično kupanje predstavlja masovni pokolj bakterija (1929: 255). Prigovor je i to da ne nudi neku pomoć prilikom sučeljavanja s konkretnim etičkim dilemama (Warren, 1997: 39). Jedino što Schweitzer nudi kao zaključak je da moralni agent mora odlučiti u svakom pojedinačnom slučaju kada je nužno nauditi drugom živom stvorenju i da se mora osloniti na vlastitu subjektivnu odluku. Na kraju, Schwitzerov stav da se svaka osoba mora osjećati krivom jer ne postoji način kako da nešto učini moralno ispravnim nije u potpunosti opravdan. Naravno da postoje slučajevi u kojima moralni agent nije ni na koji način u mogućnosti izbjeći to da nekom živom biću i u tom slučaju jednostavno ne bi bilo opravdano nametati mu krivicu (Warren, 1997: 40).

Warren zaključuje (1997: 49) da ćemo teško dokazati da je samo život dovoljan uvjet za makar i najskromniji moralni status. Tek kada uzmemo u obzir veze među živim

organizmima u biološkim sustavima, možemo pronaći dobar razlog za pripisivanje moralnog statusa čak i najnižim, jednostaničnim oblicima života. Stoga za Warren takozvano „život plus“ stajalište ima više smisla od „samo života“ jer nas moralno ne obvezuje na nerazložno poštovanje najnižih oblika života kao moralno jednakih nama. „Život plus“ nas barem tjera da moralno uvažavamo bića koja pripadaju našoj vrsti, ali opet nas ne sili da se trudimo postići nemoguće, a to je, u Schweitzerovom slučaju, zaštita svih oblika života i osjećaj krivnje ako se tako nešto ne ostvari.

Vratimo li se robotima, Schweitzerov bi uvjet praktički izuzeo iz moralnog uvažavanja – i lišio bilo kakvog moralnog statusa – bilo kakve umjetno stvorene strojeve koji nemaju u sebi *volju* za životom i koji nisu živi u onom smislu u kojem su živa ostala živa bića. Isti općeniti stav, da neživi objekti ne mogu imati moralni status, zastupa i Steve Torrence (2011: 117). Navodi kako mi imamo naviku „okriviti“ neispravne instalacije za kućni požar, iako u takvim slučajevima ne postoji moralna odgovornost, već samo kauzalna. Torrence naglašava da će neki reći da su uređaji koje pokreće umjetna inteligencija poput tog strujnog voda. Ipak, možda je najveći problem u razumijevanju Schweitzerove teorije to što nije jasno što on sve smatra životom, kako bi se znalo što se u tu jednadžbu može uklopiti. Kao što smo već vidjeli, u jednom je trenutku za njega vrijedan spašavanja i kristal leda. Ako to ostavimo po strani i uzmemo da život za njega može imati samo biološki organizam, možda bi njegova teorija mogla uključivati robote koji su dijelom ili u potpunosti stvoreni od biološkog materijala. Takvi su „kibernetički organizmi“ odnosno „kiborzi“ spomenuti u drugom poglavlju ovog rada (njihov prikaz u nastavku oslanja se na Bracanović 2018). Jedan takav robot pojavljuje se 2010. godine na Sveučilištu Reading pod vodstvom Kevina Warwicka. Robot s biološkim mozgom napravljenim od štakorskog neuronskog korteksa uspješno se kretao i izbjegavao prepreke (Warwick, 2012). Još jedan primjer je onaj sa Sveučilišta Illinois gdje su znanstvenici pod vodstvom Rashida Bashira 2014. napravili robota koji je imao biološki sustav za kretanje izrađen djelomično od mehaničkih i djelomično od bioloških dijelova štakora (srce i stanice koštanih mišića) (Cvetković i dr., 2014). Među najnovijim biološkim robotima je robotska raža koju su razvili znanstvenici na Sveučilištu Harvard 2016. s Kitom Parkerom na čelu. Uz mehaničke dijelove, biološki dijelovi su bili sastavljeni od srčanih stanica štakora koji su raži omogućili plivanje i praćenje izvora svjetlosti (Park et. al, 2016).

Svi navedeni roboti veoma su primitivni i ne mogu se mjeriti s čovjekom. Međutim, ako se vratimo na Schweitzerovo shvaćanje o svetosti svakog života, čak i takav primitivan oblik bi zavrijedio moralni status. Ako pretpostavimo da bi u nekom trenutku u budućnosti roboti bili sposobni izražavati *volju* za životom, onda im je, po Schweitzerovome stajalištu, moralni status zajamčen. Ipak, koliko god je Schweitzerov uvjet zanimljiv i dalje ne možemo pobjeći od Martinovog prigovora o problematičnoj ontologiji i metafizici koju on sa sobom nosi.

4.2. Osjećajnost kao uvjet za moralni status

Nakon „života" kao radikalnog, problematičnog i preširokog uvjeta za moralni status, sljedeći uvjet koji se može uzeti obzir kod razmatranja moralnog statusa jest osjetilnost (*sentience*). On potječe iz utilitarističke pozicije i puno je specifičniji i uži od uvjeta „život". Činjenica je, naime, da su mnoga bića sposobna osjećati bol i stoga je normalno pretpostaviti da je takvim bićima pogrešno nanositi bol bez nekog posebnog razloga (Warren 1997: 50-51). Važno je najprije definirati pojam „osjećanosti" (koji u ovom kontekstu nema svoje uobičajeno značenje). Prema Warren (1997: 51) osjećajnost je sposobnost za osjećanje ugode ili boli. Stajališta u etici koja uzimaju u obzir osjetilnost dijeli na dva gledišta: a to su „samo osjetilnost" (*Sentience Only view*) i „osjetilnost plus" (*Sentience Plus view*). Kod „samo osjetilnosti", osjetilnost je (1) nužan uvjet za posjedovanje moralnog statusa i (2) dovoljan uvjet za posjedovanje punog i jednakog moralnog statusa. Nešto drugačije „osjećajnost plus" gledište smatra osjećajnost valjanim kriterijem za moralni status, ali ne i jedinim valjanim kriterijem. Prema tome, osjećajnost je dovoljan uvjet za neki moralni status, ali ne i nužan uvjet za pripisivanje bilo kojeg moralnog statusa. Upravo zbog toga, ovo gledište pruža mogućnost pripisivanja moralnog statusa drugim entitetima koji nisu nužno osjećajući (primjerice neke biološke vrste ili ekosustavi).

Ne zalazeći u filozofiju uma i različita tumačenja osjetilnih doživljaja, može se zaključiti, kao što to čini Warren (1997: 55), da na temelju sličnosti između osjetilnih organa ljudi i drugih bića možemo donositi zaključke o njihovim mogućnostima osjećanja boli ili ugode. Kao što kaže sama Warren, jednostavno nemamo dobrog razloga pretpostaviti da smo mi ljudi po pitanju osjetilnih iskustava jedinstveni i različiti od svih drugih bića. Čini se da su

svi prirodno evoluirali organizmi sposobni imati svjesna iskustva i sposobni osjećati bol i patnju. Nemojmo zaboraviti, međutim, da je ipak moguće zamisliti biće koje ima svjesna iskustva, ali nikada nije iskusilo ugodu, bol ili bilo koji drugi pozitivan ili negativan osjećaj, raspoloženje ili emociju. Warren (1997: 56) daje primjer lika Date iz serije *Star Trek: The Next Generation* koji je svjestan, racionalan, moralno odgovoran i samosvjestan, ali je programiran tako da nije u mogućnosti iskusiti bol, ugodu ili bilo koju drugu emociju. Takvo biće, tvrdi Warren, imalo bi snažan moralni status zbog svoje sposobnosti da bude moralni agent, ali se taj status ne bi mogao temeljiti na osjećajnosti.

U filozofiji, a pogotovo u etici, dosta se raspravlja o posjedovanju dispozicije za nešto, pa tako i za osjećanje boli ili ugode. Warren kaže (1997: 56) da u slučaju osjećajnosti imati dispoziciju znači imati sposobnost ili moć koja se ne mora očitovati stalno, već je potencijalno prisutna. S jedne strane, stajalište o nepostojanju, kod ne-ljudskih životinja, dispozicije za osjećanje ugode ili boli zauzimao je već ranije spomenuti Descartes. Slično suvremeno stajalište zastupa Peter Carruthers koji se slaže s Descartesovom tvrdnjom da su ne-ljudske životinje „automati“ koji imaju iskustva, ali ih nisu nikada svjesni. Carruthers smatra da samim time što ne-ljudske životinje ne koriste jezik kako bi izrazile svoja iskustva, ne mogu tih iskustava niti biti svjesne (1992: 181). Prema tom gledištu, kao što Warren primjećuje (1997: 58), životinje bi uvijek ostale nesvjesne svog iskustva i nikada zbilja ne bi osjećale ništa bolno. Međutim, kao što ističe, tome se može uputiti nekoliko prigovora. Primjerice, pitanje je kako klasično uvjetovanje (*conditioning*) može funkcionirati ako su ne-ljudske životinje nesposobne zbilja osjećati bol. Također je sumnjivo pozivanje na jezik kod izražavanja osjećaja. Osjećaji se mogu izraziti na različite načine i jezik, naravno, tu nije nužan.

S druge pak strane, postoji četiri vrste dokaza za tvrdnju da je pojedino biće sposobno osjećati ugodu ili bol. Prva stvar koja nam može ukazivati na to je živčani sustav. Funkcionalne i strukturalne sličnosti između našeg sustava i živčanog sustava ne-ljudskih životinja mogu ukazati na sličnost s nama. Druga stvar jesu reakcije ne-ljudske životinje kada je ozlijeđena ili izložena nekoj vrsti štetnog podražaja. Treća je prisutnost osjetilnih organa ili ponašanje koje ukazuje na perceptivnu sposobnost za osjećanje boli. Četvrti dokaz za sposobnost osjećanja kod ne-ljudskih životinja je prisutnost neurokemikalija koje se kod ljudi povezuju s osjećanjem ugode, boli ili nekog drugog osjeta (Warren 1997: 60).

Ono što možemo zaključiti iz svega rečenog jest da bez obzira na sve iznesene potencijalne dokaze nije izvjesno jesu li ne-ljudske životinje zbilja sposobne osjećati ugodu i

bol, ali je jednako tako teško zaključiti i posjeduju li tu sposobnost čak i drugi ljudi oko nas. Ipak, iako ne postoje snažni dokazi za takvo što, slobodni smo pretpostaviti da ne-ljudske životinje osjećaju bol i dalje razmatrati utjecaj osjećajnosti na moralni status.

Jedan od zagovaratelja, kako ih Warren naziva, „samo osjećajnost" stajališta je Peter Singer sa svojim utilitarizmom preferencija. Njegov utilitarizam preferencija modifikacija je utilitarizma, konsekvencijalističke moralne teorije koju su utemeljili filozofi Jeremy Bentham, John Stuart Mill i Henry Sidgwick. Jednostavno rečeno, utilitarizam je moralna teorija koja postupke smatra moralno ispravnima ili neispravnima ovisno o posljedicama koje uzrokuju (1997: 64). Da bi nešto bilo ispravno kod klasičnog utilitarizma, ono u svojim posljedicama mora izazivati više sreće od njene suprotnosti. Kod utilitarizma preferencija Petera Singera (2003: 10) u prvom planu su interesi koji se shvaćaju kao „ono što se preferira". Djelovati etički bi značilo odvagnuti sve interese i djelovati na način koji će maksimizirati interese onih koji su tim djelovanjem pogođeni. Razlika u odnosu na klasični utilitarizam je ta „što se 'najbolje posljedice' uzima u značenju da, kada se sve uzme u obzir, promiče interese onih koji su pogođeni, umjesto u običnom značenju toga što povećava ugodu i umanjuje bol". Za Singera se sve valjanje moralne tvrdnje mogu izvesti iz jednog načela: načela jednakog uvažavanja interesa. Prema njegovim riječima, ono „zabranjuje da našu spremnost na uvažavanje interesa učinimo ovisnom o njihovim sposobnostima ili drugim karakteristikama, osim o karakteristici imanja interesa" (2003: 17). Ipak, Singer se zalaže za to da samo osjećajuća bića mogu imati moralni status, jer samo takva bića imaju interese. Temeljni interes za Singera je interes za izbjegavanjem boli (2003: 17).

Gledište „samo osjećajnost" ima nekoliko važnih i teških prigovora. Jedan od njih dolazi od ekoloških etičara koji mu zamjeraju što zanemaruje druge ne-osjećajuće entitete: čitave vrste, biljke i druge ne-osjećajuće elemente biosfere koji su vrlo važni za osiguravanje opstanka planeta i života na njemu. Daljnji je prigovor da utilitarizam ne pruža osnovu za pripisivanje snažnih moralnih prava individualnim ljudskim bićima pa tako ni pojedinim životinjama (1997: 77).

Singer odgovara na ove prigovore razlikovanjem klasičnog utilitarizma i utilitarizma preferencija. Za klasični utilitarizam osjećajuća bića su samo „spremnici" (*receptacles*) za sreću (1997: 77). To bi značilo da su sva bića ustvari zamjenjiva te bi njihovo ubijanje bilo moralno opravdano ako bi dovodilo do više sreće za veći broj „posuda". Ali, za utilitarizam preferencija, ljudi nisu zamjenjivi na ovaj način jer su dovoljno samosvjesni da se boje smrti i da uvelike

preferiraju nastavak vlastite egzistencije. Jedino pravo koje Singerova teorija pripisuje ne-ljudskim životinjama je pravo na jednako uvažavanje interesa (1978: 122). Načelo jednakog uvažavanja interesa tako štiti živote, slobodu i dobrobit osjećajućih pojedinaca. Ipak, neki i mu dalje, uključujući i Warren, zamjeraju labavo uzimanje u obzir prava ljudskih bića u odnosu na druga osjećajuća bića (1997: 78).

„Osjećajnost plus“, s druge strane, izbjegava prigovore upućene jednakom uvažavanju interesa osjećajućih bića jer tretira osjećajnost kao dovoljan, ali ne i nužan kriterij za moralni status i niječe da je to dovoljan uvjet za posjedovanje potpunog moralnog statusa (1997: 84). Prema tom gledištu, ostaje obveza uzimanja u obzir interesa svih bića, ali ne u jednakoj mjeri. Interesi osoba u ovom slučaju imaju veću težinu od interesa jednostavnijih bića.

Osjećajnost kao kriterij za moralni status robota može igrati ulogu kod već spomenutih bioloških robota, odnosno kiborga. Kibernetički se organizmi tako mogu razlikovati s obzirom na udio bioloških komponenata u odnosu na nebiološke. Takvu podjelu nudi Roger Clarke (2011: 261) koji razlikuje „strojno-poboljšane ljude“ (*machine-enhanced humans*) i „biološki poboljšane strojeve“ (*biologically enhanced machines*). S jedne strane, kiborzi u prvom smislu su ljudi koji su „poboljšani“ nekim mehaničkim dijelom tijela poput *pacemakera,* umjetnih udova, umjetne pužnice ili ako su na neki drugi način poboljšani robotskom tehnologijom (lasersko popravljanje vida) (Tzefastas, 2016: 179-183). S druge strane, primjere „biorobota“ koji su izrađeni od neke komponente biološkog podrijetla već smo spominjali u prethodnom poglavlju. Kao što se i tamo dalo zaključiti, ova vrsta robota daleko je od uklapanja u već spomenutu definiciju moralnog agenta. No ovdje valja primijetiti da moralni status ne uključuje nužno obje uloge kod istog entiteta. Kada bi za moralni status bilo nužno da je netko oboje – i moralni agent i moralni pacijent – tada bi to isključivalo ne-ljudske životinje, ali i neke ljude. Moralni status robota stoga ne mora biti povezan s njihovim sposobnostima za učenje, razumijevanje i donošenje odluka. Ako je „biti živ“ – kao što smo vidjeli u prethodnom poglavlju – i „biti osjećajuće biće“ – kao što smo vidjeli u ovom poglavlju – dovoljno za posjedovanje moralnog statusa, bioroboti su prihvatljivi kandidati. Tu spadaju pogotovo svi oni koji su sposobni odgovarati na podražaje iz okoline i prilagoditi se danom okolišu.

Jedan od mogućih prigovora pripisivanju moralnog statusa biorobota, kako ističe Bracanović (2018), jest „prigovor artificijelnosti“ (*artificiality objection*). Ipak, postoje primjeri mnogih umjetno (artificijelno) stvorenih bića koja uživaju puni ili barem djelomični moralni status. Neke od njih su domaće životinje koje su zaštićene od okrutnog ponašanja ili nekih

metoda industrijskog uzgoja unatoč tome što su mnoge od njih stvorene umjetnom selekcijom ili genetskim inžinjeringom. Jednako važan primjer iznosi Petersen (2012: 287) koji spominje primjere ljudi koji su stvorenim metodama poput *in vitro* oplodnje, umjetne implantancije ili surogat-majčinstva. Bilo bi stoga jako problematično tvrditi da artificijelno stvorena bića, samo zato što su artificijelno stvorena, ne mogu imati moralni status. isto se stoga može reći i za umjetno stvorene biorobote. Još jedan prigovor (Bracanović 2018) vezan je uz malu količinu tkiva koje mogu posjedovati ovi roboti. Ipak, mali komadić tkiva izaziva itekako velike rasprave u kontekstu moralnog statusa embrija ili fetusa. Mnogi ljudi imaju i moralnu averziju prema bilo kakvom eksperimentiranju s malim djelićem tkiva, npr. kod kloniranja, istraživanja matičnih stanica ili prodaje organa. Prema tome, mali komadić tkiva može igrati ulogu u dodjeljivanju moralnog statusa: ako bi ovaj „mali dio" postao „veći" (ili možda čitavo ljudsko biće) nespremnost za dodjeljivanje moralnog statusa biorobotima bi se sigurno smanjila.

Warwick (2012: 329) za ilustraciju nudi hipotetičan primjer svjesnog robota čiji je mozak sastavljen od milijardi ljudskih neurona. Tvrdi da taj robot ima više moždanih stanica od nekih životinja poput psa, mačke ili čak čimpanze koje su u većini zemalja diljem svijeta zaštićene pravilima i zakonima. Općenito se zahtijeva da se takve životinje barem tretira relativno dobro, što uključuje radnje poput omogućavanja kretanja na velikim površinama ili zaštitu života u prirodi bez ljudskog uplitanja. Zašto onda takav robot ne bi imao ista, ako ne i veća prava, pita se Warwick. Dakako, u obrnutom slučaju, kada se čovjeku dodaju mehanički dijelovi, njegov moralni status ne dovodi se u pitanje. Patrick Lin (2012: 8) tako ističe da „svjestan ljudski mozak (i njegovo tijelo) ima ljudska prava i zamjena dijelova mozga s nečim drugim, bez utjecanja na njegovu funkciju, čini se da bi sačuvala njegova prava". Prema tome, ne postoji razlog zašto istu logiku ne bismo primijenili u obrnutom slučaju: kada su kibernetički organizmi kreirani dodavanjem bioloških elemenata mehaničkoj cjelini.

Na kraju, kao što postoji dilema mogu li životinje zbilja osjećati bol ili ugodu ili su samo automati (kako je Descartes tvrdio), ista dilema postoji i u pogledu biorobota. Ali, u slučaju kreiranja robota od biološkog materijala životinja ili ljudi čiji je živčani sustav, kemijski sastav i ponašanje poznato, na biorobote bi možda bilo moguće primijeniti prethodno spomenuta četiri dokaza za tvrdnju da su životinje osjećajuća bića. Ako postoji mogućnost da je takav biorobot osjećajan (*sentient*), tada se može reći da on ima i interes za izbjegavanjem boli, čime se uklapa i u Singerov utilitarizam preferencija, a onda i u „samo osjećajnost" stajalište. Pitanje je jedino

u kojoj mjeri bi se uklopili u „osjećajnost plus“, ali bi u krajnjem slučaju imali, ako ne potpuni, a ono barem nekakav moralni status.

4.3. Osobnost kao uvjet za moralni status

Warren razlikuje dvije vrste definicija osobnosti (1997: 90), a to su (i) „maksimalistička“ definicija (*maximalist*) prema kojoj je nužan uvjet za biti osobom to da se jest moralni agent ili da se ima potencijal za postati jednim od njih. Druga definicija je (ii) „minimalistička“ koja ne uvjetuje da se mora biti moralni agent, već samo neku vrstu sposobnosti za mišljenje i samosvjesnost. Jedna od važnijih definicija osobe, ona Johna Lockea koja je spomenuta u uvodnom dijelu ovog poglavlja, glasi da su osobe svjesna bića, odnosno, „misleća inteligentna bića koja imaju razum i sposobnost refleksije i mogu promatrati sebe kao sebe, isto misleće biće, u različitim vremenima i mjestima“ (Locke, 1690/1964: 211). Lockeova definicija ostavlja prostor i za osobnost izvan ljudske vrste. No ono što osnažuje Lockeovu poziciju i njegovu definiciju ne smješta u „minimalističku“ jest dodatak prema kojem je osoba „forenzički pojam koji se pripisuje iz moralnih razloga (agentima koji su sposobni za poštivanje zakona, sreću i mizeriju)“. Stoga je „osoba“, prema Lockeu, prvenstveno moralni agent koji je moralno i legalno odgovoran za svoje postupke. Lockeova definicija spada u „maksimalističku“ definiciju osobnosti.

Primjer još jedne „maksimalističke“ definicije koju spominje Warren je ona Kantova prema kojoj se osobnost sastoji od racionalnog moralnog djelovanja. Prema njegovoj teoriji, djelovati moralno je (1) nužan uvjet za bilo kakav moralni status i (2) dovoljan i nužan uvjet za potpuni moralni status. Warren to gledište naziva „samo osobnost“ (1997: 90). U svojoj *Kritici praktičkog uma* Kant kaže da „postoje dvije stvari koje ispunjaju um s rastućim novim divljenjem i strahopoštovanjem, a to su: zvjezdano nebo iznad mene i moralni zakon u meni“ (1788/2015: 129). Upravo ovo potonje, „moralni zakon u meni“ jest nešto što, prema njegovom gledištu, čini osobu. Prema tome, samo osobe imaju moralne dužnosti, a jedina bića prema kojima osobe imaju moralne dužnosti su druge osobe. Prema Kantu, iako smo moralno dužni ne biti okrutni prema drugim osjećajućim bićima koje nisu osobe, nemamo izravne moralne obveze prema njima.

Da bismo bolje shvatili Kantovu poziciju važno je imati u vidu da se radi o deontološkoj moralnoj teoriji koja se temelji na poštivanju dužnosti, te inzistira na moralnim načelima koja su objektivna i vrijede univerzalno. Moralna ispravnost nekog djelovanja ovisi isključivo o dobroj volji moralnog agenta, a imati „dobru volju“ ovdje znači biti motiviran ispunjavati moralne dužnosti radi samih tih dužnosti. Princip koji Kant drži univerzalnim i najvažnijim je *kategorički imperativ*. On po Kantu vrijedi opće i bezuvjetno te se *a priori* nalazi u našem umu, a glasi: „čini tako da maksima tvoje volje uvijek može ujedno vrijediti kao načelo općeg zakonodavstva“ (1785/1999: 73) Prema Kantu, druga formulacija kategoričkog imperativa je „formula svrhe po sebi“. Ukratko, ona nalaže da druge osobe valja tretirati s poštovanjem ili kao da posjeduju intrinzičnu vrijednost. Budući da su osobe svrhe po sebi, njihova autonomija mora biti poštivana. I upravo zato što su svrhe po sebi, osobe imaju moralna prava koja su svi drugi moralni agenti dužni poštivati.

Prema Warren (1997: 101) snaga Kantove pozicije je osiguravanje moralnih prava osobama neovisno o nekim drugim čimbenicima koji ponekad mogu prevagnuti, kao primjerice kod utilitarizma preferencija. Ona zapravo ograničava ono što je moguće činiti pojedincima kako bi se dostigla veća količina sreće ili kako bi se zadovoljili različiti interesi. Međutim, nedostatak Kantova stajališta jest teza o statusu moralnog agenta kao nečeg što je nužno za pripisivanje moralnog statusa nekom entitetu. Prema tom gledištu, iz skupine racionalnih moralnih agenata koji su svrhe po sebi bile bi isključene ne samo životinje već i mala djeca, ljudi s teškim mentalnim oštećenjima, ali i ljudi s nekim lakšim oštećenjima koja im onemogućavaju da budu moralni agenti. Još jedan problem koji ističe Warren (1997: 103) jest i problem naravi mentalnih sposobnosti potrebnih za moralno razmišljanje da bi na temelju njih netko imao moralni status. Ako se iziskuju napredne sposobnosti, pitanje je tko će uopće imati moralni status čak i kada se radi o zdravim, odraslim ljudima. Poznato je, primjerice, Kantovo mišljenje o ženama koje je smatrao nesposobnima za apstraktno rasuđivanje (1764/2011: 39) i one čije se vrline temelje isključivo na osjećajima, ali ne na intelektu. Pitanje je stoga kako bi se one uklapale u ideju racionalnog moralnog agenta.

Kod „osobnosti plus“ Warren ističe (1997: 119) kako je moralno djelovanje dovoljan uvjet za moralni status, ali nije nužan. To gledište dopušta proširenje moralnog statusa na neka osjećajuća bića koja nisu moralni agenti. Jedan od autora koji daje definiciju osobnosti koja može biti dio „osobnosti plus“ jest John Harris. Prema Harrisu (1985: 18), „osobnost“ znači imati sposobnost cijeniti vlastiti život. Ova definicija puno je uključivija od Kantove i ne

zahtijeva od osobe da bude i moralni agent. Ipak i ova definicija ima teškoća jer također ne dopušta pridavanje moralnog statusa dojenčadi i ljudima s mentalnim oštećenjima.

Roboti za koje bi „osobnost“ mogla predstavljati ulazni kriterij za moralni status su već ranije spomenuti „kognitivni roboti“. Prema UNESCO-vu izvještaju o etici robota iz 2017. kognitivni se roboti opisuju u opreci prema „deterministčkim robotima“. Iako se, kao što je također prethodno rečeno, autonomija kod robota uzima s rezervom jer je, iz filozofskog stajališta, gotovo nepostojeća jer su oni vođeni programom koji su dizajnirali ljudi, kod kognitivnih robota to nije slučaj. Naime, autonomija se kod njih čini drugačija zbog toga što su sposobni učiti iz iskustva i prilagoditi svoje funkcije neočekivanim uvjetima. Primjerice, roboti čije je ponašanje oblikovano strojnim učenjem veoma su slični ljudima u pogledu učenja iz iskustva, točnosti njihovog razlučivanja i neočekivanim rezultatima njihova učenja i rasuđivanja.

Ipak, postoje prigovori (Bracanović 2018: 8) upućeni pridavanju moralnog statusa kognitivnim robotima. Neki će reći da su učenje i rasuđivanje kod robota samo imitacije pravih sposobnosti bez prisustva svijesti, slobodne volje i razumijevanja kao u ljudi. Te tri karakteristike slične su kao i spomenute tri u Moorovom *bright-line* argumentu (i ovdje su tri karakteristike ključne za razlikovanje između ljudi i robota). Međutim, da bi se sačuvala mogućnost moralnog statusa kognitivnih robota moguće je pozvati se na probleme koje u filozofskoj domeni predstavljaju svjesnost, slobodna volja i razumijevanje bez vidljivog rješenja. Kod rasprave o slobodnoj volji, „strogi determinizam“ posve odbacuje njeno postojanje, dok „meki determinizam“ priznaje njezino postojanje, redefinirajući slobodu kao sposobnost odabira djelovanja među ponuđenim mogućnostima (svijest i razumijevanje su složene priče za sebe u koje, u ovom kontekstu, nema potrebe ulaziti).

Još jedan argument koji se može iskoristiti za obranu moralnog statusa kognitivnih robota je „argument marginalnih slučajeva“ (Bracanović 2018). Taj argument već prethodno susreli, a on se obično pojavljuje u praktičnoj etici, a ponajviše u raspravi o pravima životinja: u jednom smislu može se odnositi na neke ne-ljudske životinje poput primata ili dupina kojima se pridaje moralni status zbog racionalnosti i razvijenih kognitivnih sposobnosti; u drugom se smislu može odnositi na ljude poput dojenčadi ili osoba s teškim mentalnim oštećenjima koje su manje racionalne od naprednih kognitivnih robota, a moralni status im se ipak priznaje.

Kod spomenute podjele robota na temelju njihove autonomije, John P. Sullins (2011: 153) autonomne robote koji samostalno donose određene odluke samo uz pomoć vlastitog programiranja vidi kao potencijalne moralne agente s moralnim statusom. Međutim, to uvjetuje zadovoljenjem nešto drugačija tri uvjeta: autonomije, intencionalnosti i odgovornosti (2011: 157). Svjestan je da je autonomiju teško filozofski odrediti, ali daje općenitu definiciju prema kojoj je robot autonoman ako „nije pod izravnom kontrolom bilo kojeg drugog agenta ili korisnika“ (2011: 158). Nadalje tvrdi da ako robot ima autonomiju, kako onda možemo reći da je praktično neovisan agent. U slučaju da mu autonomija kojom raspolaže omogućava da uspješno izvršava određene zadatke, tada robot ima efektivnu autonomiju, tvrdi Sullins. Što je autonomija efektivnija, tvrdi Sullins, to se robotu može pridati veća uloga moralnog agenta. Ipak, ističe kako isključivo autonomija nije dovoljan uvjet da bi nešto bilo moralni agent jer u prirodi nailazimo na različite autonomne organizme, poput bakterija, životinja, ekosustava i sličnih koji jednostavno nisu moralni agenti.

Za intencionalnost Sullins tvrdi kako ipak ne mora biti snažno potvrđena kao što je to slučaj kod ljudi. Tvrdi da „dokle god je ponašanje toliko kompleksno da se pojedinac mora osloniti na standardne zaključke pučke psihologije o predispozicijama ili namjerama da bi se ponašao dobro ili loše“ (2011: 158), tada možemo govoriti o postojanju naznaka upravo tog svojstva kod robota. Odgovornost Sullins karakterizira kao ponašanje koje je objašnjivo samo ako pretpostavimo da je vođeno odgovornošću prema drugim moralnim agentima. Sullins tvrdi da „ako se robot ponaša na takav način i ako ispunjava određenu društvenu ulogu koja sa sobom nosi neku pretpostavljenu odgovornost i ako ju možemo učiniti smislenom samo ako mu pripišemo 'vjerovanje' da ima dužnost brinuti se za pacijenta, možemo tom stroju pripisati moralni status“ (2011: 159). Svoja tri uvjeta Sullins zaključuje primjerom medicinske sestre koja je moralni agent u obavljanju svog posla, a što može postati i robot koji ju jednog dana može zamijeniti na tom radnom mjestu. U slučaju da je taj robot autonoman, ponaša se na intencionalan način i ako mu je program koji ga vodi toliko kompleksan da razumije odgovornost za zdravlje pacijenata pod svojom izravnom brigom, Sullins kaže da je taj robot moralni agent. Ovakav moralni agent bi bio, primjerice, već spomenuti android Data iz serijala *Star Trek: The Next Generation*.

Čini se da osobnost kao uvjet za moralni status robota ulazi u upotrebu kroz legislativu o kojoj se već raspravlja u Europskom parlamentu. Naravno, spomenom legislativa i zakona izlazi se iz okvira same etike. Tako Bryson, Diamantis i Grant ističu (2017: 277) da proglašenje

nekog entiteta pravnom osobom ne znači da je taj čin donesen nakon razmatranja etičkih teorija osobnosti ili metafizike tog pojma. Postoji mogućnost da je to pravilo doneseno samo zato da bi se izazvalo neke poželjne posljedice koje su važne donositelju toga zakona. Autori navode kako zakon koji uključuje pojam osobe ne prati neku strogu definiciju i na temelju određenih karakteristika osobe pridaje nekome ili nečemu taj status, već se zna dogoditi da je to pridavanje dosta fluidno (2017: 278). Prema *Yale Law Yournal* (Smith, 1928: 283) pravna osoba je „subjekt prava i dužnosti. Dati legalna prava ili nametnuti legalne dužnosti znači dati pravnu osobnost…" (Smith, 1928: 283). Zbog toga navode da se struktura određenog pojma u pravnom kontekstu, kao što je u ovom slučaju osobnost, naziva „fikcionalizam", a neka je domena diskursa fikcionalistička ako reprezentira nešto drugo osim doslovne istine. U Americi, primjerice, definicija osobnosti u pravnoj domeni dopušta davanje statusa osobe korporacijama i sličnim neživim objektima. Zbog toga u svojoj pravnoj definiciji osobe, autori navode, osoba je definirana kao fikcija, ali napominju da unatoč tome treba biti razmatrana kao činjenica (2017: 279). Definicija osobe tako u tom smislu ne mora davati odgovor na pitanje je li netko ili nešto osoba u društvenom, biološkom ili nekom drugom smislu. Pravna osoba jednostavno znači da je zakonodavstvo odlučilo dati osobnost nekom entitetu radi ostvarenja ili uskraćenja određenih prava.

Jedan konkretan primjer mogućnosti pridavanja statusa pravne osobe robotima se nalazi upravo u Europi. Riječ je o dokumentu *Civil Laws and Rules on Robotics*, koji je u siječnju 2017. godine donijela radna skupina za pravna pitanja robotike i umjetne inteligencije pri Europskom parlamentu. Dokument ne propisuje pravila kojima će se u budućnosti regulirati ponašanje robota već daje uputu kako da se u ovom trenutku postaviti prema robotima i umjetnoj inteligenciji. Iako se nigdje izričito ne tvrdi da bi roboti trebali dobiti status osobe u budućnosti, u jednom dijelu se spominje da bi barem najsofisticiraniji roboti trebali imati status elektroničke osobe s određenom odgovornošću za svoje ponašanje ili štetu koju bi mogli izazvati. Naravno, tu se uzima obzir i mogućnost autonomnog donošenja odluka te neovisna interakcija s trećim stranama koja također sa sobom nosi određenu odgovornost. Time se daje uputa Europskoj komisiji da bi moralni status mogao biti jedno od rješenja za problem odgovornosti autonomnih, sofisticiranih strojeva koje u budućnosti vjerojatno više nećemo moći smatrati tek pukim alatima za određeni posao.

4. 4. Odnosi kao uvjet za moralni status

Do sada obrađivane teorije u sebi su sadržavale dvije pretpostavke, ističe Warren (1997: 122). Jedna od njih je uzimanje jednog jedinog uvjeta koji je nužan i dovoljan uvjet za potpun i jednak moralni status. Pokazalo se do sada da je svaki od navedenih uvjeta – uvjet za poštivanje života, izbjegavanje okrutnosti prema osjećajućih bića i tretiranje moralnih agenata kao svrha po sebi – dobar, ali nedostatan za pripisivanje moralnog statusa. Zbog toga Warren predlaže napuštanje isključivo jednog uvjeta za moralni status i razmatranje mogućnosti kombiniranja uvjeta. Druga pretpostavka na kojoj su se temeljile prethodne teorije jest ta što su za uvjete uzimana „intrinzična" svojstva, koja u slučaju „samo odnosa" i „etike brige" ne igraju ulogu, već se pripisivanje moralnog statusa temelji na relacijskom svojstvu.

Kod „samo odnosa" teorije, intrinzična su svojstva moralno irelevantna. Glavna pozicija koju Warren iznosi za ovaj uvjet dolazi od J. Baird Callicotta (1997: 123-124), prema kojem moralni status nekog entiteta ovisi o njegovim društvenim i ekološkim vezama u društvenoj ili biološkoj zajednici. Argument za to nalazi u temelju ljudskih moralnih obveza kojeg nalazimo kod Nel Noddings. Ona ističe da imamo moralne obaveze prema bićima koja su psihološki sposobna brinuti i koja mogu, barem potencijalno, biti svjesna i odgovarati na brigu koja im se pruža. Warren uz „samo odnosi" stajalište navodi i stajalište „odnosi plus" koje ipak, uz relacijska svojstva, uzima u obzir i ona intrinzična na temelju kojih se omogućava pridavanje većeg moralnog statusa onima koji posjeduju oba svojstva.

Kod „samo osobnosti" stajališta, Callicot je inspiriran okolišnim etičarem Aldom Leopoldom koji nalazi da su naše moralne obaveze proizašle iz činjenice što smo članovi zajednice. Prema Leopoldu „pojedinac je član zajednice međuovisnih dijelova. Dok ga instinkt tjera da se natječe u zajednici, njegova etika ga tjera da surađuje… Etika zemlje (*land ethics*) širi granice tako što uključuju tlo, biljke i životinje: ili sve zajedno: zemlju" (1970: 239). Leopold je pak inspiriran Humeovom moralnom filozofijom prema kojoj se moralnost temelji na osjećaju (*sentiment*), a ne na razumu. Moralni pojmovi i principi prema tome proizlaze iz naše prirodne tendencije za odobravanje onoga što služi dobru ljudskoj zajednici. Callicott pak napominje da je slično utemeljenje moralnosti na emocijama uočljivo i kod Darwina koji smatra da su u njima utemeljeni naši ljudski moralni instinkti. „Osjećaji privrženosti i simpatije prirodno su nastali kod sisavaca kao sredstvo za osiguravanje reproduktivnog uspjeha", tvrdi

Callicott (1990: 121). Ovakvo utemeljenje moralnosti u osjećajima naziva se još i psihološkim utemeljenjem moralnosti. Tom utemeljenju Callicott dodaje ono što i Leopold spominje, a to je takozvana etika zemlje prema kojoj čovjek nije dio samo društvene, već i biološke zajednice. Time ova pozicija pripisuje moralni status entitetima na holističkoj umjesto na individualnoj razini (1997: 126). Callicott to objašnjava na primjeru pčela čije izumiranje može imati velike posljedice na cijeli ekosustav pa je stoga važnije njima posvetiti veću moralnu pažnju nego psihološki kompleksnijim zečevima ili voluharicama koje su trenutno masovno i globalno raširene i uklopljene u ekosustav (1989: 27). Takozvana zemljišna etika poprilično je biocentrična, ističe i sam Callicott, pa se može dogoditi da u njoj ljudi nemaju snažan moralni status zbog destruktivnog djelovanja koje mogu imati na ostatak biosfere. Predlaže stoga upotrebu inkluzivnije teorije iz koje je zemljišna etika nastala, a to je „biosocijalna teorija".

Prema „biosocijalnoj teoriji", tvrdi Callicott, moralne obaveze potječu iz društvenih i bioloških odnosa. Jedna stvar koju etika zemlje zanemaruje, a biosocijalna nadoknađuje, jest tretiranje domaćih životinja i kućnih ljubimaca. Kućni ljubimci postaju dio takozvanih „miješanih zajednica" – kako ih naziva Mary Midgley – i pokazuje se da imamo naviku pridavati im veći moralni status zbog trenutnih i povijesnih uloga koje u njima obnašaju. Domaće životinje su također ugrožene okrutnom eksploatacijom i uslijed toga im je uskraćena uloga u vlastitoj ili ljudskoj društvenoj zajednici. Ipak, „biosocijalna teorija" dopušta uklanjanje onih životinja koje nisu esencijalne za neku zajednicu ili biosferu ili u slučaju kada ugrožavaju članove zajednice. Prema tome, ova teorija ne zahtijeva od ljudi da postanu vegetarijanci kao što to, primjerice, zahtjeva utilitarizam preferencija.

Jedna od prednosti Callicottove verzije moralnog statusa koji se temelji na odnosima jest to što dopušta da prepoznamo moralne obveze prema biljkama, životinjama i drugim populacijama, kao i prema neživom svijetu, objektima poput rijeka, mora, planina i močvara. Zapravo, ono što se još krije u ovoj verziji bio-društvenog moralnog statusa je i pripisivanje istog moralnog statusa dojenčadi, maloj djeci koja još nisu moralni agenti ili ljudima s mentalnim teškoćama (Warren 1997: 131). Još jedna snaga ove teorije je i pragmatizam, odnosno veća praktičnost u odnosu na prethodno iznesene teorije. Ova bio-socijalna teorija priznaje, primjerice, da ljudi i ekosustavi imaju prednost pred drugim ne-ljudskim individuama, bilo da se radi o mikrobima, bilo o sisavcima.

Naravno, kao i svaka druga teorija, teorija „samo odnosi" ima svojih nedostataka. Ozbiljni problemi pojavljuju se prilikom odgovaranja na pitanje zašto bismo temelj moralnog

statusa postavili isključivo u društvene i ekološke odnose. Callicott odgovara da su kod biološke teorije prednosti „teorijska unificiranost, koherencija i unutarnja konzistentnost“ (1997: 132). Warren ističe (1997: 133), međutim, da nam unatoč tome, ova teorija ne pruža načelo za rješavanje sukoba između različitih *prima facie* moralnih obveza. Teorija također zahtijeva od nas da zaniječemo moralni status osobama i drugim osjećajućim bićima koja nisu dio društvenih ili bioloških zajednica.

Još jedan od problema koji se javlja je problem vanzemaljaca i životinja koje ne borave u zajednicama. Problem Warren ilustrira primjerom izvanzemaljskog bića poput lika iz znanstveno-fantastičnog filma E.T. koji nema neke postojeće društvene veze ili odnose s ljudskim bićima i nema prirodnu ulogu u bilo kojoj zemaljskoj biološkoj zajednici (1997: 136). S jedne strane imamo djecu koja pokušavaju spasiti to biće uviđajući njegovu dobru narav i dobroćudnost, a s druge imamo uplašene odrasle koji žele zarobiti nepoznato biće i uništiti ga. Pitanje koje se postavlja glasi: čije je ponašanje moralnije u tom scenariju? Sam Callicott nema strpljenja ulaziti u ovakve probleme tvrdeći da je vrlo nevjerojatno da ćemo ikada pronaći takva bića (ili oni nas) s obzirom na beskonačnost svemira. Zaključuje kako bio-socijalna teorija ne zahtijeva od nas da se ponašamo nasilno prema vanzemaljskim bićima, već ih uopće ne uzima u obzir. Tu je, dakako, još primjer životinja koje nisu povezane s ljudima i koje nisu nužne za ekosustav. Prema postavkama ove teorije, značilo bi da mi nemamo apsolutno nikakve moralne obveze prema takvim bićima.

Još jedna zanimljiva teorija s odnosom kao uvjetom za moralni status je verzija „etike brige“ koju zastupa već spomenuta Nel Noddings (izvornu ideju „etike brige“ predložila je Carol Gilligan). Prema Noddings, odnos je jedini važan za sve naše moralne obaveze. U brižnom odnosu „onaj koji brine“ (*one-caring*) osjetljiv je na osjećaje i potrebe „onih za koje se brine“ (*cared-for*) i stoga je spontano motiviran na njih odgovoriti. Noddings to naziva i „osjetljivost na osjećaje“ (*feeling-receptive*), što može i ne mora uključivati snažne emocije. Iako se čini da će u slučaju kada je briga jedini kriterij u obzir doći i ne-ljudske životinje ili neki granični slučajevi ljudi, Noddings tvrdi da najsnažnije moralne obveze imamo prema ljudima i to onima koji nam mogu uzvratiti brigu. Time se Noddings dovodi u nezgodnu poziciju u kojoj imamo moralne obaveze samo prema drugim ljudima, dok su ne-ljudske životinje, biljke i neživi objekti isključeni iz te jednadžbe. S obzirom na strogost uvjeta uzvraćanja brige, Noddings (1997: 144-143) bez ustručavanja odbacuje moralne obaveze prema ljudima koji nisu sposobni izraziti svoje osjećaje i brinuti za druge. Ipak, zbog toga što bi biće poput spomenutog

E.T.-ja bilo sposobno odgovoriti na brigu koja mu se pruža, Noddings izbjegava prigovor Callicottovim odnosima. Noddings tvrdi da bi bili sposobni uzvratiti jednakom mjerom i da mogli imati nekakve moralne obveze prema njemu. Još jedan problem s ovom teorijom je i odbacivanje samih moralnih pravila ili principa, čime smo ostavljeni bez moralnog vodstva u slučajevima kada nas sposobnost za brigu i primanje brige iznevjeri.

U kontekstu robota, biosocijalna teorija mogla bi pridonijeti moralnom statusu robota. Primjer jednog takvog robota koji postaje dijelom zajednice je robot „Husky“ koji služi za pomoć stanovnicima jednog indijskog sela u prenošenju kanti s vodom koju moraju svakodnevno prikupljati („Robot Becomes Part of the Community After Easing Daily Burden of Water Collection in Remote Village“, 2018.). Stanovnici koji inače nemaju previše dodira s tehnologijom i do sada se nisu susretali s robotima, osim u filmovima, prihvatili su Huskyja kao dio svoje zajednice upravo zbog pomoći koju im je svakodnevno pružao i olakšavao im posao koji bi inače sami morali raditi. Još jedan primjer je i robot pas koji se koristio u vojnim misijama u Iraku (Scheutz: 2012: 212). Njegov je zadatak bio deaktivirati mine uz pomoć svojih brojnih udova, međutim, vrlo brzo je postao dio zajednice s ostalim vojnicima.

Ovaj primjer je također primjenjiv i na „etiku brige“ jer pokazuje još jednu stranu odnosa koji je robot imao s ostatkom skupine, a to je emotivna povezanost. Naravno, u ovom slučaju nije bilo reciprociteta koji Noddings zahtijeva pa ne možemo reći da ovaj primjer u potpunosti odgovara, ali dobar je pokazatelj da ljudima ponekad i nije potreban reciprocitet da bi se emotivno povezali. Primjerice, nakon što je spomenuti robot pas bio uništen uslijed uspješno obavljene misije, jedan ga je tehničar složio iz dijelova preostalih nakon eksplozije i dao mu ime. Također je prepričao kako se vojnici izrazito vežu uz takve robote i postanu uzrujani kada im se nešto dogodi. Zanimljiv je svakako primjer studije koja uključuje istraživanje povezanosti Roomba usisavača i njihovih „gospodara“ (Scheutz 2012: 213). Neki koji ih koriste postanu toliko zahvalni na poslu koji roboti obavljaju da čak znaju „pustiti robota da se odmori“ i počistiti umjesto njega, nositi ga sa sobom na putovanja ili ih upoznavati s članovima svoje obitelji. Ipak, koliko god bile simpatične, čini se da opcije ograničavanja na naše sposobnosti za brigu o drugima i primanje iste te ograničavanje na biološke i društvene zajednice nisu najbolje opcije. Ono što i sama Warren predlaže, a što se i nameće kao zaključak nakon svih iznesenih teorija, jest potreba za kombiniranjem uvjeta za moralni status (1997: 147).

5. Zaključak

Ovaj rad dao je prikaz nekih od osnovnih teorija moralnog statusa primijenjenih u kontekstu robota kao relativno nove teme u etici robotike.

U gotovo stoljetnoj upotrebi pojma robota i uznapredovanog razvoja tehnologije, vidjeli smo da su roboti prešli put od najjednostavnijih strojeva koji su stvoreni da bi zamijenili ljude u obavljanju dosadnih ili opasnih poslova, do naprednih robota koji se koriste u etički osjetljivim područjima poput medicine, ratovanja ili istraživanja svemira. Zbog osjetljivosti navedenih područja, masovnosti upotrebe robota i nepredvidivosti okolnosti u kojima oni djeluju pojavilo se pitanje ugrađivanja etičkih načela u robote, što za sobom povlači i definiranje dvije vrste moralnih aktera u moralno relevantnim situacijama, moralnih agenata i moralnih subjekata. Među četiri vrste moralnih agenata, prema Mooru, jedna se istaknula kao nedostižna za robote, a to je potpuni moralni agent. Za takvog agenta se za sada uzima da može biti samo čovjek jer jedino on posjeduje svijest, intencionalnost i slobodnu volju koje, prema *bright-line* argumentu, moraju imati moralni agenti da bi mogli biti potpuni moralni agenti, a roboti su ti koji ih nikada neće moći imati.

Rasprava o uvjetu osobnosti kao deontološkom kriteriju za moralni status pokazala je da se te i slične karakteristike (koje se inače koriste kao razlikovne značajke između ljudi i ostalih bića) možda mogu pronaći kod nove vrste robota – kognitivnih robota. Dva argumenta staju u obranu moralnog statusa kognitivnih robota. Jedan od njih ističe da su takve karakteristike nerješivi problemi kojima se bavi filozofija uma. Drugi se oslanja na „granične slučajeve“ koje uključuju neke primate, malu djecu ili ljude s teškim mentalnim oštećenjima, koji možda nemaju spomenute karakteristike, ali im se svejedno priznaje moralni status. Još neke karakteristike svojstvene čovjeku su autonomija i odgovornost. Ipak, čak ni one nisu prepreka za pripisivanje moralnog statusa kognitivnim robotima jer se zbog načina učenja iz iskustva, točnosti njihova razlučivanja i neočekivanih rezultata učenja i rasuđivanja mogu uspoređivati s čovjekom, tako da isto vrijedi i za spomenute karakteristike.

Osjećajnost kao uvjet za moralni status dugo je vremena bio dio obrane ne-ljudskih životinja u kontekstu pridavanja moralnog statusa. Ova konsekvencijalistička pozicija potječe iz utilitarizma, odnosno utilitarizma preferencija Petera Singera. Unatoč njegovoj strožoj poziciji pripisivanja moralnog statusa samo osjećajućim bićima koja, zato što osjećaju, mogu

imati interese, u raspravi je postalo vidljivo da postoji vrsta robota koja se u to može uklopiti – „kibernetički roboti“ ili popularno nazvani „kiborzi“. Ovi „biološki poboljšani strojevi“ kako ih naziva Roger Clarke, prihvatljivi su kandidati za moralni status na temelju uvjeta osjećajnosti. Prigovori iz artificijelnosti i isticanja da je riječ o „malom komadiću tkiva“ nisu se pokazali uspješnima jer se u njih mogu uklopiti čak i neki ljudi poput male djece i ljudi s teškim mentalnim oštećenjima koji mogu također biti umjetno stvoreni, dok rasprava o „malom komadiću tkiva“ u kontekstu lišavanja moralnog statusa nailazi na odbojnost i neodobravanje.

Dodatna dva uvjeta za moralni status o kojima je bilo rasprave su život i odnosi. Život u biološkom smislu kojeg spominje Schweitzer svakako uključuje već spomenute kiborge. Ne treba zaboraviti ni na kognitivne robote koji bi se izražavanjem *volje* za životom također mogli uklopiti ovaj uvjet. Odnosi koji se temelje na Humeovom emotivnom utemeljenju moralnosti zasigurno su jedan od najuključivijih uvjeta jer je u čovjekovoj prirodi njegovati prisne odnose s bliskim ljudima, ali i ne-ljudskim životinjama te neživim objektima. Primjeri istraživanja koja su pokazala ljudsku tendenciju za vezivanje uz robote različitih vrsta i oblika, ukazuju na mogućnost pripisivanja moralnog statusa različitim vrstama robota.

Razmatranje različitih teorija u ovome radu pokazalo je kako je veoma teško da ijedna od njih izravno i sama posluži pripisivanju moralnog statusa bez većih problema. Međutim, s obzirom na to da su se neki roboti pojavljivali kao kandidati za moralni status prema više različitih teorija, čini se da bi neki oblik kombiniranja navedenih teorija mogao predstavljati rješenje za pružanje stabilnijeg okvira za dodjelu moralnog statusa robotima. Načine kombiniranja i nove poteškoće koje se iz toga mogu pojaviti tek treba istražiti, a s obzirom na napredak tehnologije i pojavu sve naprednijih robota koji će s vremenom ući u našu svakodnevnicu, vjerujemo da će dodatno potaknuti potragu za najboljim rješenjem.

6. Literatura

- „Cognitive Robotics: What is Cognitive Robotics?“ (2018.) *Cognitive Science Department – School of HASS,* (http://www.cogsci.rpi.edu/~heuveb/teaching/CognitiveRobotics/What%20Is%20Cognitive%20Robotics.html , stranica posjećena 31. kolovoza 2018.)
- „Mars Curiosity“ (2018.) *National Aeronautics and Space Administration* (https://www.nasa.gov/mission_pages/msl/index.html, stranica posjećena 31. kolovoza 2018.)
- „Robot Becomes Part of the Community After Easing Daily Burden of Water Collection in Remote Village“, *Good News Network* (https://www.goodnewsnetwork.org/robot-becomes-part-of-the-community-after-easing-daily-burden-of-water-collection-in-remote-village/, stranica posjećena: 31. kolovoza 2018.)
- „Six Things About Opportunity's Recovery Efforts“ (2018) *National Aeronautics and Space Administration* (https://www.nasa.gov/feature/jpl/six-things-about-opportunitys-recovery-efforts, stranica posjećena: 31. kolovoza 2018.)
- Abney, K. (2011) „Robot, Ethical Theory, and Metaethics: A Guide for the Perplexed“, u: Anderson, M., Leigh Anderson, S. (ur.) *Machine Ethics*, Cambridge: Cambridge University Press, str. 35-52.
- Anderson, M., Leigh Anderson, S. (2011) „Introduction“, u: Anderson, M., Leigh Anderson, S. (ur.) *Machine Ethics*, Cambridge: Cambridge University Press, str. 7-12.
- Angelo, J. A. Jr. (2007.) *Robotics: A Reference Guide to the New Technology*. Westport: Greenwood Press.
- Asimov, I. (1950.) *I, Robot*, Gnome Press.
- Bracanović, T. (2018.) “On the moral status of robots” (rukopis pripremljen za zbornik radova sa znanstvenog skupa Robots and Societies: What Transformations? What Regulations?, Laval University, Quebec 2017)
- Bryson, J. J, Diamantis, M. E., Grant, T. D. (2017) „Of, for, and by the people: the legal lacuna of synthetic persons“, *Artif Intell Law*, sv. 1 (25): 273-291.
- Callicott, J. B. (1989.) “Animal Liberation: A Triangular Affair”, *In Defense of the Land Ethics: Essays in Enviromental Philosophy*, Albany, NY: State University Press.
- Callicott, J. B. (1990.) “The Case Against Moral Pluralism”, *Enviromental Ethics*, sv. 12 (2): 99-124.
- Carruthers, P. (1992) *The Animals Issue: Moral Theory in Practice*, Cambridge: Cambridge University Press.
- Clarke, R. (2011) „Asimov's Laws of Robotics: Implications for Information Technology“, u: Anderson, M., Leigh Anderson, S. (ur.) *Machine Ethics*, Cambridge: Cambridge University Press, str. 254-296.

- Cvetković, C., Raman, R., Chan, V., Williams, B. J., Tolish, M., Bajaj, P., Sakar, M. S., Asada, H. H., Saif, M. T. A., Bashir, R. (2014.) "Three-dimensionally printed biological machines powered by skeletal muscle", *Proceedings of the National Academy of Sciences*, sv. 111 (28): 10125-10130.
- Fischer, J. M., Ravizza, S. J., M. (1998.) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge, United Kingdom, New York, USA, Melbourne, Australia: Cambridge University Press.
- Gibilisco, S. (2003.) *Concise Encyclopaedia of Robotics*, New York, San Francisco, Lisbon, London, Madrid, Mexico City, Milan, New Delhi, San Juan, Seul, Singapore, Sydney, Toronto: McGraw-Hill.
- Harris, J. (1985) *The Value of Life: An Introduction to Medical Ethics*, London: Routledge.
- Hume, D. (1777/1963) "Appendix I.: Concerning Moral Sentiment", *Enquires: Concerning the Human Understanding and Concerning the Principles of Morals*, 2. izdanje, Oxford: At The Clarendon Press.
- Johnson, E. (1999.) „Moral status", u: Audi, R. (ur.) *The Cambridge Dictionary of Philosophy*, 2. izdanje, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo: Cambridge University Press, str. 590.
- Kant, I. (1764/2011) *Observations on the Feeling of the Beautiful and Sublime and Other Writings*, prev. P. Frierson, P., Guyer, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, Dubai, Delhi, Tokyo, Mexico City: Cambridge University Press.
- Kant, I. (1785/1999) *Groundwork on The metaphysics of morals*, prev. M. J. Gregor, United Kingdom, USA: Cambridge University Press.
- Kant, I. (1788/2015) *Critique of Practical Reason,* prev. M. J. Gregor, United Kingdom: Cambridge University Press.
- Leopold, A. (1970.) *A Sand County Almanac*, New York: Ballentine Books.
- Lin, P. (2012.) „Introduction to Robot Ethics", u: Lin, P., Abney, K., Bekey, G. A. (ur.) *Robot Ethics: The Ethical and Social Implications of Robotics*, Cambridge, Massachusetts, London, England: The MIT Press, str. 3-16.
- Lin, P., Jenkins, R., Abney, K. (2017) „Preface", u: Lin, P., Jenkins, P., Abney (ur.) *Robot Ethics 2.0: From Autonomous Cars to Artifical Intelligence*, Oxford: Oxford University Press, str. ix-xiv.
- Locke, J. (1690/1964) *An Essay Concerning Human Understanding*, prev. A. D. Woozley, Cleveland, Ohio: World.
- Martin, M., W. (1993.) „Rethinking Reverence for Life", *Between the Species*, sv. 9 (4): 204-213.
- Moor, J. H. (2006.) „The Nature, Importance, and Difficulty of Machine Ethics", *IEEE Intelligent Systems*, sv. 21 (4): 18-21.

- Murphy, R. R. (2000.) *Introduction to AI Robotics*, London, England:. A Brandford Book, Cambridge, Massachusetts: The MIT Press.
- Park, S.-J., Gazzola, M., Park, K. S., Park, S., Di Santo, V., Blevins, Erin L., Lind, J. U., Campbell, P. H., Dauth, S., Capulli, A. K., Pasqualini, F. S., Ahn, S., Cho, A., Yuan, H., Maoz, B. M., Vijaykumar, R., Choi, J.-W., Deisseroth, K., Lauder, G. V., Mahadevan, L., Parker, K. K. (2016.) "Phototactic guidance of a tissueengineered soft-robotic ray", *Science*, sv. 353 (6295): 158-162.
- Petersen, S. (2012.) "Designing people to serve", u: Lin, P., Abney, K., Bekey, G. A. (ur.) *Robot Ethics: The Ethical and Social Implications of Robotics*, London: (MIT Press, str. 283-298.
- Scheutz, M. (2012.) „The Inherent Dangers of Unidirectional Emotional Bonds between Human and Social Robots", u: Lin, P., Abney, K., Bekey, G. A. (ur.) *Robot Ethics: The Ethical and Social Implications of Robotics*, London: (MIT Press, str. 205-222.
- Schweitzer, A. (1929.) *Civilization and Ethics: The Philosophy of Civilization Part II*, London: A. & C. Black.
- Singer, P. (1978) „The Fable of the Fox and the Unliberated Animals", *Ethics*, sv. 88 (2): 119-126.
- Singer, P. (2003.) *Praktička etika*, Zagreb: Kruzak.
- Smith, B. (1928.) „Legal Personality", *Yale Law*, sv. 37 (3): 283-299.
- Sullins, P. (2011.) „When Is a Robot a Moral Agent?", u: Anderson, M., Leigh Anderson, S. (ur.) *Machine Ethics*, Cambridge: Cambridge University Press, str. 151-161.
- The Royal Academy of Engineering (2009.) *Autonomus Systems: Social, Legal and Ethical Issues*, izvještaj o društvenim, pravnim i etičkim problemima razvoja i upotrebe autonomnih sustava, https://www.raeng.org.uk/publications/reports/autonomous-systems-report (stranica posjećena: 31. kolovoza 2018.)
- Torrence, S. (2005.) „A Robust View of Machine Ethics" (sažetak za izlaganje na konferenciji Association for the Advancement of Artificial Intelligence, USA 2005).
- Torrence, S. (2011.) „Machine Ethics and the Idea of a More-Than-Human Moral World", u: Anderson, M., Leigh Anderson, S. (ur.) *Machine Ethics*, Cambridge: Cambridge University Press, str. 115-137.
- Tzafestas, S. G. (2016.) *Roboethics: A Navigating Overview*, New York, Dordrecht, London: Cham, Heidelberg: Springer.
- UNESCO (2017.) "Report of COMEST on robot ethics" (UNESCO: Paris), http://unesdoc.unesco.org/images/0025/002539/253952E.pdf (stranica posjećena: 31. kolovoza 2018).
- Wallach, W., Allen, C. (2009.) *Moral Machines: Teaching Robots Right from Wrong*, New York: Oxford University Press.

- Warren, M. A. (1997.) *Moral status: Obligations to Persons and Other Living Things*, Oxford: Clarendon Press.
- Warwick, K. (2012.) "Robots with biological brains", u: Lin, P., Abney, K., Bekey, G. A. (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics*, London: MIT Press, str. 317-332.

Sažetak

U diplomskom radu prikazane su osnovne teorije moralnog statusa i stavljene u kontekst moralnog statusa robota. Za potrebe prikaza četiri teorije moralnog statusa korištena je, u najvećoj mjeri, knjiga filozofkinje Marry Anne Warren, *Moral Status: Obligations to Persons and Other Living Things* (1997). Cilj rada bio je iznijeti argumente za i protiv četiri prevladavajuće teorije moralnog statusa i kroz primjere provjeriti je li, na osnovi neke od njih, moguće pripisivanje moralnog statusa robotima. Najvećim dijelom obrađene su dvije osnovne teorije, utilitaristička i deontološka, koje za ključni uvjet moralnog statusa uzimaju osjećajnost, odnosno osobnost. Nešto manje prostora posvećeno je teoriji o životu te teoriji o uključenosti u raznovrsne odnose kao jednako zanimljivim okvirima za provjeru pripada li nekom živom ili neživom objektu moralni status.

Ključne riječi: robotika, roboti, etika robotike, moralni status, moralno djelovanje

Summary

This master's thesis provides an outline of the basic theories of moral status that are put in the context of moral status of robots. As a main reference in the explanation of four theories, book by Marry Anne Warren *Moral Status: Obligations to Persons and Other Living Things* (1997) was used the most. The goal of the thesis was to put forward the main arguments for and against each of the four theories of moral status, as well as to consider whether any of them provides a solid basis for ascribing the moral status for robots. The largest part of the thesis takes on two basic theories of moral status, the utilitarian one and deontological one, which place sentience and personhood, respectively, as the crucial condition for moral status. Somewhat less space is devoted to the theory of life and the theory of being included in significant relationships as an interesting framework for testing whether some living or non-living object can have a moral status.

Key words: robotics, robots, robotics ethics, moral status, moral agency