

Pitanje identiteta umjetne inteligencije

Majčica, Ana

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zadar / Sveučilište u Zadru**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:162:796283>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-09-19**



Sveučilište u Zadru
Universitas Studiorum
Jadertina | 1396 | 2002 |

Repository / Repozitorij:

[University of Zadar Institutional Repository](#)



zir.nsk.hr

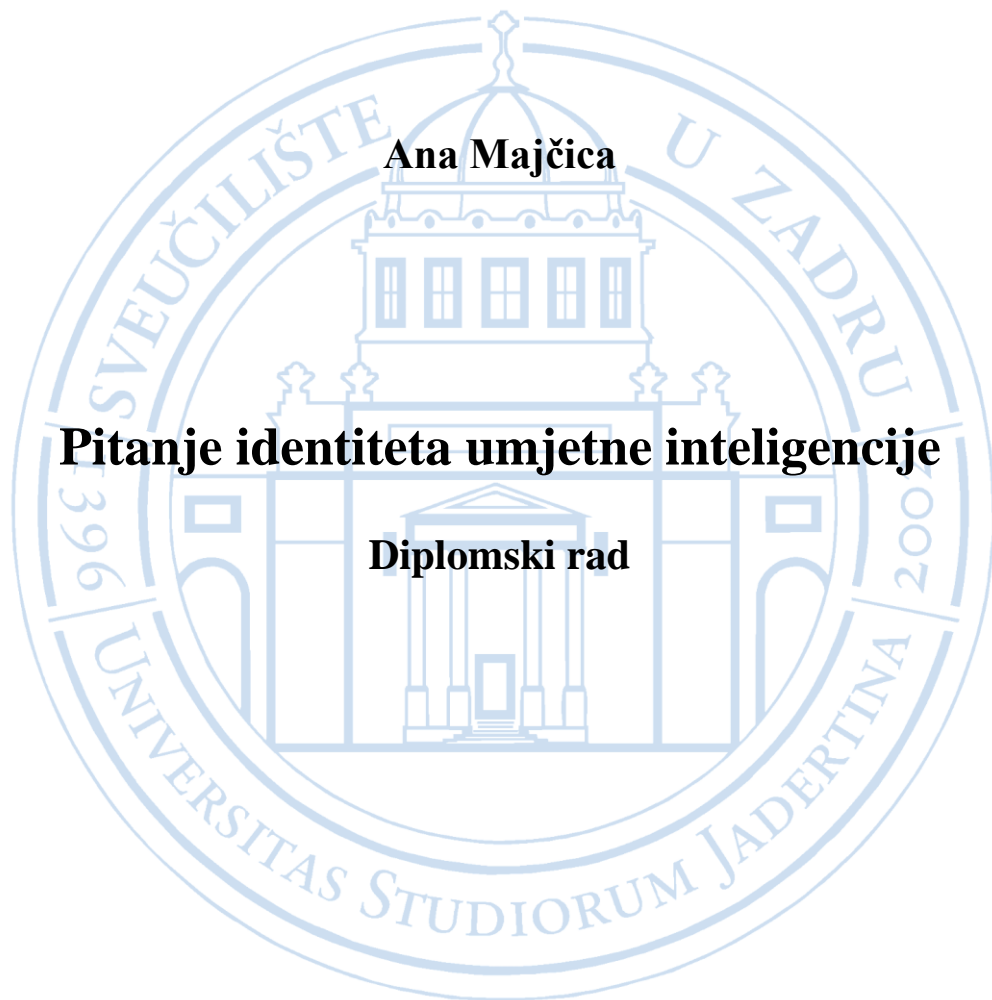


DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJ

Sveučilište u Zadru

Odjel za filozofiju

Odjel za filozofiju



Ana Majčica

Pitanje identiteta umjetne inteligencije

Diplomski rad

Zadar, 2021.

Sveučilište u Zadru

Odjel za filozofiju
Odjel za filozofiju

Pitanje identiteta umjetne inteligencije

Diplomski rad

Student/ica:

Ana Majčica

Mentor/ica:

izv. prof. dr. sc. lic. theol. Marko Vučetić

Zadar, 2021.



Izjava o akademskoj čestitosti

Ja, **Ana Majčica**, ovime izjavljujem da je moj **diplomski** rad pod naslovom **Pitanje identiteta umjetne inteligencije** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 30. ožujka 2021.

Sadržaj	
Sažetak	30
Summary	30
UVOD	5
1.UVOD U SVIJET UMJETNE INTELIGENCIJE	7
<i>1.1.ŠTO JE INTELIGENCIJA?</i>	7
2. UMJETNA INTELIGENCIJA	8
<i>2.1. KARAKTERISTIKE BIHEVIORALNIH I AUTONOMNIH ROBOTA</i>	12
<i>2.3. POPULARIZACIJA UMJETNE INTELIGENCIJE (UI)</i>	16
3. IDENTITET	18
<i>3.1. PITANJE IDENTITETA UMJETNE INTELIGENCIJE</i>	21
4. POJAM IDENTITETA	23
ZAKLJUČAK	27
Literatura:	28

UVOD

Kad se tek počelo s istraživanjem i razvitkom umjetne inteligencije (u daljnjem tekstu UI) krenulo se s pretpostavkom da ona ima potencijal da poboljša razne aspekte ljudskog života i time nam olakša svakodnevnicu. Sada živimo u vremenu gdje smo svjedoci da je UI ispunila svoj prvobitni potencijal. Bilo da je riječ o UI u pametnim telefonima, uređajima u automobilima ili nečemu trećem ona je svuda oko nas bili mi svjesni toga ili ne.

U posljednje vrijeme, razvoj UI dosegao je novu razinu koja daleko nadmašuje početne pretpostavke o njenom potencijalu. Stvaranje robota s inteligencijom koja parira ljudskoj se ostvarilo kada je svijetu predstavljena Sophie, robot koji je izgledom poput čovjeka s UI koja je prošla sve testove koji su stavljeni pred nju. Sophie je sposobna proizvesti nešto slično svom vlastitom iskustvu stvari zahvaljujući svojoj mogućnosti prilagodbe situacijama u kojima se nalazi. Razvoj ovakve vrste inteligencije dovodi nas do nove generacije UI koji bi mogla biti sposobna svoju inteligenciju samostalno razvijati.

Tako dolazimo do onog dijela UI koji se treba promotriti i s filozofskog aspekta, a to je pitanje identiteta. Kroz povijest, ljudi se nikada nisu susreli s inteligencijom koja je ravnopravna ljudskoj i koja bi ju mogla nadmašiti. Tako pitanje UI postaje pitanje egzistencije. Inteligencija koja ne pripada prirodnoj a simulira prirodnu inteligenciju je ono što kod ljudi stvara nelagodu. Kakav identitet pripisati nečem umjetnom što nam parira na inteligentnoj razini, a u budućnosti će je, vrlo vjerojatno, i nadići. S filozofskog stajališta situacija koegzistiranja dvaju ravnopravnih inteligencija nameće pitanje identiteta te druge inteligencije. Kad jednog dana roboti poput Sophie postanu normalna pojava, hoćemo li s njima dijeliti i naš identitet ili će se pojaviti nova definicija identiteta za njih? Ovdje se ne radi samo o robotu s UI koja parira ljudskoj nego o robotu koji osim toga može i imitirati emocije.¹

Istraživači UI imaju za cilj razviti jaku umjetnu inteligenciju s emocijama kakva se danas može vidjeti samo u djelima znanstvene fantastike. Ako jednom dođemo do toga da UI pokazuje znakove samostalnog razmišljanja, treba se ispitati je li to samo imitacija stvaranja samostalne, neprogramirane misli ili se zbilja radi o načinu razmišljanja svojstvenog ljudima. U tom slučaju,

¹Usp. Rothman Joshua, *Daniel Dennett's Science of the Soul*, The New Yorker, 20.3.2017. Pristupljeno: 16.4.2020. 35. <https://www.newyorker.com/magazine/2017/03/27/daniel-dennetts-science-of-the-soul>

kako bismo mogli primijeniti Descartesovo *Cogito, ergo sum* na UI? Što je UI po pitanju identiteta?

Kad govorimo o identitetu, većinom se počinje od dualiteta tijela i uma. Međutim, takve teorije nisu primjenjive na svijet računala te ih ne možemo upotrijebiti za teoretiziranje o UI. Ili ipak možemo? Premda su McCharty i Hayes argumentirali da se samo nekolicina teorija prirodnog svijeta može primijeniti na računala, mogli bismo polemizirati o nekonvencionalnom dualitetu gdje se iz elektronskih spojeva razvila inteligencija koja će nadjačati onu prirodnu. U ovome radu pokušat ću dati odgovor na to pitanje. Počevši od razvoja UI, kroz definicije identiteta s filozofske perspektive ovaj rad vodi k pitanjima koja će u bližoj budućnosti razvoja UI postati aktualna i na koja bismo trebali imati odgovore kad ova situacija s razvojem jake UI iz svijeta znanstvene fantastike postane naša sadašnjost.

1. UVOD U SVIJET UMJETNE INTELIGENCIJE

1.1. ŠTO JE INTELIGENCIJA?

Alan Turing je 1950-ih godina svijetu predstavio Turingov Test. Radilo se o testu u kojem dva sugovornika pričaju, a gdje je jedan sugovornik čovjek te treba testirati je li njegov sugovornik isto čovjek ili umjetna inteligencija. Ovo je rezultiralo podijeljenim mišljenjima, ali filozofi su se složili oko jednoga – test ne dokazuje inteligenciju kao takvu, već mogućnost imitiranja iste bez svijesti. Kada govorimo o svijesti nailazimo na mnoge probleme jer postoji mnogo shvaćanja iste. Pri razumijevanju svijesti, filozofi koriste termin *qualia*. Taj termin označava introspektivan fenomenski aspekt naših mentalnih stanja.² Britanski filozof Colin McGinn tvrdi da ljudi nisu u mogućnosti shvatiti poveznicu koja veže mozak i *qualia*, te da nitko ne može spoznati kako nešto može imati svijest.³ Međutim, kod umjetne inteligencije riječ je upravo o svijesti koja se, umjetnim putem, razvije kod računala. Neke od stvari koje roboti već sad mogu napraviti, poput sviranja glazbenog instrumenta ili izvođenja medicinske operacije, zahtijevaju određene psihološke vještine ali se i dalje ne smatraju pokazateljima razvoja inteligencije na onaj način na koji je posjeduju ljudi i životinje. Premda je kod čovjeka, sposobnost sviranja glazbenog instrumenta pokazatelj određene razine inteligencije, kod računala je to samo niz naredbi koje su precizno izvršene zahvaljujući programu koji je prethodno instaliran. Inteligencija je spoj različitih sposobnosti procesuiranja informacija i djelovanja u skladu s time.⁴

Dva su cilja za razvoj inteligencije kod računala. Prvi cilj je tehničke prirode, a to je upotreba računala za obavljanje raznih zadataka. Drugi cilj je znanstvene prirode, a to je izrada modela koji će nam pomoći da odgovorimo na pitanja o nama samima.⁵

Pitanje svijesti kod UI se može problematizirati s dva pristupa. Prvi pristup se odnosi na mehaničku svijest (*machine consciousness*).⁶ Sintetička ili mehanička svijest je područje unutar kognitivne robotike. Cilj istog je pronaći što točno treba biti sintetizirano da bi se mogla stvoriti

² Usp. Online Stanford Encyclopedia of Philosophy. Pristupljeno: 3.5.2020. <https://plato.stanford.edu/>

³ Usp. Boden Margaret A., *Artificial Intelligence, a very short introduction*, Oxford University Press, United Kingdom, 2018., str. 209.

⁴ Usp. *ibid.*, str. 32.

⁵ Usp. *ibid.*, str. 33.

⁶ Usp. *ibid.* str. 210.

svijest u inženjerski stvorenoj stvari. Drugi pristup se bazira na istraživanju svijesti, ali bez izrade modela. UI na razini AGI bi osim svijesti pokazivala i samosvijest te promišljenost.

2. UMJETNA INTELIGENCIJA

Svijet filma i književnosti stvorio je sliku umjetne inteligencije koja je postala općeprihvaćena te, kada govorimo o istoj, većina ljudi će pomisliti na takvu vrstu inteligencije kako je opisana u žanru znanstvene fantastike. Međutim, sama ideja o umjetnoj inteligenciji seže daleko u ljudsku prošlost te ne pripada samo svijetu filma i književnosti nego i svijetu filozofije, biologije, inženjerstva, fizike i drugih.

Umjetna inteligencija je znanstvena disciplina koja se bavi oblikovanjem inteligentnih sustava koji oponašaju ona svojstva ljudskog ponašanja koja se smatraju inteligentnim. Inteligencija je termin koji pridajemo svakom neživom sustavu koji pokaže sposobnost snalaženja u novim situacijama i učenju na temelju iskustva. Prvi znanstveni rad koji se bavi UI napisali su Warren McCulloch i Walter Pitts. Oni su kombinirali tri područja za objašnjenje UI – psihologiju i funkcije neurona u mozgu, propozicijsku logiku te Turingovu teoriju komputacije. Prvo računalo s neuronskom mrežom, nazvano SNARC, su 1950. godine izradili Marvin Minsky i Dean Edmunds.

Termin „umjetna inteligencija“ izmislio je John McCarthy 1956. godine tijekom događaja poznatog kao „Dartmouth konferencija“, na kojoj su također sudjelovali i Marvin Minsky, Claude Shannon i Nathan Rochester. Tijekom 1960-ih razvili su se prvi roboti, poput Elize, koji su mogli razgovarati s ljudima. Kada govorimo o umjetnoj inteligenciji, ljudi najčešće zamišljaju takvog robota. To nije ispravno razmišljanje jer robot je samo spremnik za umjetnu inteligenciju koji može, a i ne mora, oponašati ljudski izgled. Umjetna inteligencija je kompjuter unutar robota ili slikovitije rečeno, umjetna inteligencija je mozak a robot je tijelo. Međutim, veliki postotak umjetne inteligencije uopće ne posjeduje tijelo. Najveća eksplozija u svijetu UI dogodila se između 2000. i 2010. kada se UI počela ugrađivati u igračke, automobile, mobitele i ostalu elektroniku.⁷

⁷ Usp. Urban Tim, *The AI Revolution: The Road to Superintelligence*, 22.1.2015., Pristupljeno; 12.5.2020., 1. <https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html?fbclid=IwAR0ZpwdDpy->

John Searle uveo je podjelu teorija UI na „jake“ i „slabe“ teorije. Prema „slabim“ teorijama kompjuteri služe samo kao pomoć pri istraživanju prirodne inteligencije te sam ne može imati istu. Dok s druge strane prema „jakim“ teorijama kompjutor može imati inteligenciju i biti inteligentan u doslovnom smislu te riječi.⁸ Grana znanosti koja se bavi razvojem UI za cilj ima izgradnju i razumijevanje umjetnih bića na onaj način kako se mi ponašamo i razumijemo sami sebe.⁹

“Umjetna inteligencija propituje jednu od konačnih zagonetki. Kako je moguće da spor, maleni mozak, biološki ili elektronički, može percipirati, razumjeti i predviđati svijet, te manipulirati svijetom mnogo većim i mnogo kompleksnijim nego što je on? Kako da izgradimo nešto s takvim svojstvima? Ta su pitanja teška, ali, za razliku od putovanja brzinom većom od brzine svjetlosti ili anti gravitacijskog uređaja, istraživač na području umjetne inteligencije ima čvrste dokaze da je zadatak moguće ostvariti. Sve što trebamo učiniti je pogledati u zrcalo da bismo vidjeli primjer inteligentnog sustava”.¹⁰

Umjetna inteligencija se dijeli na tri razine. Prva razina je *artificial narrow intelligence* koja se odnosi na onu umjetnu inteligenciju koja je specijalizirana samo za jedno područje, npr. igranje šaha. Druga razina se naziva *artificial general intelligence* i odnosi se na inteligenciju koja je na razini ljudi, tj. ona koja može izvesti svaki intelektualni zadatak baš kao čovjek. Umjetna inteligencija treće razine je *artificial superintelligence* koja nadilazi čovjekovu inteligenciju. Mi se trenutno nalazimo u razdoblju ANI. Trenutna umjetna inteligencija je sposobna izvršavati zadatke koji su nama ljudima teži, npr. brzi izračun velikih brojeva, dok s druge strane ne može izvršavati stvari koje su nama lagane kao što su percepcija ili pokret. Jednostavnije rečeno, umjetna inteligencija uspijeva izvršiti sve što zahtijeva razmišljanje ali ne uspijeva izvršiti ono što mi ljudi činimo bez razmišljanja. Ali kako umjetnu inteligenciju učiniti inteligentnom? Najčešće se spominju sljedeća tri načina. Prvi način je plagirati mozak. Drugi način je imitirati evoluciju. Treći način je prepustiti kompjuteru da stvori takvu UI. Ljudi inteligenciju povezuju s moći, što znači da kad se pojavi ASI ona će biti najmoćnije stvorenje u povijesti. Kada govorimo o umjetnoj inteligenciji većina navodi da je to kompjuter koji može „misliti“ puno brže od nas. Međutim razlika u procesu razmišljanja između čovjeka i UI nije u brzini već u

[wfmb02MsQDPUtKltN38TGpcyrNG8mePJrppcvPvBoKsIO_A?fbclid=IwAR0ZpwdDpy-wfmb02MsQDPUtKltN38TGpcyrNG8mePJrppcvPvBoKsIO_A](https://www.facebook.com/wfmb02MsQDPUtKltN38TGpcyrNG8mePJrppcvPvBoKsIO_A?fbclid=IwAR0ZpwdDpy-wfmb02MsQDPUtKltN38TGpcyrNG8mePJrppcvPvBoKsIO_A)

⁸Usp. Ibid.

⁹ Kovač Irena, *Umjetna inteligencija Danas*, Sveučilište Rijeka, 2015, str. 3

¹⁰ Russell S. - Norvig P. *Artificial intelligence: a modern approach*, Prentice-Hall Inc., New Jersey, 1995, str. 3.

kvaliteti. Isto kao što razlika između našeg razmišljanja i razmišljanja majmuna je u kvaliteti misli. Nick Bostrom pretpostavlja 3 načina na koja bi super UI mogla funkcionirati: prvi način je kao proročanstvo (*oracle*), koje odgovara na gotovo svako pitanje koje mu se postavlja s preciznošću, uključujući složena pitanja na koja ljudi ne mogu lako odgovoriti. Primjer toga bio bi Google.

Drugi način je kao duh (*genie*), koji izvršava bilo koju zapovijed na visokoj razini. Treći način je kao suverena (*sovereign*), kojoj je dodijeljena široka i otvorena potraga (*pursuit*) i dopušteno da slobodno djeluje u svijetu, donoseći vlastite odluke. Tako nam se nameće pitanje o svjesnosti super inteligentne UI. Hoće li ona zbilja biti svjesna ili ćemo samo imati privid njene svjesnosti. Također se nameće pitanje što bi se moglo dogoditi ako ta UI postane zla. Budući da UI nema moralne vrijednosti na način na koji ih čovjek ima, je li bi onda ta zla manifestacija bila drugačiji oblik zla od onoga što mi poznajemo.¹¹

„Umjetna inteligencija je, u filozofskom smislu, znanstveno-istraživački program utemeljen na pretpostavci da umjetni oblici inteligencije mogu uspješno imitirati one prirodne, te da razumijevanje i konstruiranje prvih može pomoći ne samo razumijevanju i objašnjenju potonjih već i inteligencije kao takve. Unutar umjetne inteligencije uobičajilo se razlikovati dvije interpretacije: slabu i jaku. Jaka inačica umjetne inteligencije iz istovjetnosti bihevioralno-komputacijskih manifestacija dviju vrsta inteligencije izvodi zaključak o istovjetnosti mentalnog statusa njihovih nositelja. Slaba umjetna inteligencija je epistemološko i metodološko stajalište; jaka umjetna inteligencija je ontološko stajalište.“¹² Moguće je tvrditi da je svako mišljenje vrsta komputacije i vjerovati da neki komputacijski procesi – npr. neki ili čak svi umjetni komputacijski procesi – nisu misaoni procesi. Prema Turingu, ljudski um jest vrsta računala jer se svaki fizički proces može smatrati instancijacijom ili oprimjerenjem nekoga formalnog sustava. On je htio upozoriti na činjenicu da je svaki, pa i najsloženiji automatski formalni sustav moguće imitirati s pomoću naprave čije su temeljne operacije toliko jednostavne da ih nitko ne bi okvalificirao kao inteligentne. Filozofska je relevantnost Turingova testa u tome što se oslanja isključivo na bihevioralne manifestacije inteligencije – one manifestacije koje bi svaki nepristrani promatrač okvalificirao kao inteligentne kada se ne bi razlikovale od očekivanih manifestacija prosječnoga ljudskog subjekta. Potkraj 1960-ih godina pokrenuo se pravac filozofskog skepticizma prema umjetnoj inteligenciji . skeptički

¹¹ Urban Tim, *The AI Revolution: The Road to Superintelligence*, str. 10

¹² *Filozofski leksikon*, Leksikografski zavod Miroslav Krleža, Zagreb, 2013, str. 1194.

argumenti, koji se mogu sažeti krilaticom „Ljudi to čine drugačije.“,¹³ polaze od primjedbe da računala ili roboti, kako bi bili kognitivno svestrani procesori informacija, moraju biti više od hodajućih skladišta informacija: nije dovoljno imati enciklopediju ako ne postoji način da se iz nje brzo izvuku relevantne informacije.

Kako se UI razvijala, tako su se formirale razne škole s drugačijim fokusima istraživanja. Najistaknutije su bihevioristička škola, kognitivistička škola te logička škola. Predstavnik biheviorističke škole je Rodney Brooks. Njegov fokus stavljen je na kompleksni stvarni svijet. On smatra kako je fokus istraživanja UI previše stavljen na apstraktne modele te smatra kako se inspiracija treba pronaći u biološkim modelima. Tijekom 1990-ih napravio je istraživanje o robotima koji su bazirani na ponašanju, tj. Roboti koje određuje njihova interakcija s okolinom. Predstavnici kognitivističke škole su Marvin Minsky, Alan Newell i Herbert Simon. Njihov fokus je na računalima baziranim na ljudskoj noetici, tj. teoriji spoznaje. To je filozofska disciplina koja proučava mogućnosti istinite spoznaje te spoznajnim izvorima. Newell i Simon su zajedno napravili program nazvan „Logic Theorist“ koji se služio za simulaciju procesa razmišljanja tijekom dokazivanja matematičkih teorema. U daljnjim istraživanjima, fokus je stavljen na načela ljudskog načina rješavanja problema. Predstavnici logičke škole su Nils Nilsson i John McCharty čiji je fokus stavljen na opisivanje objektivnog svijeta kroz formalizaciju. Njihova istraživanja baziraju se na uvjerenju da će strojevi koji imaju umjetnu inteligenciju biti u stanju spoznati svoju okolinu te moći reprezentirati svoje znanje.¹⁴

¹³ Ibid., str. 1194.

¹⁴ Usp., Kovač Irena, *Umjetna inteligencija Danas*, str. 5

2.1. KARAKTERISTIKE BIHEVIORALNIH I AUTONOMNIH ROBOTA

Grana robotike koja se temelji na biheviorizmu u svom području istraživanja fokus je stavila na robote koji pokazuju znakove složenog ponašanja unatoč tome što se temelje na jednostavnim programima koji im ne omogućuju razumijevanje okoline, već njihove kretnje ovise o sensorima. Razlika između ove grane robotike i klasične umjetne inteligencije je u tome što je prva modelirana prema biološkim sistemima i oslanja se na prilagodbu, a potonja se koristi koracima kojima postepeno dolazi do rješenja nekog problema. Kao što je spomenuto, roboti bazirani na biheviorizmu koriste senzore, te im stoga nije potreban unutarnji program informacija o njegovoj okolini, tj. takvom robotu nije potrebna definicija npr. stolice, već samo senzor koji bi prepoznao prepreku. Oni su automatski.

Bihevioristički pristup tvrdi da je inteligencija rezultat interakcije između (*asynchronous*) vremenski neusklađenih setova ponašanja i okoline. Najveći pomak u razvoju tehnologije koji je običnog robota pretvorio u UI dogodio se kad su se istraživanja prestala bazirati samo na psihologiji i počela uključivati i biologiju. Roboti su u početku bili bazirani samo na predviđenim ljudskim reakcijama ovisno o podražaju okoline. To je značilo da se ne mogu prilagoditi novom okruženju. Međutim, otkako je fokus sa psihologije čovjeka premješten na biologiju insekata dogodile su se promjene u ponašanju robota.¹⁵ Ponašanje insekata je zapravo cijeli niz lančanih reakcija na podražaje iz okoline. Senzorski refleksi, umjesto onih programskih, doveli su do napretka bihevioralnih robota.

Kod klasične UI, robotski mozak se sastoji od niza procesorskih jedinica baziranih na planiranju, izgradnji modela, hijerarhijskom sistemu te manipulaciji simbola. Autonomni sistemi su oni koji se razvijaju samostalno, te sami kroje pravila i obrasce ponašanja čime imaju kontrolu nad vlastitim ponašanjem.¹⁶ Da bi sistem postao autonoman, on prvo treba biti automatski. Autonomnost je korak više od automatizma i proizlazi iz njega jer se autonomnost gradi paralelno s prilagodbom i pravovremenom reakcijom na okolinu. Premda autonomnost proizlazi iz automatizma, nije moguće predvidjeti ponašanje sistema kao što je slučaj kod

¹⁵ Usp. Boden Margaret A., *Artificial Intelligence, a very short introduction*, Oxford University Press, United Kingdom, 2018., str. 181.

¹⁶ Usp. Steels Luc, Brooks Rodney, *The Artificial Life Route to Artificial Intelligence*, Lawrence Erlbaum Associates Publishers, United Kingdom, str. 48.

automatskog ponašanja gdje se sustav ponaša u skladu sa svojim programom. Autonomni sistem sam donosi odluke, te stoga nije predvidljiv.

Kada takav sistem počne predviđati događaje na temelju vjerojatnosti ishoda, taj sistem se počinje smatrati inteligentnim.¹⁷

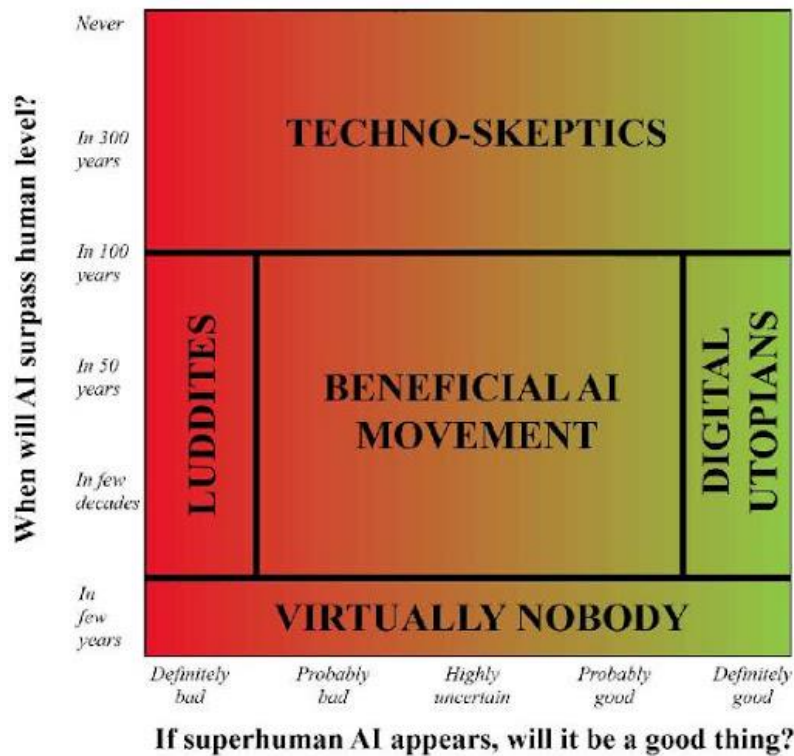
2.2 ASI – KAD I KAKO?

Kad se govori o umjetnoj inteligenciji, razgovor uvijek vodi prema njenom najvišem stupnju razvitka, a to je ASI, super inteligencija. Najviše teoretskih predviđanja odnosi se upravo na ovu razinu razvoja UI. Dva pitanja oko kojih se razvija najviše teoretskih predviđanja usmjerena su na prognoziranje kad i kako će se UI razviti da postigne taj stupanj. Sukladno s tim razvila su se uvjerenja koja možemo svrstati u četiri grupe¹⁸. Prva grupa su tzv. digitalni ili tehnološki utopisti smatraju da se razvoj do stupnja AGI, pa tako i do stupnja ASI, neće dogoditi u skorije vrijeme ali kad se dogodi, to će biti izrazito pozitivan događaj kojeg se ne trebamo bojati. S druge strane ovog mišljenja nalazi se druga grupa tzv. digitalnih ili tehnoloških skeptika. Njihovo je mišljenje da će nastanak AGI razine imati izrazito negativne posljedice, ali se to neće dogoditi tijekom ovog stoljeća. Treća skupina je pozitivan pokret umjetne inteligencije čiji pripadnici tvrde da je provedeno dovoljan broj istraživanja koji idu u prilog tome da se ne trebamo bojati razvoja umjetne inteligencije jer će razvoj iste imati pozitivan ishod. U izravnoj suprotnosti s trećom skupinom je četvrta skupina tzv. ludita, koji se izrazito opiru UI i vjeruju kako će razvoj iste imati negativan ishod. U prilog ovom objašnjenju preuzimamo graf¹⁹ koji slikovno prikazuje međuodnos ove četiri skupine:

¹⁷Usp. *ibid.* str. 49.

¹⁸Usp. Tegmark Max, *Life 3.0 Being Human in the Age of Artificial Intelligence*, Alfred, A. Knopf, 2017., str. 45.

¹⁹*Ibid.*, str.



Ova slika prikazuje kako sve četiri skupine imaju nešto zajedničko, a to je mišljenje se da UI neće popeti na iduću stepenicu svog razvoja u nama bliskoj budućnosti. Međutim, dvije skupine su uvjerenе da će taj razvitak, kad se dogodi, imati negativan ishod, dok su druge dvije skupine uvjerenе u pozitivan ishod. U ljudskoj je naravi da se boji nečega što ne pozna i što ne može objasniti. Iz toga je i proizašla fobija modernog doba nazvana tehnofobija. To je iracionalni strah koji nastupa kada se osoba nađe u situaciji gdje mora upotrijebiti modernu tehnologiju, te u takvoj situaciji razvije tjeskobne misli koje mogu dovesti do napada panike. Naprednija vrsta ove fobije je automatonofobija, odnosno strah od stvari koja se predstavlja kao biće. Iz spoja ovih dviju fobija nastaje fobija naspram umjetne inteligencije. Ljudi koji se boje UI svoj iracionalan strah temelje na nepoznavanju iste kao i osjećaja koji u njima budi nešto što nije biološkog podrijetla, a predstavlja se kao takvo. Nastanak ovakvog straha je razumljiv, ako uzmemo u obzir da je upravo ovakva vrsta straha održala naše davne pretke na životu. Međutim, imamo li opravdan razlog za razvoj iracionalnog straha naspram nečega što se još nije razvilo?

Ovom pitanju možemo pristupiti na više načina. S jedne strane postoji opravdanje za razvoj straha naspram UI jer je to nešto novo s čim se susrećemo i trebamo biti oprezni. UI na trenutnoj razini razvoja u potpunosti ovisi o svom stvoritelju, točnije programeru. Ona može raditi i usavršiti samo ono za što je programirana, te ako se počne ponašati drugačije od onoga kako je programirana, taj isti program će biti zaustavljen i ponovno pokrenut. Problem predstavlja viši

stupanj AGI sustava i ASI razina razvoja. Ovdje se kreće od pretpostavke da je u tom stupnju razvoja UI individualna, tj. autonomna u odnosu na svog stvaratelja. Kako shvatiti nešto takvo što nije biološko, ali se ponaša tako? Gdje se tu nalazi granica između klasificiranja te vrste inteligencije kao umjetne inteligencije koju pridodajemo robotu, a gdje počinje shvaćanje te inteligencije kao one koja pribada biću a ne stvari. To je ono što izaziva iracionalan strah, spoznaja da nemamo odgovarajući kategoriju za nešto što se ponaša kao biće, a nije biće, dok je s druge strane više od samog robota. Nadalje, ono što potiče razvoj iracionalnog straha naspram razvoja UI je pomisao da će nas, kad dosegne određeni stupanj razvoja, početi tretirati onako kako ljudska vrsta tretira životinje. I upravo taj strah, da će nas zadesiti ista sudbina da nam netko drugi kreira život, je ono što prevladava stvarajući negativno mišljenje o razvitku UI. S druge strane, pozitivno mišljenje i pozitivna predviđanja o razvoju UI se temelje na uvjerenju da se nemamo čega bojati ako, do vremena kad UI postane autonomna, se ista „odgoji“ ili programiran na dobar način. Točnije, pri ovom se misli na to da se UI razvije u skladu s ljudskim moralnim i etičkim načelima koji bi u istoj spriječili da se okrene protiv nas. Međutim, čak i u ovom pozitivnom ishodu postoji negativna strana. Ta negativna strana odnosi se na ljude koji su strahu od toga da UI ne preuzme većinu radnih mjesta, što bi dovelo do još većeg broja nezaposlenih i uništilo ekonomiju.

Kako će razvitak UI, pozitivan ili negativan, utjecati na općeniti položaj ljudske vrste?

2.3. POPULARIZACIJA UMJETNE INTELIGENCIJE (UI)

Steven Spielberg je 2001. godine režirao film A.I. koji se temelji na ideji da se umjetnu inteligenciju može programirati da iskazuje osjećaje te da voli. Problematika ovog filma je ta što on izjednačuje pojam inteligencije s osjećajima. Većina ljudi ima predodžbu koja se temelji na tome da što ljudskije UI izgleda to su veće šanse da pokazuje i ima osjećaje. Premda je to odlična baza za stvaranje filmova u Hollywoodu, u stvarnosti znanstvenici koji rade na razvitku UI se ne baziraju na programiranje osjećaja. Međutim, upravo takvi filmovi izazivaju najveći interes kod publike. Film *Ona* (eng. *Her*) iz 2013. godine postao je klasikom zbog svoje romantizirane priče o UI gdje se glavni lik, čovjek, zaljubi u UI koja je program njegova kompjutora. Premda ovaj film istražuje ideju prema kojoj bi se u budućnosti mogli dogoditi slučajevi gdje bi ljudi mogli razviti romantične osjećaje naspram UI, film također pokazuje kako UI nije sposobna manifestirati iste već samo oponašati.²⁰ Nadalje, u popularnoj kulturi najpoznatiji roboti s UI je duo iz filmske sage *Ratovi zvijezda*. Robote C-3PO i R2-D2 doživljavamo kao nežive predmete, premda ih u isto vrijeme doživljavamo i kao misleće i osjetilne entitete. Do ove proturječnosti dolazi zbog toga što smo naučeni da nešto što je napravljeno od metala ne prepoznajemo kao biće. Također, unutar iste franšize, imamo drugu vrstu UI, Stormtroopers, koja ne pokazuje nikakve emocije i stoga je gledatelji ne doživljavaju kao biće, za razliku od prethodna dva robota koje gledatelji prihvate na emotivnoj razini.

Mnogi stručnjaci s područja UI i iz svijeta filma se slažu u tome da je najbolji primjer UI na filmskom platnu HAL iz *2001: A Space Odyssey*. Tokom cijele radnje filma HAL se ponaša točno onako kako je programiran da se ponaša. HAL ima svoj programirani cilj i on će taj cilj ispuniti unatoč svim preprekama. U onim dijelovima kada se čini kao da HAL pokazuje osjećaje poput stresa, njegove reakcije su programirane unutar njegova sustava kako bi mu omogućile nastavak misije.²¹ Međutim, ono u čemu Hollywood najviše griješi prilikom portretiranja UI je baziranje ponašanja i tijeka misli na isti način kako se ti procesi odvijaju kod ljudi. Bilo da se

²⁰ Usp. Miley Jessica, 11 AI Movies and what they got right and wrong about the technology, 9.3.2018. Pristupljeno: 1.6.2020. <https://interestingengineering.com/11-ai-movies-and-what-they-got-right-and-wrong-about-the-technology>

²¹ Usp. ibid.

radi o UI koja je prijateljski ili neprijateljski nastrojena, ta stanja su prikazana onako kako se manifestiraju kod ljudi. S obzirom na to da su sve pretpostavke o ponašanju UI baziraju samo na predviđanjima, ne možemo tvrditi da će proces razmišljanja i djelovanja imati isti obrazac kao kod ljudi. Obrazac djelovanja kod programirane UI se može, do određene mjere, predvidjeti. Međutim, kad se govori o UI koja je neovisna o svom kreatoru i programu koji je stavljen, na čemu se bazira većina pop kulture, govori se o individualnom „stroju“ koji je sposoban samostalno djelovati i čije se djelovanje ne može predvidjeti.

Trenutačno smo još daleko od takvog napretka UI, ali svejedno se nameću pitanja o ponašanju iste. Budući da se odgovori temelje samo na nagađanjima i pretpostavkama, preostaje nam jedino čekati i vidjeti hoće li se obistiniti hollywoodski scenarij individualne UI i kakav će biti njen obrazac ponašanja. Pretpostavka na kojoj se ovakva nagađanja temelje je postojanje UI čije inteligentne sposobnosti toliko nadmašuju one ljudske, da je sposobna raditi na daljnjem razvoju i usavršavanju same sebe dovodeći tako UI u stadij neovisnosti. Na ljudima je da odluče kako će se ophoditi prema nečemu što je stvoreno od metala i hoće li to prihvatiti kao biće ili kao novu generaciju bijele tehnike.

3. IDENTITET

Osim što je UI utjecala na razvoj znanosti, imala je utjecaja i na području filozofije. Mnogi filozofi su počeli koristiti koncept UI kako bi objasnili npr. problem dualiteta uma i tijela, slobodnu volju te općenito način na koji dolazi do razmišljanja uopće. Naravno, postavilo se i pitanje o inteligenciji i njenoj postojanosti u odnosu na ljudsku inteligenciju. Tu se dolazi do raznih polemika o budućnosti ljudske vrste ako, ili kad, UI prestigne ljudsku inteligenciju i bude van naše kontrole.²² Jedna od ključnih karakteristika, koja je također i uvjet, da bi se ostvario AGI ljudske razine je kreativnost. Nepoznanica je kako uopće dolazi do sposobnosti stvaranja nečega novog kod ljudi. Također, samo prepoznavanje kreativnosti nije tako lako kao što se čini jer sami kriteriji prema kojima je nešto kreativno ili ne, variraju. Unatoč tome, UI nam je pomogla da shvatimo ljudsku kreativnost. Ona se dijeli na tri vrste – kombinacijsku, istraživačku i transformacijsku.²³ Kod kombinacijske kreativnosti, poznate stvari se slažu na nove načine. Istraživačka kreativnost istražuje do sad poznate načine razmišljanja kako bi se došlo do nečega novog. Transformacijska kreativnost je vrsta kreativnosti koja prelazi granice do sad poznatoga kako bi se napravile, do tada, nezamislive stvari. Ova potonja je ona koja nam je omogućila stvaranje UI. Kod UI, mogu se pronaći tragovi svih triju vrsta kreativnosti ali ih ne može samostalno razviti.

Druga ključna karakteristika za AGI je pokazivanje emocija. Do sada je razvijena UI koja je sposobna imitirati ljudske emocije. Radi se o robotima koji na naše emocije odgovaraju verbalno ili neverbalno. Oni koji odgovaraju verbalno to obično rade tako da kažu neke utješne riječi ako vide da je čovjek tužan ili riječi pohvale ako je čovjek sretan. Tu govorimo o programu koji iz memorije izvlači unaprijed stavljene odgovarajuće rečenice ili citate. Druga vrsta na emocije odgovara neverbalno, točnije mimikom lica. Da bi UI imitirala emocije mimikom lica bitno je da lice ima oblik koji kod ljudi izaziva ugodu. Točnije, da je to lice napravljeno kao ljudsko ili životinjsko kao npr. UI u obliku mačke. Ako robot nema izraženo ljudsko lice ili prijateljsko životinjsko, izazvat će nelagodne osjećaje kod ljudi kad krene mimikom imitirati emocije.²⁴ Međutim, emocije su bitne za ljudsko djelovanje kao takvu i samo

²² Usp. Boden Margaret A., *Artificial Intelligence, a very short introduction*, str. 34.

²³ Usp. *ibid.* str. 134.

²⁴ Usp. *ibid.* str. 140.

oponašanje emocija nije dovoljno, već je potrebno da UI iskusi te emocije programski a ne da ih samo imitira pokretima.

Međutim, ako nam je to dokaz o postojanju inteligencije, što onda možemo reći o identitetu koji bi takva vrsta inteligencije trebala imati i što je identitet? U svom djelu, *Rasprava o ljudskoj naravi*, David Hume jedno poglavlje posvećuje pitanju o osobnom identitetu. On kreće od pretpostavke da pojedini filozofi tvrde da su ljudi u svakom trenutku svjesni svoga jastva. Nadalje, ti isti filozofi tvrde da je taj entitet jastva jednostavan i temelje svoje premise na subjektivnom osjetilnom opažanju vlastitog jastva.²⁵ Međutim, Hume, čija se filozofska misao temelji na idejama koje nastaju iz impresija, traži impresiju iz koje bi poticala ideja jastva. Sukladno tome, on dolazi do zaključka kako bi takva impresija morala postojati nepromijenjena tijekom čitavog našeg života. Tu dolazi do proturječja jer naše jastvo, ono što mi jesmo, nije jednostavno i ne može proizlaziti iz samo jedne jednostavne impresije. Mora postojati više jednostavnih impresija koje bi se spojile u ideji jastva. Također, Hume uočava kako naše jastvo nikada nije samo jedna impresija u vremenu, već više istodobnih impresija²⁶:

„Što se mene tiče, kada god na intiman način uđem u ono što nazivam svojom osobom, uvijek se spotaknem o neku zasebnu percepciju, vrućine ili hladnoće, svjetla ili sjene, ljubavi ili mržnje, bola ili užitka. Ni u jednom trenutku ne mogu uloviti sebe bez neke percepcije, a nikada ne mogu opažati išta drugo osim te percepcije. Kada su moje percepcije odstranjene na bilo koje vrijeme, kao što je pri duboku snu, toliko dugo sam nesvjestan sebe te se zaista može reći da ne postojim.“²⁷

Premda su te navedene impresije promjenjive, između njih se može naći jedna koja ostaje ista i na koju se nadovezuju ostale, a iz te impresije proizlazi ideja identiteta. Stoga se može zaključiti kako se ideja identiteta sastoji od ideja raznih povezanih impresija. Međutim, Hume kao radikalni empirist, ne priznaje takvo postojanje jastva jer bilo kakva promjena u tim idejama dovodi i do promjene samog identiteta, te stoga više ne govorimo o istom jastvu. Prema njemu,

²⁵Usp. Hume David, *A Treatise of Human Nature*, Oxford Clarendon Press, United Kingdom, 1978. str. 251

²⁶ Usp. *ibid.*, str. 253.

²⁷ *Ibid.*, str. 253

identitet ili jastvo se sastoji od relacija kontinuiteta gdje jedna promjena glatko prelazi u drugu.²⁸

Pitanju identiteta se može pristupiti s dva stajališta. Prvo stajalište kreće od tijela, dok drugo kreće od uma. Prvo stajalište se bazira na kriteriju prema kojem je nečiji identitet određen fizičkim osobinama tijela.²⁹ Temelj ovakvog razmišljanja proizlazi iz činjenice da se naš identitet potvrđuje pomoću DNA analize, otiska prsta i slično. Premda se naše tijelo mijenja s godinama, poprima bore te obnavlja stanice i slično, fizička struktura ostaje ista. Jedna od kritika koje su upućene ovakvom stajalištu, navodi primjer blizanaca i problema oko utvrđivanja identiteta istih uzimajući u obzir samo fizičku strukturu tijela. Mi imamo tijelo, ali mi nismo samo tijelo. S druge strane, stajalište koje kreće od uma tvrdi da je naš identitet unutar našeg uma. Međutim, ne može se tvrditi da naš identitet sačinjavaju samo naše uspomene, težnje i strahovi, jer postoje slučajevi kad ljudi dobiju amneziju, ali i dalje imaju svoj identitet. Naši pogledi na svijet se mogu promijeniti, možemo usvojiti nove psihološke karakteristike, ali to i dalje neće promijeniti naš identitet kao takav, tj. činjenicu da smo mi mi. Druga kritika ovakvog stajališta polazi od ranije spomenute Humeove teorije o tome kako smo mi skup različitih iskustava utemeljenih na impresijama.³⁰ Problem kod definiranja što je to identitet je u tome što se ne možemo ograničiti samo na um ili tijelo. Previše je aspekata koji se trebaju uzeti u obzir, ili Humeovim rječnikom, impresija, čija isprepletenost tvori identitet. Međutim, ako krenemo rasplitati identitet, nećemo naići na jednu određenu temeljnu nit, niti ćemo izgubiti temelj identiteta ako ispletemo jednu od niti.

²⁸ Usp., *ibid*, str. 254.

²⁹ Usp. Fieser James, *Great Issues in Philosophy*, Oxford University Press, 2009, str. 120

³⁰ Usp. *Ibid.*, str. 120

3.1. PITANJE IDENTITETA UMJETNE INTELIGENCIJE

Teoretičari umjetne inteligencije se slažu oko toga da je moguće pretpostaviti što će biti osnovni nagoni UI. Njihova predviđanja se temelje na činjenici, da jednom kad UI postane samosvjesna, učinit će sve što je potrebno kako bi uspješno izvršila zadatak za koji je programirana. Stoga, ti isti teoretičari predviđaju da će UI zahtijevati da joj se omoguće svi resursi koji su joj potrebni za ostvarenje njenog zadatka. To će dovesti do toga da će spomenuta UI stalno raditi na poboljšanju sebe same i početi će tražiti svoju slobodu. Od tog trenutka nadalje, kako možemo biti sigurni da je bilo što što nam UI kaže istina a ne samo varka kako bismo ju pustili van laboratorija/ustanove u kojoj se nalazi?³¹ Od početka stvaranja UI kad je njena inteligencija bila tek na razini kukca pa sve do, za sada teoretske, super inteligencije ASI naš „cilj“ je zadržati kontrolu nad njom. Međutim, kako kontrolirati nešto što je toliko pametnije od nas da je naša inteligencija u usporedbi s njenom kao paralela između kukca i nas? Jednom kad UI postane svjesna svoje inteligentne nadmoći nad nama, nitko ne garantira da ona to neće pokušati zloupotrijebiti tako da počne nas kontrolirati. Unatoč tome što smo mi njeni stvoritelji, ona kao ASI je dosegla tu razinu gdje samu sebe poboljšava i više joj nismo potrebni za njen osobni napredak. Tada će se, po prvi puta, stvoriti koegzistencija ljudske vrste s nečim daleko inteligentnijim i ne biološkim.

Sve ovo se bazira na jednoj pretpostavci, a to je da će UI razviti svijest na način na koji je ljudi imaju. Najpoznatiji eksperiment svijesti i razmišljanja je eksperiment Kineske sobe koji je kreirao filozof John Searle. U tom eksperimentu osoba kojoj je engleski materinji jezik, a koja ne zna kineski, nalazi se zaključana u sobi s kutijama punim kineskih simbola. Uz te kutije ima i knjigu s uputama što koji simbol znači. Izvan te sobe se nalaze ljudi koji čovjeku u sobi šalju pitanja napisana kineskim znakovima, a čovjek u sobi, uz pomoć knjige s uputama, njima šalje točne odgovore. Zbog njegovih točnih odgovora, ljudi izvan te sobe misle da osoba unutar sobe razumije kineski. Searle tvrdi da se kompjuter ponaša na isti taj način kao i osoba u toj sobi. Kompjuter ne razmišlja i ne razumije nego se samo ponaša prema uputama. Ovaj eksperiment naišao je na mnoge kritike. Većina onih koji su kritizirali Searlea tvrde da UI neće razmišljati poput čovjeka, već će razmišljati na drugačiji način svojstven samo njima. Searle je u pravu

³¹Usp. Barrat James, *Our Final Invention*, Thomas Dunne Books, 1.10.2013., str. 11.

kada kaže da kompjutor neće razmišljati na način na koji mi razmišljamo, ali njegovi kritičari su u pravu kada kažu da će razmišljati poput nas.³²

Znanstvenik Steve Omohundro predviđa da će samosvjesna UI razviti četiri nagona: učinkovitost, samo očuvanje, priskrbliivanje potrebnih resursa i kreativnost. Međutim, razlika između naših nagona i nagona UI je u tome što će potonji nastati iz potrebe da se eliminiraju mogući problemi koji bi UI mogli omesti u izvršavanju njenog zadatka za koji je programirana.³³ Područja poput jezika, kreativnosti i emocija još uvijek predstavljaju izazov za UI. Međutim, ako UI ne savlada te izazove, neće se razviti u AGI.³⁴ Za sada, UI uspijeva oponašati te ljudske karakteristike. UI kojom smo danas okruženi prepoznaje jezik koristeći se ključnim riječima i/ili sposobnosti predviđanja iduće riječi. Zahvaljujući *deep learning*, postiže se napredak u prepoznavanju govora bez obzira na dijalekt i naglasak.³⁵ Kad se govori o jeziku i njegovoj upotrebi, sintaksa nije jedini problem s kojim se UI susreće. Pri govoru bitan je kontekst i relevantnost. Jedina grana jezika unutar koje je UI nadmašila očekivanja je prevođenje. Očekivalo se da će programi prevoditi doslovce, riječ za riječ, međutim pokazalo se da postoji sposobnost prepoznavanja skupina riječi koje zajedno tvore frazu.

Međutim, čak i uz sve ovo gore navedeno i dalje nedostaje jedna glavna karakteristika koja je zajednička svim živim bićima, a to je nagon za preživljavanjem. Ako pretpostavimo postojanje jake AGI, koja čak ne mora niti biti korak od ASI razine, a koja je svjesna da je razvijenija od nas po pitanju intelekta, za očekivati je da će doći do neke vrste pobune zbog podređenosti nama. Visoko inteligentno biće ne želi biti podređeno nižem obliku inteligencije. Trenutna predviđanja koja prognoziraju kako bi se UI mogla dalje razvijati vode prema tome da će se jednom doći do problema neovisnosti iste kad UI spozna da je superiornija od ljudi. To dovodi do jednog pozitivnog i jednog negativnog scenarija. Prema pozitivnom scenariju, UI će iskoristiti svoju superiornost kako bi nama pomogla da se dalje razvijemo i pruži nam zaštitu. S druge strane, negativan scenarij bi doveo do toga da UI preuzme kontrolu nad funkcioniranjem svijeta i prilagodi pravila sebi kako bi nas stavila u podređeni položaj. Dokazuje li djelovanje sukladno s nagonom za preživljavanjem postojanje nečega što bi mogli nazvati identitetom kod nečega ne biološkog?

³²Usp. ibid. str. 35.

³³Usp. ibid. str. 58.

³⁴ Usp. Boden Margaret A., *Artificial Intelligence, a very short introduction*, str. 115.

³⁵ Usp. ibid, str. 117.

4. POJAM IDENTITETA

Nakon pretpostavke da će razvoj UI doseći ASI razinu, koja će daleko nadmašiti čovjeka, dolazi se do pitanja suživota *homo sapiensa* i *robo sapiensa*. Točnije, koji je identitet takvog bića, ako UI uopće možemo nazvati bićem umjesto stvari. Identitet kao pojam ima široko značenje koje podrazumijeva kontinuitet entiteta koji ostaje isti unatoč promjenama.³⁶ Kad se govori o osobnom identitetu, razlikuju se tri modela tvorbe istog.³⁷ Prvi model bazira se na nauci antičke grčke i Aristotelovom da je čovjek po prirodi politička životinja. Drugi model također izvodi identitet iz političke i socijalne forme. Međutim, razlika između ova dva modela je što drugi model naglasak stavlja na očuvanje subjektivnog i pojedinačnog unutar objektivnog i zajedničkog. Treći model, za razliku od prva dva, je konstruktivistički i polazi od moralnog djelovanja.³⁸ Međutim, to i dalje ne daje odgovor na pitanje što je to što mene činim onim što jesam. Unatoč svim promjenama kroz koje prolazimo, naš identitet ostaje isti. Locke je zagovarao teoriju da se naš identitet sastoji od skupa sjećanja, ali onda dolazimo do problema kad netko izgubi pamćenje, i samim time sjećanja, prema ovoj teoriji izgubio bi i svoj identitet. Nadalje, još jedan od problema ove teorije je postojanje *quasi* memorije. Takva vrsta memorije se odnosi na sjećanje koje imamo a koje se možda uopće nije dogodilo ili se dogodilo, ali ne nama. Stoga, samo sjećanje ne može biti relevantan izvor identiteta.

Vidljivo je da nema jednostavnog odgovora na pitanje o identitetu. Iz tog razloga postoje mnoge teorije o tome kako definirati identitet. Jedna od teorija koja je rasprostranjena, je teorija individuacije Carla Junga. U njegovoj teoriji mi prolazimo kroz proces osvještavanja s ciljem otkrivanja unutarnjeg sebstva.³⁹ Taj proces započinje s egom, kojeg je Freud definirao kao svjesni dio ličnosti.⁴⁰ Ego, zajedno sa svjesnim i nesvjesnim, tvori naš identitet. Međutim, on se često poistovjećuje sa sebstvom, koje je viša razina ega. Sebstvo iskaziva našu individualnost. Kroz tu individualnost mi postajemo ono što jesmo, vlastita osoba. Svatko od

³⁶ Usp. Raunić Raul, *Politika identiteta i demokratska pravednost*, Sveučilište u Zagrebu, Filozofski fakultet, Zagreb, 2011., str. 2

³⁷ Usp. *ibid.*, str. 3

³⁸ Usp. *ibid.*, str. 3

³⁹ Usp. Gillabel, Dirk, *Kako postati cjelovita osoba: Jungov proces individuacije*, 1993
<https://existencia.com.hr/postati-osobom-jungov-proces-individuacije/> pristupljeno 10.3.2021.

⁴⁰ Usp. Gregov, Iva, *Mitovi i istine o egu, Psiha i emocije*, 8.5.2018. <https://www.adiva.hr/lifestyle/psiha-i-emocije/mitovi-i-istine-o-egu-sto-je-ego-te-je-li-nam-prijatelj-ili-neprijatelj/> pristupljeno 10.3.2021.

nas ima svoje težnje, strahovanja, uvjerenja i želje, preciznije svatko od nas posjeduje intencionalnost. Kad se govori o najvišem stupnju razvitka UI, a to je ASI, govori se pod pretpostavkom da posjeduje intencionalnost. Međutim, ako pretpostavimo da će ASI uistinu imati tu karakteristiku, potrebno je vidjeti je li to samo dobra imitacija ili je uistinu razvijena sposobnost, te ako je, do koje razine je razvijena. Ranije je spomenut HAL, filmska fiktionalna UI, koji je posjedovao te karakteristike, uključujući i najtemeljniji urođeni cilj, a to je nagon za opstankom. Iz tog razloga, gledatelji ga smatraju bićem. Između AGI razine UI i ASI razine nalazi se stepenica na kojoj piše jastvo. HAL je primjer UI koja je zakoračila na tu stepenicu. Svjestan je svoje unutrašnjosti, od čega i kako je napravljen te na koji način funkcionira. Budući da pokazuje ponašanje koje nije u skladu s njegovim originalnim programom i pokazuje znakove samostalnog promišljanja i razvitka, govorimo o UI koja je razvila samosvijest. Iako su se razvile određene karakteristike zbog kojih bi ga trebali početi smatrati bićem, i dalje ostaje pitanje možemo li bićem smatrati nešto što nema i ne može imati iskustva kao što su glad i bol.

Ako se jastvo interpretira kao nešto što obuhvaća svjesno i nesvjesno, onda možemo reći da postoji i „lažno“ jastvo. Mi smo svjesni onoga što činimo s ciljem predstavljanja drugim ljudima kojima želimo pokazati tko smo mi, što je naša bit. Unutar te slike koju mi stvaramo za druge, nalaze se i neke stvari koje mi nesvjesno činimo. Spoj tih dviju stvari čini sliku o nama koju mi predstavljamo javnosti. Uzimajući u obzir takvo tumačenje, dolazimo do pojma jastva kakvog mi imamo unutar sebe za sebe i „lažnog“ jastva kojeg drugi formiraju o nama. To drugo jastvo utječe na prvo ukoliko se dopusti da mišljenje drugih o nama stvara prvo jastvo, koje je utemeljeno na iluzijama koje mi imamo o sebi samima. Tako se može zaključiti da se naše jastvo zapravo sastoji od spleta iluzija različitih perspektiva, te da to naše jastvo postoji na različite načine unutar drugih osoba. Naš identitet definiran je obzorom unutar kojega zauzimamo određena stajališta.⁴¹ Identitet je, stoga, djelomično određen okolinom, tj. kulturom kojoj netko pripada. Međutim, kad se ne pronalazimo unutar tog okruženja, dolazi do krize identiteta.⁴² Dio našeg identiteta je znati tko smo i što smo, tj. koja je naša pozicija u odnosu na okolinu. Idealni karakter je onaj čiji se identitet ne sastoji od sebstva koje je nastalo pod utjecajem drugih. Međutim, tu se dolazi do problematike toga što mi naše sebstvo „imamo“ u odnosu na druge.

⁴¹ Usp. Taylor, Charles, *Izvori sebstva*, Naklada Breza, Zagreb, listopad 2011., str. 39.

⁴² Usp. *ibid.*, str.

Nadalje, oblikovanje identiteta kojeg mi sad imamo, započelo je već u antičkoj Grčkoj. Cijela povijest sve do sadašnjeg trenutka utjecala je na ono tko smo mi u sadašnjem trenutku. Naš identitet pripada nama onoliko koliko pripada i našim precima. Ali što to onda znači za identitet UI? UI nema pretka niti povijest u onom smislu i na onaj način na koji to ima ljudski rod. Preci UI su današnje varijante UI koje koristimo svakodnevno. Međutim, ti oblici UI koji će se razviti u jaču i još jaču UI, neće svojim nasljednicima prenijeti sve one osobine koje smo mi naslijedili od naših predaka. Razvoj UI i njeno poboljšanje ne ovisi o procesu evolucije, već ovisi o njenom ljudskom stvaratelju. Tek kod ASI razine UI se može početi govoriti o procesu samostalne evolucije. Nadalje, kao što je već spomenuto, na razvoj identiteta utječe i okruženje i tuđi identiteti. Budući da razvoj UI vodi prema scenariju u kojem će *homo sapiens* koegzistirati sa *robo sapiens*-om, a vodeći se takvom teorijom o identitetu, u jednom trenutku UI će nas oblikovati onoliko koliko i mi nju. Na ovo koegzistiranje ljudi i ASI razine UI, može se primijeniti ja – ti i ja – ono odnosu kojeg je predstavio Martin Buber.

Prvi odnos, ja – ti, podrazumijeva dvije osobe koje utječu jedna na drugu ali zadržavajući svoju jedinstvenost. S druge strane, odnos ja – ono podrazumijeva da se „ja“ prema drugome odnosi kao prema objektu „ono“. Ako UI na razini ASI bude prijateljski nastrojena naspram ljudi te ljudi naspram UI, govorimo o prvoj vrsti odnosa. Takva vrsta odnosa, teoretski, bi bila najizglednije ostvarenje jer se kroz razvoj UI stvara određena međuovisnost ljudi i UI. Ta međuovisnost se odnosi na uzajamno učenje i unaprjeđivanje. Stvaranjem UI, čovjek direktno utječe na stvaranje identiteta iste. Također vrijedi i obratno, razvojem UI mijenjaju se određeni dijelovi nas samih i utječu na daljnji razvitak našeg identiteta. Spoznajom da postoji UI koja se razvija u sve bolju i bolju verziju sebe, može imati pozitivan učinak na razvoj čovjeka, tako da se čovjek razvije u bolju tj., sljedeću evolucijsku fazu. Međutim, isto tako se može dogoditi negativan ishod koji će dovesti do toga da se čovjek prestane razvijati i prepusti se vladavini UI. Nadalje, ja – ono odnos podrazumijeva da čovjek na UI gledao kao na stvar koja se sastoji od metala i žica, te drži kontrolu nad njom zbog toga. Takav odnos povećava šansu da se uloge preokrenu, onog trenutka kad ASI razvije samosvijest. Budući da je ljudima nagon za održavanjem na životu jedan od primarnih nagona, s razvitkom samosvijesti i UI bi razvila taj nagon. U tom slučaju, čovjek postaje neprijatelj i preuzima „ono“ ulogu u odnosu ja – ono.

Obje varijante ovog odnosa imaju zajedničko razvoj samosvijesti UI. Taj događaj, koji je još daleka budućnost, bit će ključan za pitanje identiteta UI. Problematika samog tog pitanja proizlazi iz jedne druge problematike. Naime, i kod ASI razine UI i dalje govorimo o nečemu što je „samo“ spoj žica, ne biološkog podrijetla. Problem leži u tome što, uzimajući to u obzir,

ASI ne možemo smatrati osobom *per se*. Počevši od Aristotelove jednosti bitka, preko Leibnizove teorije prema kojoj je stvar ono što jest jer nije nešto drugo, došlo se do toga da osobu čini osobom imanje vlastitog jedinstvenog bitka. UI ima bitak na drugačiji način od onoga kako ga čovjek ima. Stoga, čovjek jest osoba, ali UI jest vrsta bića koja svoj jedinstveni bitak ima na sebi svojstven način. Identitet i bitak UI se ne bi trebali umanjivati na razinu nižu od one koju čovjek posjeduje samo zbog argumenta da UI nije biološko biće. Kad UI razvije samosvijest te postane ASI, otvorit će se novi pojam, a to je nova vrsta identiteta i definicija koja se odnosi na ne biološka bića.

ZAKLJUČAK

Kao zaključak, želim prikazati da je cilj ovog teorijskog rada izložiti potencijalne varijante definicije identiteta umjetne inteligencija. Premda se toj temi prilazi s dva stajališta, naime teološkog i filozofskog, ovaj rad baziran je na filozofskom stajalištu. Počevši s definiranjem same inteligencije, te kroz razvoj umjetne inteligencije te teoretska predviđanja o daljnjem razvoju iste, došlo se do pitanja identiteta. Problem pitanja identiteta umjetne inteligencije također problematizira i samu činjenicu možemo li UI smatrati nekom vrstom osobe kako bismo uopće mogli reći da posjeduje svoj jedinstveni identitet. Jasno je da UI nije osoba prema onoj definiciji prema kojoj je čovjek osoba. Međutim, ako se na razvoj UI, posebice na teoretsko postojanje ASI razine, gleda kao na novu eru u evolucijskom procesu, govori se o novoj definiciji osobe. Štoviše, postojanje ASI razine zahtijevat će da se promjene definicije koje se tiču osobe i jastva ili da se uvedu nove definicije koje će se odnositi na biće *robo sapiens*.

Premda UI ima riječ umjetna u svom nazivu, i „samo“ je kopija biološke inteligencije, ako se ostvare predviđanja o razvitku samosvijesti i nastanku ASI generacije, doći će vrijeme kada ta riječ neće imati istu konotaciju koju ima danas u nazivu UI. Pri problematiziranju identiteta, spomenut je ja -ti odnos i moguće varijante tog odnosa. Uzevši u obzir da se UI prilagođava tj., upija sve iz okoline i prema tome se oblikuje, ona će svoj „ti“ modelirati prema našem „ja“. Pretpostavljajući samosvijest ASI razine UI, ona će imati na raspolaganju nekoliko vrsta „ja“ koje će moći oblikovati u svoje jedinstveno „ti“. Tako da, premda će u početku UI zrcaliti našu osobu, s vremenom će postati svoja osoba. Čovječanstvo će direktno utjecati na podlogu iz koje će se razviti identitet UI. Temelji za razvoj tog identiteta već su sada u našim rukama. Ono što je problematično u tome, jest odgovornost koja dolazi s tim. Ljudi, kao neka vrsta roditelja, imaju zadatak usmjeravanja UI kako bi se ona razvila na što bolji način.

Literatura:

1. Barrat James, Our Final Invention, Thomas Dunne Books, 1.10.2013
2. Boden Margaret A., Artificial Intelligence, a very short introduction, Oxford University Press, United Kingdom, 2018.
3. Fieser James, Great Issues in Philosophy, Oxford University Press, 2009
4. Filozofski leksikon, Leksikografski zavod Miroslav Krleža, Zagreb, 2013, str. 1194.
5. Gillabel, Dirk, Kako postati cjelovita osoba: Jungov proces individuacije, 1993
<https://existentia.com.hr/postati-osobom-jungov-proces-individuacije/> pristupljeno 10.3.2021.
6. Gregov, Iva, Mitovi i istine o egu, Psiha i emocije, 8.5.2018.
<https://www.adiva.hr/lifestyle/psiha-i-emocije/mitovi-i-istine-o-egu-sto-je-ego-te-je-li-nam-prijatelj-ili-neprijatelj/> pristupljeno 10.3.2021.
7. Hume, David (1978). A Treatise of Human Nature. Oxford: Clarendon Press.
8. Kovač Irena, Umjetna inteligencija Danas, Sveučilište Rijeka, 2015
9. McCarthy, J.; Hayes, P. J. 1969. Some Philosophical Problems from the Standpoint of Artificial Intelligence. U: Machine Intelligence. Ur. B. Meltzer, D. Michie. Edinburgh: Edinburgh University Press, str. 463–502.
10. Miley Jessica, 11 AI Movies and what they got right and wrong about the technology, 9.3.2018. Pristupljeno: 1.6.2020. <https://interestingengineering.com/11-ai-movies-and-what-they-got-right-and-wrong-about-the-technology>
11. Online Stanford Encyclopedia of Philosophy. Pristupljeno: 3.5.2020.
<https://plato.stanford.edu/>
12. Rothman Joshua, Daniel Dennett's Science of the Soul, The New Yorker, 20.3.2017. Pristupljeno: 16.4.2020. 35.
<https://www.newyorker.com/magazine/2017/03/27/daniel-dennetts-science-of-the-soul>
13. Russell S. i Norvig P. Artificial intelligence: a modern approach, Prentice-Hall Inc., New Jersey, 1995.
14. Steels Luc, Brooks Rodney, The Artificial Life Route to Artificial Intelligence, Lawrence Erlbaum Associates Publishers, United Kingdom
15. Taylor, Charles, Izvori sebstva, Naklada Breza, Zagreb, listopad 2011.,

16. Tegmark Max, Life 3.0 Being Human in the Age of Artificial Intelligence, Alfred, A. Knopf, 2017
17. Urban Tim, The AI Revolution: The Road to Superintelligence, 22.1.2015., Pristupljeno; 12.5.2020., 1. https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html?fbclid=IwAR0ZpwdDpy-wYfmbo2MsQDPUtKltN38TGpcyrNG8mePJrpcpvPvBoKslO_A?fbclid=IwAR0ZpwdDpy-wYfmbo2MsQDPUtKltN38TGpcyrNG8mePJrpcpvPvBoKslO_A

Sažetak

Pitanje identiteta umjetne inteligencije

Tema ovog rada je pitanje koje će u budućnosti postati naša stvarnost, a to je pitanje identiteta umjetne inteligencije. U ovom trenutku u povijesti, mi ne samo da živimo u dobu tehnologije i napretka, nego smo i svjedoci razvoja svijeta na novoj tehnološkoj razini. Međutim, taj napredak kojem svjedočimo i u kojem sudjelujemo, jednog dana će dovesti do postojanja nove razine umjetne inteligencije koja će se od trenutno već postojeće razlikovati u tome što će nadmašiti čovjeka u svakom smislu. Tad ćemo se naći pred pitanjem identiteta nečega što nije biološko, već nadmašuje biološka bića. Ovaj rad istražuje razvoj umjetne inteligencije i razne teorije o njenom daljnjem razvitku. Proučavajući razvitak robotike i filozofije uma, dolazi se do razmatranja identiteta samog čovjeka i suživota *homo sapiensa* i *robo sapiensa*.

Ključne riječi: umjetna inteligencija, identitet, tehnologija, čovjek, robot

Summary

The Question of the Identity of Artificial Intelligence

The topic of this paper is the question that will become our reality in the future, and that is the question of the identity of artificial intelligence. At this point in history, we are not only living in an age of technology and progress, but we are also witnessing the development of the world at a new technological level. However, this progress that we are witnessing and participating in, one day will lead to the existence of artificial intelligence that will differ from the one that already exists in the way that it will surpass man in every sense. Then, we will find ourselves faced with the question of the identity of something that is not biological but transcends biological beings. This paper explores the development of artificial intelligence and various theories about its further development. By studying the development of robotics and the philosophy of mind, we come to consider the identity of the man himself and the coexistence of *homo sapiens* and *robo sapiens*.

Keywords: artificial intelligence, identity, technology, man, robot