MOŽE LI VJEŠTAČKA INTELIGENCIJA IMATI SAMOSVIJEST?

Nemanja Tubonjić

Univerzitet u Banjoj Luci, Filozofski fakultet, BiH
Katedra za filozofiju
nemanja55@mail.com

https://doi.org/10.7251/FPNDP2304077T

APSTRAKT:

Kada govorimo o mogućnosti vještačke inteligencije veoma često se susrećemo sa onim strahom od posljedica razvoja ove inteligencije. Ovaj strah ponekad jeste iracionalan, ali otvara pitanja ne samo etičkog tipa u smislu primjene vještačke inteligencije nego i u aspektu stava da vještačka inteligencija može razviti samosvijest i u određenom momentu zamijeniti čovječanstvo. Trenutni razvoj i brzina kojom vještačka inteligencija operiše ponovo je aktualizirao ova pitanja i ponovo postavio jedno od najstarijih pitanja filozofije – šta je svijest, tj. šta je samosvijest? Da li iza određenog programa samo stoji kompleksnost algoritama koji funkcionišu prema principu realizacije onoga što je u njih učitao programer ili postoji potencijal za nešto više? Ovo je ujedno i pitanje prirode ljudske samosvijesti, jer kako je moguće da kažemo ono Ja ukoliko je kompleksnost ljudskog mozga samo manifestacija bioloških procesa koji se dešavaju u mozgu i generalno ljudskom tijelu? I može li se po ovom istom principu, ukoliko smo npr. sposobni da reprodukujemo ljudski mozak u smislu tehnoloških aparata, ujedno formirati i samosvijest unutar AI algoritama? O kakvoj se dijalektičkoj prirodi ovdje radi? Ovaj rad neće ponuditi definitivan odgovor jer takav odgovor ne može ni nauka u trenutnoj fazi da ponudi. Ali ovo postaje pitanje čovjeka u tehničkom dobu i cilj rada biće upravo da otvori pitanja ljudske prirode u tehničkom dobu zajedno sa time može li ova priroda da oformi tehničku prirodu i manifestuje kroz čisto tehničku volju.

Prethodno saopštenje

UDK:
007.52:133.529.6

Ključne riječi:
vještačka inteligencija, samosvijest, algoritam, tehnologija, potencijalnost

Uvod

Mogućnosti vještačke inteligencije toliko su fascinantne i toliko brzo napreduju da u čovjeku stvaraju određenu vrstu straha. Strah koji se ovdje javlja nije uvijek opravdan, više proizilazi iz domena nekih fantastičnih razmatranja, ali u svojoj srži sadrži određene etičke probleme, kao i probleme koji se javljaju u sferi mogućnosti razvoja samosvijesti unutar vještačke inteligencije. Ali, šta je svijest? Da li je svijest skupina učitanih informacija ili posjeduje akt samostalnosti? Kako

je moguće reći Ja? Može li tehnika da reprodukuje svjesnost? Da li je ovdje riječ o logičkoj ili dijalektičkoj prirodi?

Ovdje nema definitivnih odgovora. Osnovni problem nije više da se definiše ljudski bitak kroz neposredan odnos sa tehnikom, nego u smislu da se ovim zajedničkim bitkom čovjeka i tehnike stvori novo dijete tehničkog doba – samosvijest tehnike. Također, ovdje se javlja i etičko pitanje mogućnosti ove samosvijesti – kako čovjek treba postupati prema njoj i može li se ovdje stvoriti jedna transhumanistička sinteza? Čovjek je pokretač vlastite evolucije. Tako će ovaj rad otvoriti pitanje koliko su mogućnosti tehnologije danas ograničene, pitanje ljudske samosvijesti, kao i krajnje etičko pitanje odnosa tehnologije i čovjeka i potencijalnih oblika ovog odnosa. Ovdje je važan i pristup ovom pitanju i udaljavanje od straha i panike jer ukoliko možemo govoriti o samosvijesti onda moramo govoriti o njenoj dvojnoj prirodi – prirodi koja čini dobro i prirodi koja čini zlo. Ukoliko potencijal za prvu prirodu postoji onda nužno esencija našeg kretanja mora biti taj put.

Već decenijama se govori o mogućnostima računarske tehnike, odnosno govori se o tome koliko ova tehnika može da usavrši ljudski život, ali kroz ova pitanja mogu se razmatrati i mnoga pitanja unutar filozofije samosvijesti. Samosvijest je jedna od najkompleksnijih fenomena unutar prirodnog svijeta. Samosvijest ne samo da je po sebi fascinantna i jedna od najkompleksnijih stvari, nego je otvorila niz kako kritičkih tako i dogmatskih pitanja. Nerazumijevanje ovog fenomena dovodilo je do razvoja raznih religijskih teorija, pri čemu je ljudski duh pokušao objasniti sam sebe i dati definitivne odgovore o ustrojstvu stvarnosti, ali nikada nije uspio da dođe do potpune spoznaje o samom sebi i vlastitoj prirodi. Ukoliko ljudski duh ne zna šta je njegova vlastita samosvijest, ukoliko ne zna kako ona funkcioniše i kako se manifestuje, koji zakoni iza nje stoje, onda čak i ukoliko stvori tu mogućnost samosvijesti tehnike i dalje neće znati kako je to napravio. Na taj način ne otvara se samo pitanje može li čovjek napraviti nešto, nego se otvara pitanje da li on razumije ono što je napravio i može li iskoristiti svoja tehnička sredstva ne samo da razumije samoga sebe nego da razumije i ono što je izvan njega.

1. Trenutne mogućnosti računarske tehnike

Osvrt na pitanje o mogućnostima računarske tehnike davno je dao i Hubert L. Dreyfus u svojoj knjizi *"Šta računari ne mogu"* koja je objavljena 1972 godine. U okviru ovog poglavlja osvrnućemo se na neka od njegovih zapažanja iz navedene knjige, tj. uporedićemo ih sa nalazima nekih modernijih autora. Iako se od 1970-ih godina do danas mnogo toga značajno izmijenilo u računarskoj tehnologiji, ovdje se javljaju pitanja filozofske prirode koja po nekom svom sistemu i jesu najosnovnija pitanja koja zahvataju cjelokupnu ljudsku istoriju. Možda će pravac kojim će Dreyfus krenuti biti pravac odnosa između logičkih sistema

koji operiraju unutar računarske tehnologije i čovjeka kao po prirodi dijalektičkog bića koje u sebi sadrži element neodređenosti. Ovaj momenat neodređenosti esencijalan je za shvatanje i principa slobodne volje, ukoliko predpostavimo da takav princip može postojati usljed mogućnosti da se kaže ono Ja. Da li je ljudska inteligencije, pa i sama bit ljudske prirode, nereciplirajući fenomen? Da li računari mogu da na određen način razumiju kontekst unutar koga se dešava specifično iskustvo i da na osnovu toga daju zadovoljavajući rezultat ili transformiraju svoju logičku strukturu u dijalektički red? Može li računar postati misleća stvar?

U svojoj knjizi *''Prednost umjetne inteligencije"* (2018) autor Thomas H. Davenport će objasniti da pod AI tehnologijama razumijevamo sposobnosti tehnologije koje su ranije posjedovali samo ljudi. Riječ je o određenim znanjima i uvidima, obradi specifične vrste poslovnih zadataka i procesa, dosegnut nivo automatizacije putem algoritama, eliminišući određen nivo ljudskog napora. I čini se da Davenport ne naglašava potencijalnost AI tehnologije za samostalnosti, koliko je svodi na čistu automatizovanu ljudsku praksu. Time se suština AI-a nužno svodi na ljudsku praksu, tj. na ostvarivanje ljudskih ciljeva. Ali, američki naučnik i profesor informacionih tehnologija na Univerzitetu Berkley iz Kalifornije, Stuart Russel će u svojoj knjizi *''Kao čovjek: Umjetna inteligencija i budućnost naših vrsta"* (2019) postaviti pitanje o prirodi same inteligencije. On će objasniti da je inteligencija ona suština koja nas određuje i da su ljudska bića inteligentna ukoliko teže ostvarivanju vlastitih ciljeva. Svijest o vlastitim ciljevima jeste akt samosvijesti. Stuart će reći da su mašine inteligentne samo utoliko ukoliko njihovi postupci vode ka ostvarivanju njihovih ciljeva, a pošto mašine nemaju vlastite ciljeve i nisu sebi svrha, one nužno ostvaruju ljudske ciljeve. Iz ovoga možemo zaključiti da dok god mašina ili određeni algoritam ispred sebe ne postavi vlastiti cilj u službi samoga sebe nije moguće govoriti o nekoj njegovoj samostalnosti ili čak svjesnosti.

1.1. Pojedinačni elementi AI tehnologije

Primjećujemo da su akti AI tehnologije samo fragmentirani akti ljudske prakse, a esencijalnost ljudske prakse jeste razumijevanje konteksta u kome se ova praksa manifestuje, što je prije svega zanimljiv fenomen unutar kulture ljudskog jezika kao sredstva prenošenja informacija i razumijevanja. Ovo pitanje jezika i komunikacije pogotovo je aktuelno u smislu AI softvera za prevođenje. Kada čovjek prevodi tekst sa jednog jezika na drugi on ne može prevoditi taj tekst na čisto logički, bukvalan način. On prevodu mora dati kontekst, mora razumjeti specifična pravila, tj. mora poznavati prirodu samog jezika na koji se prevodi. Dugi niz godina programi za prevođenje nisu davali toliko zadovoljavajuće rezultate, nego su više pomagali prevodiocima u usavršavanju njihovih prevoda. Ali trenutno svjedočimo da automatski prevodioci, poput Google Translate-a, svakim danom bivaju sve bolji i bolji.

Iako ovaj program koristi veliku bazu podataka za prevode, kao i najrazvijenije algoritme, on i dalje neće biti u mogućnosti da prepoznaje kontekst u jednom generalnom smislu. Čak i ukoliko se u njega učitaju jezička pravila sklapanja rečenica unutar prevoda ovdje je riječ o čisto matematičkim formulama.

Može li automatski prevodilac da razumije fraze, umjesto da fraze prevodi onako kako su sistemski u njega učitane, koliko god one bile kompleksne? Na ova pitanja će se osvrnuti i Dario Stojanovski (2021). U radu o mogućnosti razumijevanja konteksta u softverima za prevođenje Stojanovski ukazuje da kvalitet prevoda više zavisi od povećanja količine unesenih podataka nego od pokušaja dodavanja samog konteksta. Drugim riječima, ukoliko bi se ovakav softver usavršio na način da se obrađuju informacije iz riječi u rečenici ili čak čitavih odlomka teksta na način da se rekonstruiše sam kontekst onda bi kvalitet prevoda bio daleko bolji. Ali to opet ne bi bilo ''razumijevanje" konteksta koliko mogućnost unaprijeđenog softvera da vadi fragmente informacija i shodno algoritmu ponovo ih slaže u ono što mi percipiramo kao specifičan kontekst. Na osnovu toga možemo uvidjeti da dalji razvoj ovakvih programa zavisi isključivo od čovjeka, tj. da njihova ograničenja počivaju na logičkim zakonitostima – uslov za razvoj samosvijesti jeste mogućnost proizvođenja nečega novog, potencijalno samostalno, a logika ne može dati ništa novo.

1.2. Ka pojmovnom određenju (vještačke) inteligencije

Na osnovu prethodno navedenog javlja se potreba i za preispitivanjem pojma ''inteligencija". Prema definiciji koju nudi „Encyclopedia Britannica" ljudska inteligencija predstavlja sposobost „za učenje na osnovu iskustva, prilagođavanje novim situacijama, razumijevanje i korištenje apstraktnih koncepata, kao i korištenje znanja u svrhu manipulacije okruženjem" (2022). Inteligencija bi bila sveukupnost mentalnih procesa jednog bića, ali opet veoma je teško utvrditi šta ti procesi jesu. Da li onda računarska tehnologija ima inteligenciju ukoliko se u nju učitavaju specifični algoritmi za rješavanje određenih problema i dostizanje rezultata? Inteligencija bi trebala predstavljati određenu samostalnost, svjestan akt, mogućnost da ono Ja kaže Ja bez da se taj govor o Ja apriori učita. Da li onda danas možemo govoriti o vještačkoj inteligenciji? Dreyfus (1972) tvrdi da bi vještačka inteligencija postala stvarnost ukoliko bi bili u mogućnosti da stvorimo ili repliciramo vještački nervni sistem koji bi bio sličan ljudskom. Tek tada bi bilo riječi o pravoj vještačkoj inteligenciji. Sve do tog momenta govorimo o simulaciji inteligencije, govorimo o iluziji inteligencije, tj. samo o fragmentiranoj ljudskoj inteligenciji prenesenoj u određene algoritme da olakšaju ljudsku egzistzenciju. U to vrijeme on će zaključiti da pri trenutnom stanju nauke takav poduhvat nije izvodljiv, a svjedoci smo da računarska tehnologija još uvijek nije na tom nivou razvoja.

Repliciranje svijesti može se desiti na dva načina – svjesno i slučajno. Svjestan

akt repliciranja znači razumjeti prirodu svijesti, čitav konglomerat mentalnih procesa, što još uvijek uveliko izmiče neuronauci. Drugi put jeste slučajan akt – da putem usavršavanja računarske tehnologije i uvezivanjem njene fragmentarnosti slučajno dođemo do akta svjesnosti. Pošto ne znamo prirodnu genezu svijesti unutar organske materije isto tako bili bi zbunjeni genezom svijesti unutar tehnološke ''materije". Možda faktor pri tome može biti i usavršavanje računarske tehnologije kroz sintezu sa samim čovjekom, tj. već započeta faza transhumanizma. Tako imamo jedan aktuelni fenomen gdje se putem AI tehnologije mogu čitati ljudske misli – a dešifrovanje ljudskih misli i razumijevanje njihove manifestacije jeste korak ka razumijevanju prirode svjesnosti. O ovoj tehnologiji pišu Tang, LeBel, Jain i Huth (2023), pri čemu se kroz analizu EEG snimaka fokusiraju na razumijevanje i rekonstrukciju sadržaja govora kao složenog moždanog procesa. Ali i u ovom slučaju algoritam koji prati moždane talase nije sposoban na akt nepredvidljivosti. I na osnovu tog primjera možemo uvidjeti koliko smo još daleko od toga da AI stvarno transcendira svoje logičke operacije.

Drugi zanimljiv fenomen samosvijesti jeste igra. Igra je jedna od elementarnih sposobnosti ljudske svijesti i predstavlja akt inteligencije kao intencije prema objektu izvan nas. Ona počiva na analiziranju stvarnosti, ali podrazumijeva i mogućnost greške. Čak i tamo gdje najbolji majstori određene igre mogu da daju savršene analize potencijalnih ishoda postoji mogućnost greške, neka nepredvidiva opcija. Ali, ukoliko računarska tehnologija ima mogućnost da analizira, vremenski veoma brzo, svaki mogući ishod i potencijalni potez, utoliko ona unaprijed determiniše ishod igre, tj. utoliko ju je teže pobijediti. Drajfus (1972) će napomenuti da nismo u mogućnosti da preispitamo sve pravce i ishode određene igre, poput šaha, a da bi to za mašinu takođe predstavljalo (vremenski) izazov. Ali, 2016. godine izaće članak u časopisu „Nature" u kome će biti opisan razvoj sistema šahovske igre na računaru gdje će kompleksnost računarske tehnologije nadmašiti ovu ograničenost, tj. kako usavršavanje algoritama omogućava računarskom sistemu da nadmaši kapacitete ljudi u ovom specifičnom kontekstu. Tako će Hassabis, Huang, Maddison i drugi autori (2016) za primjer uzeti drevnu igru Go, tj. situaciju u kojoj je računar pobijedio evropskog šampiona ove igre. Razlog ovakvog ishoda jeste u tome što se novi oblik softvera, obučen kroz veliki broj partija, razvio u toj mjeri da velikom brzinom analizira strategije, poteze i potencijalne ishode iz prijašnjeg igračkog iskustva. I taj primjer ukazuje da tehnologija s vremenom može značajno da uznapreduje, a da opet iz te kompleksnosti ne razvije ništa više od onoga što je već upisano u nju.

Dreyfus (1972) će u jednom misaonom eksperimentu predstaviti čovjekovu svijest iz perspektive računarske tehnologije; tj. prema njegovoj biološkoj osnovi – funkciji neurona koji obrađuju informacije, psihološkoj osnovi – prema kojoj računar služi kao model uma gdje bi bila riječ o određenoj vrsti softvera koji daje

pravila, epistemološkoj osnovi – prema kojoj se znanje može formalizovati i izraziti putem logički relacija, te ontološkoj osnovi – koja bi uvezala niz logičkih činjenica u sliku svijeta. Ako redukujemo svjesnost na ove faktore, tj. ako je fragmentiramo, vidjećemo da nijedan od ovih faktora ne sadrži akt svjesnosti unutar sebe. Nijedan od ovih faktora ne može za sebe reći Ja. Nijedan neuron, nijedan fragment ljudskog mozga, ne sadrži u sebi ono Ja, jer je ogoljen od svake volje i podložan je nužnim zakonima kretanja organske materije u specifičnom stanju. A pošto je nauka jedno kompleksno područje, koje usljed toga mora ostati fragmentirano, ostaje pitanje može li se svijest objasniti pojedinačnom analizom dijelova ljudskog mozga? Može li nauka koja se bavi izučavanjem neurona objasniti šta je svijest ili može samo konstatovati da su neuroni osnov svijesti, pa dalje opisivati njihov rad i funkciju?

Italijanski naučnik Giulio Tononi je razvio teoriju svijesti kao konglomerata integrisanih informacija (2008). On će dati primjer kineskog pisma i objasniti da računar može prepoznati kineska slova i čitati ih, ali ne može biti svjestan svog djelovanja jer ne posjeduje dovoljno integrisanih informacija. Ukoliko pretpostavimo da je svijest konglomerat integrisanih informacija, tj. ukoliko pretpostavimo da se ovo dešavalo milionima godina evolucije, onda moramo pitati gdje počinje, a gdje prestaje granica svjesnog i nesvjesnog? Da li količina upisanih informacija utiče na to da li je neki softver svjesniji? Na osnovu prethodno navedenih pretpostavki o inteligenciji ovo ne bi bilo moguće, jer što je kompleksniji algoritam i što više informacija ima u sebi on postaje samo svestraniji, što nije akt svjesnosti, odnosno mogućnost onog Ja. Akt svjesnosti ne mora unutar sebe imati nikakve informacije sem apriori prirodno datih mehanizama koji manifestuju ono Ja. ChatGPT unutar sebe ima višestruko više informacija nego jedno dijete, a opet dijete ima ono Ja za razliku od ChatGPT – ili bilo kog drugog AI rješenja. Šta znači ovo – biti svjestan? Osjećati bitak za sebe, biti u bitku, osjećati individualnost, biti u mogućnosti reći Ja – ne samo u ljudskom smislu. Mnoge životinjske vrste imaju određen nivo svjesnosti, npr. neke od njih anticipiraju smrt jer je se boje, što je opet dokaz svjesnog bitka. Iako i dalje ne znamo konkretne uzroke tih mentalnih procesa, možemo ih opisivati, te u(s)tvrditi da ih postojeća softverska rješenja tek oponašaju. Drugim riječima, još uvijek kompleksnost nekog računarskog programa ne može manifestovati, već samo simulirati akt svjesnosti – jastvo.

2. Ontološki problem slobodne volje na primjeru simulirane stvarnosti

Ukoliko govorimo o samosvijesti moramo govoriti i o konceptu slobodne volje. U tom smislu, moguće je postaviti pitanje da li je ovaj koncept slobodne volje moguć unutar računarske tehnologije, odnosno i da li je slobodna volja ontološki moguća u realnom svijetu. Jedan od najinteresantnijih oblika računarske tehnologije kroz koju se može sagledati uticaj AI-a jeste simulirana stvarnost.

Simulirana stvarnost, koja kroz video igre svakodnevno postaje sve naprednija i naprednija, unutar sebe sadrži niz kompleksnih algoritama. Na ovo će se osvrnuti i Safadi, Fonteneau i Ernst (2015) u svom tekstu *"Artificial Intelligence in Video Games: Towards a Unified Framework"* gdje će pokušati objasniti da stvaranje kompleksnog autonomnog subjekta koji prelazi određenu video igru van samog čovjeka jeste moguće, ali je veoma kompleksno i značilo bi repliciranje velikog dijela ljudske inteligencije. Također, stvaranje određenog AI za prelaženje određene igre značilo bi nemogućnost primjene tog AI-a na druge igre i opet bi pokazalo nemogućnost da se replicira ljudska svestranost. Ukoliko kažemo da su video igre oblik računarske tehnologije visoko kompleksnog nivoa unutar kojih se učitava određena priča i njen grafički prikaz, odnosno niz kompleksnih algoritama koji simuliraju stvarnost, možemo li onda govoriti o liku unutar video igre, tj. određenom programu koji bi imao potenciju za samosvijest?

2.1. Metafizika virtuelnog svijeta

Unutar svake video igre postoji više-manje linearni put kretanja glavnog lika čiji je cilj ispričati njegovu priču. On unutar te igre simulira autonomnost svojih odluka iako su svi mogući pravci njegovog kretanja već unaprijed učitani u igru i predodređeni. On može simulirati govor o svojoj volji, on može simulirati samosvijest, ali sve što se dešava u rukama je onog bića koje upravlja igrom. Čak se sve ono što se dešava transcendira iz ruku onog koji igra na polje već učitanog programa, koji unutar sebe ima svaki mogući ishod, uključujući krajnji ishod same igre. Da bi vjerodostojnije simulirali svijest i autonomnost igrača kreatori igara sve češće odbacuju koncepciju „linearnog dešavanja" , tj. omogućavaju igraču biranje između više opcije i na taj način stvaraju privid slobodne volje. Ishodi sada postaju različiti i tako se određena igra ne mora završiti na samo jedan ili dva načina, nego čak na tri, četiri ili više načina. Ali svaki od ovih ishoda i dalje je nužno određen programiranjem, unaprijed je određen algoritmima, odnosno već je (za)dat prije nego je igrač krenuo da igra. Zamislimo igru koja ima beskonačan niz potencijalnih ishoda koji se proizvode kontinuiranim učitavanjem i igrača koji ide putem beskonačnog niza različitih pravaca. On bi imao čistu simulaciju svoje autonomije u biranju tih pravaca i stvaranju ishoda i simulirao bi slobodnu volju, iako ona unutar te igre nužno ne postoji. Ukoliko ovu logiku prenesemo na realni svijet, ukoliko kažemo da se sve dešava po nekim nužnostima, da su naše odluke determinisane onim što stoji ispred nas i da je akt svjesnosti akt svjesnog bitka, ukoliko pritom kažemo da je realitet kontinuirani bitak u promjenjivosti, onda je svaki izbor, svaki ishod, svaka potencija, nužno apriori određena nečim, nužno apriori moguća i nužno će se desiti. Ukoliko bitku dajemo status apsoluta i time isključujemo nebitak, ukoliko bitku dajemo status permanentne promjenjivosti, utoliko nismo u mogućnosti

da isključimo niti jednu potencijalnost bitka da bude i ako kažemo da je sve što biramo određeno motivacijom prema onome što je izvan nas, čak i ako odlučimo da ne biramo ništa, onda koncept slobodne volje pada.

2.2. Mogućnost slobodne volje u (simuliranoj) stvarnosti

Na tu problematiku referiše i Benjamin Curtis u tekstu *''Free Guy's philosophy: could we just be lines of code in a grand simulation?"* (2021). Na primjeru filma Free Guy, gdje je riječ o liku iz igre koji postaje samosvjestan, Curtis postavlja pitanje imamo li mi slobodnu volju i koja je razlika između ''uma" koji funkcioniše prema programskim zakonima jednog računara i uma koji funkcioniše prema biološkim zakonima mozga? Curtis poredi jedan program čije su aktivnosti rezultat kompjuterskih operacija nad kojim taj program nema nikakve kontrole sa biološkim operacijama koje se dešavaju unutar našeg mozga nad kojima mi također nemamo nikakve kontrole i postavlja pitanje u čemu je onda razlika. Ukoliko ne želimo da zapadnemo u religijski dogmatizam i kažemo da za razliku od programa ljudi imaju nešto ''vantjelesno" – dušu, gdje se otvara pitanje definisanja biti duše i njenog realiteta koji opet ne može biti apstraktan, jer bivstvuje i nečim je određen, jedini zaključak može biti da je svjestan um proizvod biološke materije čiju funkciju u potpunosti još ne razumijemo. Curtis pita kako onda možemo reći da svjestan um ne može proizaći i iz nebiološke materije ukoliko ona može biti jednako kompleksna kao i biološka? On ne daje odgovor na to pitanje, ali ono može poslužiti kao dobar osnov za promišljanja o našem realitetu.

Sve naše odluke jednake su odlukama lika iz video igre. Sve naše odluke određene su Drugim u najširem smislu. Za nas taj Drugi jeste bitak izvan nas, za lika iz video igre to je algoritam čiji je on sam dio kao što smo mi dio bitka. I čini se da niko od nas nema potenciju za samosvijest, samo simulaciju samosvijesti. Ali za razliku od lika iz video igre čovjek je svjestan svoga bitka, on misli i može da negira sve sem da misli. On ima potenciju da kaže za sebe Ja – jastvo mu je esencijalno i ono se osjeća. Čak i da je svijet čitava simulacija akt svjesnog bitka nije moguće simulirati. I ukoliko se odredimo prema nekom determinizmu, ukoliko kažemo da smo određeni onim što je izvan nas, akt spoznaje svijeta, akt sumnje i akt onog Ja nužno u sebi sadrži neki oblik autonomije. Ukoliko se određujemo prema ovome onda naš zaključak ne može biti ništa drugo do da je sloboda spoznata nužnost. Ukoliko govorimo o mogućnosti za istinsko jastvo unutar AI ono mora funkcionisati prema istom principu prema kom funkcioniše čovjek. Nauka taj princip još ne može objasniti, a filozofija može ostati samo na polju ispitivanja.

Pored toga, moguće je povući još jednu ontološku paralelu između računarskih simulacija i stvarnosti. Što više ishoda jedna video igra ima utoliko je ona kompleksnija i zahtjeva više računarske memorije i snage za procesiranje svojih programa.

Koji je onda kapacitet bitka? Vidimo da se na primjeru simulacije bitka otvaraju pitanja samog bitka, tj. da potraga za sviješću i slobodnom voljom unutar računara prestaje da bude strogo tehnološko pitanje – ono u suštini postaje osnov za temeljno ontološko i antropološko (samo)propitivanje. Iz te perspektive smatramo da ne treba toliko strahovati od potencijalnosti samosvijesti tehnologije, već prije od nepoznavanje prirode same samosvijesti.

Transhumanistička transcendencija ili umjesto zaključka

Ukoliko ovo pitanje samosvijesti vještačke inteligencije trenutno postavimo kao jednu nemogućnost i ukoliko kažemo da postoji neka potencijalnost u razvoju ovog koncepta u budućnosti, ostaje i dalje neodgovoreno pitanje o tome šta vještačka inteligencija znači za čovjeka. Da li je možda vještačka inteligencija samo produženje čovjekove inteligencije u smislu njegove brzine u spoznaji svijeta? Može li čovjek da pređe ograničenja vlastite prirode? Pošto je čovjek usavršio vlastite mehanizme spoznaje kroz razvoj tehnike, tehnika je uvijek bila određena vrsta oruđa koje čovjek koristi kao nešto izvan njegove tjelesnosti. Više kao produženje čula iz vana prije nego produženje čula u samom čovjekovom biću. Ali, razvoj moderne tehnologije uspio je da pređe i ova ograničenja. Tako vidimo da već sada postoje projekti poput stapanja organske materije sa tehnologijom, od jednostavnijih tehnoloških implantata pa sve do Muskovog Neuralinka. Može li čovjek zapravo da pređe ograničenja vlastite prirode?

Pojam transhumanizma ovdje razumijevamo kao proces putem kog se nadilazi prirodni bitak čovjeka kroz njegovo stapanje sa tehničkim svijetom. U svom članku o budućnosti transhumanizma McKie (2018) će objasniti da ideja o unapređenju ljudskog tijela nije nova i da smo kroz vijekove radili na tome da tehničkim sredstvima nadoknadimo neke fizičke nedostatke ili ograničenja. Ono što nam budućnost donosi jeste potencijalno spajanje čovjeka sa mašinom što bi rezultiralo ujedno transformisanjem njegove cjelokupne prirode. Povećanje inteligencije, snage, dužine života, odnosno unapređenje drugih organskih procesa, predstavljeni su kao temeljni zadaci unutar transhumanisitčkog pokreta. Uzimajući u obzir aktuelne pomake u okviru tehnologija za unapređenje ljudi (eng. human enhancement technologies) smatramo da se više i ne može raspravljati o tome da li ih (običan) čovjek želi ili ne želi. One su već neizostavan element ljudske evolucije i napretka. Jedino šta preostaje jesu etička pitanja načina primjene specifičnih tehnoloških pronalazaka. Ukoliko se vještačka inteligencija zadrži na svojim logičkim okvirima koji rješavaju specifične ljudske probleme utoliko ona može samo ubrzati procese ljudskog saznavanja svijeta, a ujedno transformisati i njegovu prirodu.

Tu više nije riječ o tome može li se proizvesti samosvijest, nego o tome da se ljudska svijest proširuje izvan vlastitih prirodnih određenja. Ukoliko smo vidjeli

da svi koncepti vještačke inteligencije funkcionišu po principima logičkih zakona van neke samostalnosti ili dijalektičnosti onda se iz tog začaranog kruga logičkih zakona nije moguće transcendirati. Za ostvarivanje tehnološke samosvijesti potrebno je nešto kvalitativno više, tj. u okvirima tehnologije moraće se desiti kvalitativni skok poput onog koji se desio na nivou kvalitativnog skoka organske materije pri razvoju ljudske inteligencije. Ukoliko se opet osvrnemo na Dreyfusa – potrebno je saznati funkcionisanje ljudskog mozga i cjelokupnog sistema koji proizvode svijest prije nego se upustimo u proizvođenje svijesti u onim okvirima materije gdje ona nije prisutna. A to znači da naši naučni okviri moraju postati širi i kompleksniji, odnosno mora im se dati vremenski prostor kako bi se razvili. Ovaj razvitak najprije mora doći u smislu razumijevanja šta je to čovjek kao organsko biće, odnosno razumijevanja njegove svjesnosti, prije nego što se upustimo u tehnička ostvarenja kompleksnijih sistema.

Ukoliko se zadržimo u okviru iskaza da ne znamo šta je svijest i da možda nikada nećemo ni saznati onda ispred nas stoji jedan specifičan put poboljšanja onoga što je već dato – stapanje organske materije sa tehnologijom, tj. stvaranje novog čovjeka tehničkog doba. Ali time smo već upali u jednu vrstu dogmatizma gdje smo rekli da nas jedno pitanje ne interesuje samo zato što još ne poznajemo njegov odgovor. Drugi korak jeste korak kritičke svijesti koja teži istini – šta je samosvijest? Odgovor na to pitanje bi podrazumijevao da se sve naučne i tehnološke snage moraju upregnuti u tom pravcu. Tokom tog procesa ne samo da bismo sticali nova saznanja o tome šta je čovjek kao organsko biće, nego bismo i stvarali pretpostavke za novog čovjeka iz transhumanističke perspektive.

LITERATURA

Curtis, B., 2021. Free Guy's philosophy: could we just be lines of code in a grand simulation?. *The Conversation*, August 20. https://theconversation.com/free-guys-philosophy-could-we-just-be-lines-of-code-in-a-grand-simulation-166389.

Davenport, H. T. (2021). *Prednost umjetne inteligencije*. Mate.

Dreyfus, L. (1977). *Šta računari ne mogu*. Nolit.

McKie, R., 2018. No death and an enhanced life: Is the future transhuman?. *The Guardian*, May 6. https://www.theguardian.com/technology/2018/may/06/no-death-and-an-enhanced-life-is-the-future-transhuman.

Russel, S. (2022). *Kao čovjek: Umjetna inteligencija – napredak ili prijetnja?*. Planetopija.

Safadi, F., Fonteneau, R., Ernst, D., 2015. Artificial Intelligence in Video Games: Towards a Unified Framework. International Journal of Computer Games Technology, 1, 1-30. doi: 10.1155/2015/271296.

Silver, D., Huang, A., Maddison, C., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484 – 489

Stojanovski, D. (2021). *Modeling Contextual Information in Neural Machine Translation* [Doctoral dissertation]. Univerzitet Ludwig Maximilian. https://edoc.ub.uni-muenchen.de/28411/.

Strenberg, R. J. (2022). Human intelligence. *Encyclopedia Britannica*. Pristup na: https://www.britannica.com/science/human-intelligence-psychology.

Tang, J., LeBel, A., Jain, S., & Huth, A. (2023). Semantic reconstruction of continuous language from non-invasive bran recordings. *Nature Neuroscience*, 26, 858 – 866. Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol Bull*, 215(3), 216-242. doi: 10.2307/25470707.

CAN ARTIFICIAL INTELLIGENCE HAVE SELF-AWARENESS?

Nemanja Tubonjić

University of Banja Luka, Faculty of Philosophy, BiH
Department of Philosophy
nemanja55@mail.com

ABSTRACT:
When we talk about the possibility of artificial intelligence, we often encounter fear of the consequences of its development. This fear may sometimes be irrational, but it raises questions not only of an ethical nature in terms of the application of AI, but also in terms of the belief that AI may develop consciousness and eventually replace humanity. The current development and speed at which AI operates has once again brought up these questions and has posed one of the oldest questions in philosophy - what is consciousness, or rather, what is self-awareness? Is the complexity of algorithms behind a particular program simply a manifestation of what the programmer has inputted, or is there potential for something more? This is also a question about the nature of human self-awareness, as how is it possible for us to say "I" if the complexity of the human brain is merely a manifestation of biological processes occurring in the brain and the human body in general? And can we, based on the same principle, form self-awareness within AI algorithms if we are, for example, able to reproduce the human brain in technological devices? What kind of dialectical nature is at work here? This work will not offer a definitive answer, as such an answer cannot be provided by science in its current stage. However, this becomes a question for humans in the technological age, and the aim of this work is to open up questions about human nature in the technological age, along with whether this nature can shape technological nature and manifest through pure technological will.

Keywords:
artificial intelligence, self-awareness, algorithm, technology, potentiality