

LOVRO GRGIĆ

Filozofski fakultet Sveučilišta u Zagrebu

Filozofija/Povijest, 4. godina

lgrgic@ffzg.hr

UDK: 007.52:004.8

17.026:007.52

Što ako androidi zaista sanjaju električne ovce?

U ovom radu pokušao sam ispitati implikacije koje bi znanstveno tehnička postignuća na području robotike, po nekima dohvatljiva u skoroj budućnosti, mogla imati na status robota i međuodnos robota i ljudi. Osnovna pitanja koja se u tekstu obrađuju su sljedeća: Što je dovoljno da se neko biće smatra živim? Koja prava proizlaze iz takvog statusa? Koji su preduvjeti da živo biće bude i moralno biće? Je li moguće da robot bude moralno biće? Kakav status zaslužuju bića s moralnim potencijalom? U susretu misaonog eksperimenta i etičke refleksije kao literatura su poslužili klasici znanstvene fantastike (Asimov, Dick, Roddenberry), i filozofi čija se moralna filozofija smatra općim mjestom (Aristotel, Kant, Rawls). Tradicionalni koncepti primjenjuju se na buduće probleme kako bi se dobila slika jednog mogućeg svijeta i njegovih moralnih normi. Kraj ostaje otvoren, odražavajući tako problematičnost početnih pretpostavki.

KLJUČNE RIJEČI
ROBOTI, ŽIVOT,
MORAL, UMJETNA
INTELIGENCIJA,
PRAVNI STATUS

POVOD S ISTOKA

Uz znatizeljlu, mašta je vjerojatno jedan od glavnih pokretača čovječjeg djelovanja, i kao takva visoko je vrednovana među ljudima. Kada se za nekoga kaže da je maštovit, to se, u principu, doživljava kao kompliment. Maštati se može o mnogim stvarima; neke se smatraju mogućima, neke, pak, nemogućima. Kada mašta prijeđe tu granicu (pretpostavljeno) mogućega (ili dozvoljenog), ona prelazi u sferu fantazije i fantastičnog. I među fantazijama se mogu raditi rang-liste. Dok bi se na vrhu trebale naći razne fantazije koje su "opće dobro" u smislu da ih svi (ili barem većina) imaju (npr. seksualne fantazije), znanstvena fantazija (ili, kako se uvriježilo, "fantastika"), vjerojatno bi bila u donjoj polovici. Kao književni žanr, znanstvenu fantastiku (u daljnjem tekstu ZF) obično se podcjenjuje, tj. tretira kao šund literaturu bez umjetničke, a i bilo kakve druge vrijednosti. U većini slučajeva to i jest tako. Ipak, klasici ZF-a su polivalentno vrijedne knjige. Književnoteorijski dio na stranu, ideje koje su se znale pojaviti u ZF-u bile su često vrlo zanimljive. Kao posljedica razvoja znanosti i tehnike u devetnaestom stoljeću pojavio se i žanr koji je popularizirao i eksploatirao duh epohe. I dok je znanstveno-tehnički razvoj utjecao na žanr ZF-a, i žanr je, povratno, utjecao na sam razvoj, barem kao poticaj, ono "golicanje" mašte koje vuče čovjeka naprijed. Mnoge ideje su se prvi puta pojavile upravo u pričama, a tek su kasnije (ponekad i puno kasnije) bile ozbiljene pomoću znanosti i tehnike. Ideje, formirane s one strane mogućega, u sferi fantazije, "uvedene" su natrag u svijet zbilje. Gotovo da se može kazati kako im se ozbiljenjem promijenio ontički status. Znanstvena fantastika reducirana je natrag na znanost.

Veliki pisci ZF-a se, međutim, nisu ograničili na znanstveni aspekt. Upravo suprotno! Postulirana postignuća znanosti često su služila da se pokaže kako bi u takvim, putem znanosti i tehnike izmijenjenim uvjetima postojanja, funkcionirao čovjek.¹ Pored čovjeka, tu se pojavljuju i razna druga bića, a upravo je njihova međusobna interakcija najznačajniji moment. Danas smo u situaciji da i interakcije o kojima se "fantazira" u okviru ZF-a najavljuju svoj skori prelazak u zbilju, u sferu zbiljski mogućeg. Jedna vijest koju su 7. svibnja 2007. prenijele značajne novinske agencije iz cijelog svijeta kao da je najava ovog ozbiljenja.

Ministarstvo trgovine Južne Koreje izdalo je priopćenje u kojem najavljuje osnivanje dvanaesteročlane komisije čiji će zadatak biti iznalaženje etičkog

1 Neka prošlo vrijeme upotrijebljeno u govoru o ZF-u ne bude krivo shvaćeno. Naravno da je ona kao žanr i danas postojeća. Ipak, autori uzeti u obzir za ovaj rad pripadaju klasicima, i u tom smislu tradiciji, dakle prošlosti. Teme kojima su se oni bavili danas postaju aktualne u drugom kontekstu. Pretpostavljam da teme današnjih ZF priča neće još dugo postati aktualne u tom smislu. Kakva bi to fantazija bila koja bi već sutradan postala stvarnost!

kodeksa za robote!² “Očekujemo da će uskoro doći dan kada će roboti s intelektualnim potencijalom moći djelovati na temelju vlastitih odluka. Stoga predstavljamo ovo kao pokušaj davanja etičkih smjernica o ulozi i sposobnostima robota”, izjavio je predstavnik ministarstva (Reuters).³ Najavljena je i mogućnost da se u izradi etičkog kodeksa kao predložak uzmu tri zakona robotike koja je u svojim pričama postavio znameniti autor ZF-a Isaac Asimov (Asimov 1976). Na čelu tijela kojeg sačinjavaju znanstvenici, odvjetnici, doktori i psiholozi nalazi se profesor Kim Dae-Won sa Sveučilišta u Myongjiu: “Postavljamo pravila koja će odrediti granice razvoja robotike, kao i način na koji će ljudi živjeti zajedno s robotima. Društvo u kojem roboti i ljudi žive zajedno može nastupiti prije nego se nadamo, vjerojatno unutar deset godina. Robotska etika dio je ljudske etike. Svrha ovog kodeksa je iznalaženje načina koegzistencije ljudi i robota, a ne ograničavanje razvoja robotike” (IOL technology).⁴ Naglasak će biti stavljen na zadržavanje ljudske kontrole nad robotima, sprečavanje ilegalne uporabe robota (za neželjene ili opasne svrhe), zaštitu robotski prikupljenih podataka, osiguravanje njihovog jasnog identificiranja i mogućnosti da ih se pronađe (da im se uđe u trag).⁵ U svjetlu ovakvih događanja čini se zanimljivim upustiti se u samostalni pokušaj razrade jednog etičkog kodeksa za robote.⁶ On bi obuhvaćao eventualna prava i/ili obveze robota u odnosu na ljude i druge robote, ali i prava ljudi u odnosu na robote. Nužno je istražiti na osnovu kojih odlika bi roboti mogli ili trebali uživati određena prava. Druga tema koja se nameće je eventualna moralnost samih robota, tj. pitanje može li robot biti moralni subjekt, sa svim posljedicama takvog statusa. Vrijedni sugovornici u ovakvom nastojanju su, naravno, pisci ZF-a, uz pokojeg filozofa.

ČEMU
8.16/17
OŽUJAK
2009.

² Nije slučajnost da se ovakav događaj zbio upravo u Južnoj Koreji. Riječ je o trećoj po redu zemlji u razvoju robotike (odmah iza SAD-a i Japana) i šestom po redu proizvođaču robota na svijetu. Korejska vlada planira robotiku učiniti glavnim sredstvom privrednog razvoja. Sama država ulaže oko sto milijuna dolara godišnje u razvoj robotike, a procjenjuju da će do 2013. u svakom korejskom kućanstvu biti po jedan robot. Zadaće koje im se namjenjuju su iznimno raznolike, od brige za djecu i nemoćne, predavanja u školama, sve do čuvanja granice.

³ *South Korea plans code of ethics for robots*, na <http://www.reuters.com/articlePrint?articleId=USSEO1665712007057> (14. 8. 2007.). Moj prijevod s engleskog originala.

⁴ *SKorea writing code of ethics for robots*, na http://www.ioltechnology.co.za/article_print.php?iArticleId=5019263 (14. 8. 2007.). Moj prijevod s engleskog originala.

⁵ Za robote koji će biti korišteni u vojne svrhe predviđa se poseban kodeks. U trenutku pisanja ovoga rada nije bilo nikakvih daljnjih informacija o vojnom kodeksu.

⁶ Kako bi se izbjegli eventualni terminološki nesporazumi, potrebno je odmah na početku strogo definirati pojmove. Robot je svaka mehanička multifunkcionalna naprava koja se može programirati za različite operacije. Samu riječ "robot" prvi je upotrijebio češki pisac Karel Čapek 1920. godine. Droid je robot s umjetnom inteligencijom. Android je čovjekoliki droid (Anić et al. 2004.). Ove općenite distinkcije dobro je imati pred očima; međutim, one nisu presudne. Bitno je da je uvijek riječ o biću (u najširem značenju te riječi), stvorenom od strane čovjeka (ili bilo kojeg drugog živog bića), a putem tehnike omogućene znanstvenim spoznajama. Materija od koje je sačinjeno ne mora nužno biti anorganska (ili pretežno anorganska), ali momenti poput kloniranja ne ulaze u vidokrug ove teme.

JA, ROB(OT)

Ne čudi da je Južna Koreja najavila mogućnost “posuđivanja” zakona robotike od Isaaca Asimova. Ako se na robota, bez obzira na njegove eventualne osobine, gleda kao na oruđe tek nešto sofisticiranije od čekića, tri zakona koja se pojavljuju u Asimovljevoj zbirci pripovijedaka *Ja, robot* čine se idealnima: “1.) Robot ne sme da povredi ljudsko biće ili da svojom neaktivnošću dozvoli da ljudsko biće bude povređeno. 2.) Robot mora da se povinuje naređenjima dobijenim od ljudskih bića, izuzev u slučaju ako su ta naređenja u suprotnosti s Prvim zakonom. 3.) Robot mora da štiti sopstveno postojanje sve dok takva zaštita nije u suprotnosti s Prvim ili Drugim zakonom” (Asimov 1976: 7). Zakoni su ustrojeni po principu algoritma. Vrijednost im je hijerarhizirana, pa se tako uvijek pri odluci za djelovanje mora prvo u obzir uzeti Prvi zakon; tek ako je ovaj ispunjen prelazi se na djelovanje u skladu s Drugim, a isti princip vrijedi i za Treći.

Asimov se u samim pričama majstorski poigrava upravo trima zakonima i načinom na koji roboti djeluju u skladu s njima. Pojavljuju se zanimljive teze. U jednoj od priča, robopsiholog iznosi tvrdnju kako je zapravo nemoguće razlikovati robota od najboljeg ljudskog bića jer su tri zakona robotike ustvari rukovodeći principi većine etičkih sistema u svijetu.⁷ Treći zakon odraz je ljudskog nagona za samoodržanjem. Drugi zakon odraz je “dobrog” čovjeka koji ima društvenu svijest i osjećanje odgovornosti prema drugima, i koji uviđa nužnost poštivanja autoriteta, a za opću korist. Prvi zakon bila bi zapravo interpretacija one biblijske “ljubi bližnjega svoga kao samoga sebe” na način da je nužno izložiti sebe opasnosti da bi se spasilo drugoga.

Ipak, Asimov nije toliko naivan. Njegovi roboti su inteligentniji i snažniji od ljudi i toga su svjesni; posjeduju sposobnost učenja; izražavaju osobine poput znatiželje;⁸ imaju funkcionalne korelate emocija; posjeduju nagon za samoodržanjem. Unatoč svim ovim odlikama, robot je zapravo rob, a takvim ga čini Prvi zakon. On je u vječnom vlasništvu tvornice koja ga je proizvela, nema nikakva osobna prava, niti može posjedovati imetak. U skladu s Aristotelovom formulacijom, on je živo oruđe.⁹ Da bi se bilo osoba, time i moralni subjekt, po Aristotelu je nužna mogućnost kreposti kao okosnice ljudskog (su)života. Oni koji tu mogućnost ne posjeduju su ili životinje, ili barbari, ili robovi, dakle nisu

⁷ U ovom kontekstu je, naravno, riječ o čovjekolikom inteligentnom robotu, tj. androidu.

⁸ Upravo ova osobina okosnica je priče *Logika kartezijanskog okusa*. Robot pokaže znatiželju o svome porijeklu, ali ga ljudski odgovor kako su ga oni stvorili ne zadovolji. U potrazi za sigurnim temeljem za razmišljanje robot na kraju stiže do toga da zasigurno postoji pošto misli. Budući dakle postoji, nužno je da bude i neki uzrok njegovog postojanja. Međutim, uvidjevši da je po svemu superiorniji ljudima, zaključuje da ga oni nisu mogli stvoriti, držeći se pri tome kao aksioma tvrđnje kako nijedno biće ne može stvoriti biće bolje od sebe samog! Kao svog tvorca tada postulira Gospodara. Pretpostavivši jasno svjetlo strogog razmišljanja senzornim podacima koji mogu i prevariti, sve pojave koje mu se ne uklapaju u sistem odbacuje kao iluzije.

⁹ O kvaliteti života, tj. pitanju što jest živo(t), bit će riječi kasnije. Naravno, dobar dio Asimovljevih ljudskih likova ne doživljava robote kao živa bića.

ljudi u punom smislu te riječi. Samim time, ne priznaje im se status moralnog subjekta, niti sva prava koja iz njega proizlaze. Krepost je, kako se iščitava iz *Nikomahove etike* (Aristotel 1982.), stanje s izborom tj. promišljenom žudnjom za stvarima koje su u našoj moći, a na način na koji bi to učinila razborita osoba. Za prosudbu nečije kreposti ključna je mogućnost izbora, tj. moment slobode. Robot tu mogućnost nema. Prvi zakon (a i druga dva), ugrađen u njegovu matricu kao sastavni dio, oduzima mu opciju slobodnog izbora. On mora djelovati tako i tako, premda kao inteligentno biće vidi i druge opcije i njihove posljedice, a posjeduje i fizičku moć izvršavanja; robot, dakle, nije moralno biće, niti to može biti. Kao oruđe, on je kod Asimova predmet kao i svaki drugi, i ne posjeduje nikakva prava. S njim se, po potrebi, može raditi apsolutno sve. Na dovođenje robota u opasnost ili njegovo uništenje gleda se samo kao na materijalni gubitak za vlasnika, a ljudi prema robotima nemaju nikakvih obveza. Prema njima je, naravno, moguće razviti simpatije, ali one se načelno ne razlikuju od sentimentalne vezanosti za npr. komad pokućstva. Pitanje o emocijama, po prirodi vrlo zanimljivo, vodi od Asimova do jednog drugog osebujnog autora.

ANDROIDI I ELEKTRIČNE OVCE

Naslov rada očita je parafraza naslova znamenite ZF knjige Philipa K. Dicka *Sanjaju li androidi električne ovce?*¹⁰ Za razliku od Asimova, čiji su roboti uglavnom goleme metalne mašine, Dick čini svoje robote izrazito antropomorfnima. Činjenica da njegovi androidi nalikuju čovjeku (ne samo fizički) uvodi dodatnu nelagodu u pokušaj svođenja njihove egzistencije na status pukih mašina. Izuzev mozga, koji se izgrađuje i ugrađuje u zasebnim procesima, android je u potpunosti organsko biće, sve do koštane srži (koja se ipak razlikuje od ljudske).¹¹ Imaju sve fiziološke potrebe i procese poput ljudi; konzumiraju hranu, mogu stupati u spolne odnose, smiju se, tuširaju se, puše. Njihov mozak čini ih inteligentnima barem na nivou prosječnog čovjeka. Sposobni su naučiti čitati i pisati, a razvijaju i potrebu za čitanjem. Povrh toga, mogu iskusiti emocionalna stanja poput usamljenosti, boli i depresije.¹² Neki od njih čak niti nisu svjesni da su androidi (imaju ugrađenu umjetnu memoriju s lažnim sjećanjima na prethodni život).

Unatoč svemu tome, oni se ipak razlikuju od ljudi. Životni vijek im je četiri godine. Razlog tome je što androidi, za razliku od ljudi, nemaju prirodni sistem zamjene

ČEMU
8.16/17
OŽUJAK
2009.

¹⁰ Po knjizi je snimljen i kulturni klasik *Blade Runner* Ridleya Scotta iz 1982.

¹¹ Dick zapravo ne specificira od kakvog je materijala izgrađen njihov mozak, tako da je otvorena mogućnost da i on bude organski.

¹² Neki androidi se čak i drogiraju zbog depresije.

istrošenih tjelesnih stanica, tj. njihovog trajnog samoobnavljanja; jednostavno se istroše i prestanu funkcionirati. U međuvremenu ne stare i ne oboljevaju. Ne rađaju se, ne odrastaju (proizvode ih u željenoj dobi), niti mogu imati djecu. lako su u stanju doživljavati određene emocije, one su kod njih u pravilu kratkotrajne, te androidi lako zapadaju u nezainteresiranost: "... imala [je] više spoljašnje smirenosti, ali unutrašnja napetost i frenezija još su bili tu. Ipak, mračna vatra je jenjavala; životna sila je isticala, curila iz nje; toliko puta je to gledao kod drugih androida. Klasična rezignacija. Mehaničko, intelektualno prihvatanje onoga što istinski organizam – koji iza sebe ima dve milijarde godina pritiska da živi i da odmiče putevima evolucije – nikad ne bi mogao pomireno prihvatiti." (Dick 1984: 127). Glavni razlog koji doprinosi njihovom statusu je njihova nemogućnost iskustva empatije. Empatija je, po definiciji, "sposobnost razumijevanja značenja i značajnosti emocija i ponašanja druge osobe; uživanje u tuđe emocionalno stanje" (Anić et al. 2004: 144; sv. 3). U postapokaliptičnom društvu kakvog opisuje Dick, empatija prema svim oblicima života (tj. i drugim ljudima i svim ostalim životinjama) zalog je mirne budućnosti i opstanka. Kao takva, ona je ključna ljudska odlika. "Empatija, očigledno, postoji samo u ljudskoj zajednici, za razliku od inteligencije koja je u nekom stupnju zastupljena u svim životinjskim filama i klasama, čak i kod arahnida. Pre svega, empatijska sposobnost verovatno zahteva postojanje razvijenog grupnog instinkta; organizam – usamljenik, kao što je naprimjer pauk, ne bi imao korist od empatije; ...Empatijom će živo biće sklono skupljanju u stada, čovjek naprimjer, steći uvećanu sposobnost opstanka." (Dick 1984: 24). Uskraćeni za mogućnost empatičnosti, androidi su isključeni iz društva, tj. svedeni na predmet koji ima svog vlasnika. Ne smiju ništa posjedovati, ne smiju stupati u spolne odnose, ne smiju nositi oružje, niti se smiju samostalno kretati (tj. bez pratnje ili znanja ljudskog vlasnika). Štoviše, oni zakonski niti nisu živi. Njihov specifičan status u svijetu kojeg Dick "fantazira" valja detaljnije proučiti. Ako je kao glavna ljudska odlika postulirana empatičnost, jasno je zašto android ne može biti subjekt ravan čovjeku – u pravnom, a niti u bilo kojem drugom smislu. Međutim, ostaje nejasno zašto mu se ne priznaje status živog bića. Organske građe, android ima većinu funkcija koje se pripisuju živim bićima, a povrh toga posjeduje inteligenciju i kakvu takvu emocionalnost. Sve mu to ne garantira nikakav pravni status. Dok je ubijanje i nasilje prema ljudima i životinjama najstrože zabranjeno, androida se smije "povući iz prometa". Androidi se koriste za najteže poslove, dok su životinje toliko dragocjene i čuvane da ne rade ništa. Razumljivo je da su androidi stvoreni da budu korisni, ali nije jasno zašto ne bi mogli imati barem nekakva prava, poput npr. životinja u našem današnjem svijetu. Zanemarujući kontekst u koji ih Dick stavlja, čini se da njegovi androidi posjeduju

sve preduvjete da budu živa i moralna bića, a na temelju toga i da eventualno uživaju ista prava kao i ljudi. Iako se to može učiniti nepriličnim, ovdje u pomoć treba zazvati Kanta. "Dakle taj princip ćudorednosti, upravo zbog općenitosti zakonodavstva, koja ga čini formalnim najvišim odredbenim razlogom volje bez obzira na sve njegove subjektivne različnosti, um ujedno proglašava zakonom za sva umna bića, ukoliko uopće imaju neku volju, tj. neku moć da svoj kauzalitet određuju predodžbom pravila, dakle ukoliko su sposobna za djelovanje prema načelima, a po tome i prema principima a priori (jer samo ovi imaju onu nužnost, koju um zahtijeva za načela). **On se, dakle, ne ograničava samo na ljude, nego na sva konačna bića, koja su obdarena umom i voljom...**" (Kant 1974: 65; moj kurziv!). Odavde je jasno da svako umsko biće ima moralni potencijal. Problem je jedino može li se pokazati da androidi posjeduju um i volju. Prethodno je rečeno da posjeduju inteligenciju (Dick dodaje i sklonost k apstraktnom mišljenju), a za Kanta je to dostatno: "Biće pak, koje je sposobno za djelovanje prema predodžbi, jest **inteligencija** (umno biće), a kauzalitet je takvog bića prema toj predodžbi zakona njegova volja." (Kant 1974: 173)¹³ Prema tome, Dickov android je, unatoč tomu što je umjetno stvoren putem znanosti omogućene tehnike, moralno biće par excellence.¹⁴

S takvom bi se ocjenom zacijelo morao složiti i John Rawls, barem na temelju tvrdnji iznesenih u svome djelu *A Theory of Justice* (Rawls 1971.). Ako je android dovoljno inteligentan da je sposoban imati koncepciju vlastite dobrobiti (izražene racionalnim planom života), tada on može sudjelovati u zasnivanju moralnosti putem izbora moralnih normi u izvornom stanju, a time i biti moralna osoba.¹⁵

Dok se ovdje pokazalo da i umjetno proizvedeno biće, pod pretpostavkom da je moguće proizvesti umjetnu inteligenciju, može biti moralno biće, vrijedno je ispitati i kriterije koji kvalificiraju neko biće kao živo. Život se čini kao jedan od faktora koji bitno utječu na nečiji pravni status, a da bi ga se pomnije istražilo, nužno je *krenuti gdje niti jedan čovjek do sada nije išao...*

TRAŽENJE NOVIH OBLIKA ŽIVOTA

Popularna ZF serija *Zvezdane staze* (Roddenberry 1966.), u skladu sa svojim najavnim motom, "...tražiti nove oblike života i nove civilizacije, hrabro ići gdje niti jedan čovjek do sada nije išao!", često se znala upustiti u stvarno

¹³ Originalni kurzivi! Naravno, problem definiranja inteligencije vrlo je složen. Budući da ima puno različitih opisa, niti jedan nije opće prihvaćen.

¹⁴ Budući da ima barem ograničeni emocionalni potencijal, android je sposoban razviti i ono što Kant naziva moralnim osjećajem.

¹⁵ Kod Rawlsa, za razliku od ostatka kontraktualističke tradicije, nije riječ o zasnivanju društva ili države, već upravo morala, tj. općepriznatih moralnih normi.

zanimljiva spekuliranja. Jedna od epizoda, pod nazivom *Mjera čovjeka*, bavi se slučajem koji se direktno nadovezuje na prethodno poglavlje ove radnje. Radi se o sudskom procesu, čija je svrha odrediti pravni status jednog androida, odnosno, utvrditi je li on osoba ili privatno vlasništvo. Tužiteljstvo, kojem je cilj dokazati kako je riječ o privatnom vlasništvu, koristi sljedeće argumente: Android je automat stvoren da slični ljudskom biću; on je proizvod ljudskog uma i izgrađen je od strane čovjeka (i software i hardware), a svrha mu je služiti čovjeku. Kao takvo biće, android nema nikakva prava već se mora pokoravati naredbama vlasnika. Federacija Planeta (velika država u kojoj se događa većina zbivanja u seriji) garantira pravni status svim bićima koja su u stanju zamjećivati i osjećati.¹⁶ Obrana nastoji pokazati da android spada u tu skupinu i da na temelju toga ima zagwarantirana prava kao i ostala bića iz nje; u ovom slučaju, da ima pravo ne pokoravati se naredbama proizvođača. Kriteriji koje Federacija postavlja su jasni: da bi se biće smatralo subjektom (dakle, i pravnim subjektom), ono mora posjedovati inteligenciju, svijest i samosvijest. Android u pitanju je definitivno inteligentan jer je sposoban učiti i snalaziti se u novim situacijama. Aktivno reagira na izvanjske podražaje i komunicira s okolinom, što su odlike svjesnog bića. Kao kriterij za samosvijest određuje se svijest o vlastitom postojanju i djelovanju. Kako je android očito svjestan uloga koje on i ljudi imaju na suđenju, i dovoljno je zabrinut za vlastito postojanje da se inicijalno odlučio usprotiviti potencijalno opasnom naređenju, čini se da posjeduje i samosvijest.¹⁷ Na temelju toga donesena je presuda kako android mora uživati sva prava kao i ljudi i ostala bića iz te skupine.¹⁸ Riječ je, u stvari, o pojavi nove rase, novog oblika života, čije bi podčinjavanje značilo zapravo porobljavanje, a to je nedopustivo. Premda je ovaj android biće od anorganskih materijala, njegove odlike (u koje spada i inteligencija nadmoćna ljudskoj, kao i mogućnost verbalnog komuniciranja) nedvojbeno mu pružaju zasluženi pravni status. Puno veći problem javlja se pri susretu s nečim manje nalik na čovjeka.

Upravo takav slučaj opisuje se u epizodi *Kvaliteta života*. Jedan mali nehumanoidni robot, dizajniran da obavlja različite popravke, dolazi u središte pažnje kada ga se počne smatrati potencijalno živim! U općenitoj definiciji koju u seriji daje lik doktorice, život je ono što omogućava biljkama i

¹⁶ Termin koji se pojavljuje u originalu je *sentient being*, ali njega je nemoguće direktno prevesti na hrvatski.

¹⁷ Samosvijest bi se mogla opisati kao svijest o svijesti, tj. o svojim svjesnim stanjima. Osnovna je teza da životinje, iako svjesna bića, nisu svjesne toga da su svjesne. I kod ljudi su trenuci samosvijesti relativno rijetki; bilo bi krajnje neekonomično da svaki podražaj analiziramo na način da sami sebi kažemo: "Jesi li svjestan toga da upravo sada gledaš predmet koji je zelene boje?". Nevolja sa (samo)sviješću je u nemogućnosti njenog elegantnog definiranja; stoga preostaju samo neelegantni opisi poput ovog. Kao nadopunu tvrdnji o samosvijesti, android ističe da njegova iskustva imaju kvalitnost, tj. da njegova memorija nije samo puk zbir činjenica, već da one (činjenice) za njega imaju posebnu kvalitetu subjektivnog dojma. Povrh zabilješke o tome da je u tom i tom trenutku na tom i tom mjestu ugledao specifičan predmet određene nijanse zelene boje, njegova memorija bilježi i njegovo osobno iskustvo gledanja (i viđenja) tog predmeta, dojam, koji je samo njegov, i koji ne može biti iskaziv u znanstveno strogo prihvatljivim kategorijama i kao takav zabilježen. Takva tvrdnja je, naravno, nedokaziva, ali time i neoboriva.

¹⁸ Svi argumenti iz Kanta i Rawlsa vezani uz raspravu o androidu kakvog opisuje Filip K. Dick primjenjuju se i na ovaj slučaj.

životinjama da konzumiraju hranu, izvlače iz nje energiju, rastu, razmnožavaju se i prilagođavaju okolišu. Sve što zadovoljava ove uvjete (ili barem većinu njih) smatra se živim.¹⁹ Robot u pitanju ima sposobnost učenja, tj. mogućnost da svaki obavljeni popravak (problem i njegovo rješenje) pohrani u memoriji, kako bi ubuduće bio učinkovitiji. Pri izvršenju zadata robot je pokazao začuđujuće odlike: reprogramirao bi postavke koje su mu ljudi zadali kako bi efektivnije izvršio dobiveni zadatak; nije izvršio za sebe potencijalno opasnu radnju (spalio je svoje upravljačko sučelje kako ga se ne bi moglo teledirigirati, a poslije ga je sam popravio); u grupnom zadatku, jedan od robota je ostao obavljati svoju zadaću pod cijenu vlastitog uništenja kako bi se ostali roboti mogli sigurno evakuirati.²⁰ Postavlja se pitanje: je li biće koje pokazuje takve odlike živo? Ono definitivno "konzumira hranu i izvlači iz nje energiju". Štoviše, za razliku od čovjeka, ono je u prednosti jer preskače jedan suvišan korak i konzumira direktno energiju (pretpostavljeno električnu). Prilagođava se okolišu jer može modificirati postavke obavljanja dobivene zadatke ovisno o okolnostima. Povrh toga, svojim pokušajem spašavanja samog sebe od uništenja pokazalo je svijest o svom okruženju i sposobnost da se tom okruženju prilagodi, barem na razini na kojoj to demonstriraju mnoge životinjske vrste. Kao stroj opremljen raznim alatima, ovo bi biće čak bilo u stanju proizvoditi nove jedinke na svoju sliku i tako se razmnožavati. U skladu s prethodno navedenim odlikama, ono bi se moglo smatrati živim! Za razliku od androida iz prethodnog slučaja, ono nije inteligencijom na ljudskoj razini (naravno, niti iznad nje) i ne može komunicirati verbalno, ali je ipak živo. Temeljem toga, ono, vjerojatno, ne bi moglo uživati sva prava kao ljudi, androidi, i ostala bića tog ranga (ne bi npr. moglo biti kapetan broda), ali bi trebalo imati barem neka prava. Federacija zabranjuje da se živa bića iskorištavaju kao radnici, odnosno, da se prisiljavaju na rad. Ona mogu biti korištena za rad, ali ne i eksploatirana. To bi značilo da takav robot, dakle anorgansko, putem tehnike, a pomoću znanosti omogućeno, ali unatoč tome živo biće, u Federaciji ima pravo uživati status barem jednak onome organskog živog bića u našem, današnjem svijetu. Jedan konj, npr., nema slobodu kretanja, ne može posjedovati vlasništvo, glasati na izborima ili biti sveučilišni profesor; može biti korišten (konj, ne sveučilišni profesor; sic!) za oranje ili jahanje bez svoje izričite privole. Ipak, pri tome ne smije biti eksploatiran ili zlostavljan; ne smije biti iznurivan do smrti; vlasnik ga je dužan hraniti i brinuti se o njemu (npr. zdravstveno); ne smije ga se tući ili oštro fizički kažnjavati. Za razliku od konja, jedan traktor, koji nije živ, može biti korišten do mile volje. Kada se ne ponaša onako kako bi vozač želio, traktor može biti

ČEMU
8.16/17
OŽUJAK
2009.

¹⁹ Za usporedbu se mogu razmotriti odlike koje navodi A. Meyer, a prenosi poznati antropolog Helmuth Plessner: "... A. Meyer u svojoj već citiranoj *Logici Morfolgije* (Berlin 1926) [daje] jedan sažetak posvuda prihvaćenih životnih oznaka: prehrana (izmjena tvari) – razmnožavanje – razvoj – nasljeđivanje – rast – podraživost – regulacija – kretanje (izmjena energije) – struktura." (Plessner 2004: 120)

²⁰ Suvišno je posebno napominjati da se ovaj robot samostalno kreće, tj. da počelo svoga kretanja ima u sebi samom.

“premlaćen” i “izvrijeđan” bez težih posljedica (barem po traktor). Iako se koriste za istu svrhu, konj je živ, a traktor nije. Slična distinkcija može se povući između gore spomenutog robota i današnjeg tvorničkog robota koji npr. slaže automobile na traci. Iako su oba anorganska, jedan od njih je živ, i na temelju toga zaslužuje određena prava, koja drugi nema.

RODDENBERRY PROTIV ASIMOVA

Pri pokušaju sažimanja do sada iznesenog, kao najveći problem nameću se sadržaji pojmova *život* i *inteligencija*. I dok u današnjem svijetu vlada prešutni konsenzus oko toga da sva živa bića imaju barem nekakva prava zagwarantirana pozitivnim zakonima, a na temelju raznih moralno vrijednosnih sustava (kršćanskog, hinduističkog...), pitanje je što se sve može okarakterizirati kao živo.²¹ Ako bi ljudi uspjeli putem tehnike (makar i nehotično) proizvesti biće koje bi zadovoljavalo značajan dio gore navedenih životnih značajki, imali bi dvije opcije: promijeniti definiciju života i iz nje “izvlastiti” potencijalno žive robote²², ili priznati ih kao živa bića i dati im na temelju toga prava, barem na razini koju danas uživaju neke životinje. Prihvatanje drugog slučaja značilo bi da se takve robote vjerojatno ne bi smjelo koristiti u po život (njihov život!) opasnim situacijama. Budući da je jedna od svrha pravljenja naprednih humanoidnih robota upravo ta da oni umjesto čovjeka obavljaju opasne zadatke, jasno je zašto se vlada Južne Koreje odlučila za pozivanje na Asimovljev koncept, u kojem roboti nemaju nikakva prava.

U suprotnosti s Asimovljevim je svijet kakvog fantazira Gene Roddenberry u svojim *Zvezdanim stazama*. Tamo je svaki život šticećen, a svaka inteligencija poštovana. Ako pretpostavimo da će jednom biti moguće proizvesti umjetnu inteligenciju (što je ipak vrlo problematična pretpostavka), čini se da iz toga proizlazi da će takvo biće, barem potencijalno, biti i moralno biće, subjekt istih

ČEMU
8.16/17
OŽUJAK
2009.

21 Činjenica je da mikroorganizmi poput bakterija i virusa (granični slučaj između živog i neživog), ili od njih nešto većih insekata, u principu ne uživaju nikakva prava na temelju toga što su živi. Ipak, postoje sustavi moralnih normi koji smatraju da su i oni jednako vrijedni i imaju jednaka prava kao i sva ostala živa bića. Jedan takav je npr. religijski sustav džainizma.

22 Ta, izmijenjena definicija, mogla bi zvučati kao ona koju je početkom dvadesetog stoljeća, ne predviđajući razvoj robotike, dao W. Roux: “Živa bića ... su prirodna tijela koja se najmanje jednim zbrojem određenih, elementarnih funkcija, koje direktno ili indirektno služe samoodržanju ... kao i samoregulacijom ... razlikuju u vršenju svih ovih funkcija od anorganskih prirodnih tijela...” (Plessner 2004: 120). Kao protuargument može poslužiti tvrdnja zastupnika teorije višestruke ostvarivosti, koji naglašavaju da su za procese bitne uloge koje oni obavljaju, a ne načini na koje ih oni obavljaju. Na nivou osjeta boli, nebitno je da li je ovaj ostvaren kao aktivacija c vlakana organskog sastava, ili se odvija u silikonskom (anorganskom) sklopu. Za bol je samo bitno da je ono stanje koje omogućuje organizmu da bude svjestan vanjske povrede i da na nju reagira. Kao zorni primjer može se uzeti hobotnica. Iako je njen živčani sustav drastično različit od ljudskog, velik broj ljudi smatra kako i hobotnica može iskusiti bol; ako po ničemu drugom, to se može zaključiti po njenim reakcijama na potencijalno bolne podražaje. Ako načelno prihvaćamo da i vrlo različit živčani sustav od onog ljudskog može poslužiti kao nužni preduvjet za iskustvo boli (a čini se da nam primjer s hobotnicom to sugerira), onda se i opcija s anorganskim sustavom kao takvim nužnim preduvjetom može činiti barem mogućom.

moralnih normi i etičke refleksije kao i mi.²³ Kantovi i Rawlsovi navodi u tom smjeru nikako ne predstavljaju pervertiranje njihovih učenja, već jednostavno mogućnosti koje ovi nisu uzeli u obzir, ali koje potencijalno proizlaze iz njihovih koncepata. Biću koje je na intelektualnom i moralnom planu jednako nama uskratiti da uživa ista prava kao i mi bio bi jedan od najvećih rasističkih ispada u ljudskoj povijesti (a konkurencija je tu žestoka).

ČEMU
8.16/17
OŽUJAK
2009.

23 Treba još jednom naglasiti kako nema čvrste definicije pojma inteligencije. Opisno i općenito, inteligencija je ono što bićima omogućuje snalaženje u novim situacijama i otkrivanje pravilnost u zbivanjima kao nužnog sredstva opstanka i eventualnog napretka. U ovom kontekstu, inteligenciju i treba shvatiti tako, tj. opširno i općenito. Problem umjetne inteligencije užasno je kompliciran i višeslojan, te zahtjeva opširniju razradu na drugom mjestu; ovdje je moguće samo ukazati na pojedine aspekte problema. Oni koji tvrde da je moguće proizvesti umjetnu inteligenciju (i da će to kad-tad biti učinjeno), načelno zastupaju tezu da je mentalni potencijal svediv na tjelesni potencijal, tj. da su mentalna stanja zapravo stanja u kojima se nalaze određeni fizički entiteti (dakle, da je fizička podloga ne samo nužan, već i dovoljan preduvjet mentalnih zbivanja). Sve što nas u tome trenutno sprječava je nedostatan nivo na kojem se nalaze znanost i tehnika. Njihovim napretkom bit ćemo u mogućnosti da proizvedemo sve bitne fizičke entitete i posložimo ih na način koji će nam omogućiti stvaranje umjetne inteligencije. Jedan od mogućih povoda da se nešto proglasi umjetnom inteligencijom jest da to nešto položi tzv. *Turingov test*. Da bi npr. jedno računalo prošlo Turingov test, ono bi moralo moći komunicirati s čovjekom putem ekrana tako da ta osoba nakon određenog vremenskog perioda (minimalno pet minuta) još uvijek misli da putem veze komunicira s drugom osobom, a ne s računalom. Takav test može se učiniti proizvoljnim, ali je zapravo intuitivno dosta prihvatljiv. Za sada niti jedno računalo nije uspjelo proći Turingov test. Oni koji drže da je umjetnu inteligenciju načelno nemoguće proizvesti, smatraju da ono mentalno nije moguće svesti na ono fizičko, tj. da nikakvo proizvođenje fizičkih entiteta i njihovo dovođenje u određenu vezu nije dovoljno da se proizvede umjetna inteligencija. Ona nije neproizvodiva zbog trenutnog stanja znanosti i tehnike; nikakav napredak na tom polju neće je omogućiti. I citat iz *Kritike praktičkoga uma* (Kant 1974) može u ovom kontekstu biti vrlo problematičan. Sadržaji pojmovi *inteligencija* kod Kanta i (*umjetna*) *inteligencija* u kontekstu suvremene rasprave nisu nužno isti. S obzirom na transcendentalni identitet onoga *Ja*, Kant se vrlo vjerojatno ne bi složio s mogućnošću njegova proizvođenja dovođenjem određenih fizičkih entiteta u specifičnu vezu. Oni koji smatraju da je umjetnu inteligenciju moguće proizvesti, zadržat će mogućnost da ona (umjetna inteligencija) djeluje prema nekoj predodžbi pravila i utoliko ima moralni potencijal; iako ta pravila nemaju apriornu nužnost, biće s umjetnom inteligencijom ipak bi moglo željeti da njegove maksime postanu opći zakoni, i time bi zadovoljilo "zlatno pravilo" koje je u temelju većine moralnih kodeksa (Ne čini drugima ono što ne želiš da drugi čine tebi). Problem je, u krajnjoj liniji, ontološkog karaktera, i može se svesti na dva pitanja: a) Na koji način inteligencija jest (postoji)?, i iz njega izvedivo početno b) Je li tako postojeću inteligenciju moguće proizvesti, ili ona nije proizvodiva, već naprosto jest?

LITERATURA:

"South Korea plans code of ethics for robots", <http://www.reuters.com/articlePrint?articleId=USS EO1665712007057> [14. 8. 2007.]

"South Korea writing code of ethics for robots", http://www.ioltechnology.co.za/article_print.php?iArticleId=5019263 [14. 8. 2007.]

ANIĆ, VLADIMIR, et al. (2004). "Hrvatski enciklopedijski rječnik", (12 svezaka), Zagreb: EPH, Novi Liber.

ARISTOTEL (1982). "Nikomahova etika", Zagreb, FPN.

ASIMOV, ISAK (1976). "Ja, robot", Beograd: Izdavački zavod "Jugoslavija".

DIK, FILIP K. (1984). "Sanjaju li androidi električne ovce?", Beograd: Samostalno autorsko prevodilačko izdanje Branislava Brkića, Aleksandra B. Nedeljkovića, Zorana Rapajića i Branka Đakovića.

FRAKES, JONATHAN (1992). "Star Trek: The Next Generation"; epizoda 135, "The Quality of Life", Paramount Television.

KANT, IMMANUEL (1974). "Kritika praktičkog uma", Zagreb: Naprijed.

PLESSNER, HELMUTH (2004). "Stupnjevi organskoga i čovjek", Zagreb: Naklada breza.

RAWLS, JOHN (1971). "A Theory of Justice", Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

SCHEERER, ROBERT (1989). "Star Trek: The Next Generation"; epizoda 035, "The Measure of A Man", Paramount Television.

LOVRO GRGIĆ

What if Androids Really Do Dream of Electric Sheep?

SUMMARY

In this paper, the author has tried to explore the possible implications, which scientific and technical breakthroughs in the field of robotics (held by some to be at hand in the near future) could have on the status of robots and the relations between robots and humans. The main problems addressed are the following: What are the pre-conditions of accepting an entity as a living being? What legal status should a creature, which is considered alive have? What are the pre-conditions of being a moral being (subject)? Is it possible for a robot to be a moral being? What status should a being, which meets all pre-conditions of being a moral subject, have? In an attempt to present an answer to these questions, the author has combined hypothetical reasoning with the theory of ethics, drawing from both science fiction classics (Asimov, Dick, Roddenberry) and moral philosophy milestones (Aristotle, Kant, Rawls) in the process. In doing so, he tried to apply traditional concepts on future problems, which resulted in forming a picture of a possible world, along with its moral code. The paper ends without a decisive conclusion, thus reflecting the uncertainty of the premises with which it began.

KEYWORDS

Robots, life, morality, artificial intelligence, legal status

ČEMU
8.16/17
OŽUJAK
2009.