

Etički i socijalni aspekti umjetne inteligencije

Strmečki, Gloria

Undergraduate thesis / Završni rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University North / Sveučilište Sjever**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:122:072728>

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

Download date / Datum preuzimanja: **2024-04-20**



Repository / Repozitorij:

[University North Digital Repository](#)



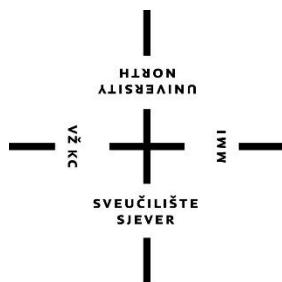


**Sveučilište
Sjever**

Završni rad br. 179/PMM/2019

Etički i socijalni aspekti umjetne inteligencije

Gloria Strmečki, 1409/336



Sveučilište Sjever

Odjel za Ime odjela
Završni rad br. 179/MM/2019

Etički i socijalni aspekti umjetne inteligencije

Student

Gloria Strmečki, 1409/336

Mentor

Sead Alić, prof. dr.

Prijava završnog rada

Definiranje teme završnog rada i povjerenstva

ODJEL	Odjel za ekonomiju		
STUDIJ	preddiplomski stručni studij Poslovanje i menadžment u medijima		
PRISTUPNIK	Gloria Strmečki	MATIČNI BROJ	1423/336
DATUM	15. 9. 2019	KOLEGIJ	Socijalna filozofija
NASLOV RADA	Etički i socijalni aspekti umjetne inteligencije		
NASLOV RADA NA ENGL. JEZIKU	Ethical and social aspects of artificial intelligence		

MENTOR	Sead Alić	ZVANJE	prof. dr.
ČLANOVI POVJERENSTVA	1. doc. dr. Dario Čerepinko, predsjednik		
	2. doc. dr. Željko Krušelj, član		
	3. prof. dr. Sead Alić, mentor		
	4. doc. dr. Lidija Dujić, zamj. član		
	5. _____		

Zadatak završnog rada

BROJ	179/PMM/2019
OPIS	

Umjetna je inteligencija već dio naše svakodnevice. Trenutno se intenzivno radi na različitim oblicima povezivanja snage i mogućnosti obrade podataka računalnih programa i ljudskog mozga koji u tom aspektu rada zaostaje za svojim vlastitim proizvodom.

Zadatak ovog rada je obraditi pojam umjetne inteligencije te analizirati moralna i socijalna pitanja koja se neminovno vežu uz uvođenje svake nove tehnike u ljudske oblike poslovanja, a posebno uz umjetnu inteligenciju koja uvodi i nova razmišljanja o definiciji čovjeka i ljudskoga.

Zadaci rada su, prije ostaloga:

- Definiranje razlike između poslovne i umjetne inteligencije
- Pozitivni efekti uvođenja umjetne inteligencije
- Opisati skepsu koja se pojavljuje s najavama novih oblika umjetne inteligencije
- Obraditi etičke dvojbe vezane uz umjetnu inteligenciju
- Analizirati razliku filozofskog, sociološkog, te socijalno-filozofskog pristupa fenomenu
- Opisati pojam superinteligencije
- Iznijeti vlastiti stav o umjetnoj inteligenciji u zaključku rada.

ZADATAK URUČEN	25/9/2019	POTPIS MENTORA	
----------------	-----------	----------------	---

PREDGOVOR

„Pravo pitanje ovdje je, kada ćemo izraditi povelju o pravima umjetne inteligencije? Od čega će se sastojati? I tko će odlučivati o tome?“ – izjavio je Grey Scott, futurist i tehno-filozof, jedan od vodećih svjetskih stručnjaka u polju novih tehnologija.

Ova tri pitanja će biti premisa mojeg znanstvenog rada, upravo zbog važnosti donošenja etičkih i moralnih zakona prije implementacije najvećeg izuma modernog čovječanstva.

Ovim bih se putem htjela zahvaliti, za početak svojem mentoru prof. dr. Seadu Aliću, koji mi je proširio mnoge vidike svojim filozofskim pristupanjem različitim temama.

Također, veliko hvala mojem samohranom ocu, koji mi je omogućio da nastavim školovanje, te mojoj pokojnoj maci koja će mi zauvijek biti veliki uzor i unutarinja snaga koja me potiče na razumnije odluke u životu.

SAŽETAK

Tehnologije umjetne inteligencije rapidno se razvijaju sa mnogo potencijalnih beneficija za ekonomiju, društvo, zajednicu i individualca. Spoznavanjem takvog potencijala, potrebno je iskoristiti te koristi što rasprostranjenije moguće, brže i sa što umjerenijom tranzicijom.

Diljem sektora, UI tehnologije nude obećanja poput rasta produktivnosti i stvaranja novih proizvoda i usluga. Neke od tih tehnologija već se koriste u raznim sektorima, poput prodaje, proizvodnje te zabavne industrije i postoji značajan potencijal za daljnje korištenje u npr. farmaciji, edukaciji i transportu.

Sa takvim potencijalom dolaze i pitanja vezana za utjecaj UI tehnologija na posao, radni život te javne i političke debate o automatizaciji u bližoj budućnosti. Već se naziru indikacije da su takva pitanja stupila u javnu svijest, što je primarno ukoliko postoji mogućnost da se kolektivno razglaba o mogućim pozitivnim i negativnim efektima prije same implementacije umjetne inteligencije i zatim još važnije, superinteligencije.

KLJUČNE RIJEČI: Umjetna inteligencija, Društvo, UI tehnologije, Automatizacija, Implementacija, Superinteligencija

ABSTRACT

Artificial intelligence technologies are rapidly evolving with many potential benefits for the economy, society, community and the individual. Recognizing such potential, it is necessary to exploit these benefits as widely as possible, as quickly as possible and with as moderate transition as possible.

Across the sector, AI technologies offer promises such as productivity growth and the creation of new products and services. Some of these technologies are already being used in various sectors, such as sales, manufacturing and the entertainment industry, and there is considerable potential for further use in pharmacy, education and transportation.

With such potential questions appear regarding the impact of Ai technologies on work, work life, and public and political debates about automatization in the near future. There are already indications that such issues have entered the public awareness, which is elementary if there is a possibility to collectively disclose the possible positive and negative effects before the actual implementation of artificial intelligence and then more importantly, superintelligence.

KEY WORDS: Artificial Intelligence, Society, AI Technologies, Automatization, Implementation, Superintelligence

Sadržaj

UVOD	7
Poslovna inteligencija	9
Implementiranje umjetne inteligencije u različite sektore poslovanja	10
Pozitivni učinci uvođenja UI sustava u poslovanje	11
Zabrinutost pionira umjetne inteligencije	12
ETIČKI PRISTUP U INTERAKCIJI ČOVJEKA SA UI	13
FILOZOFSKO-SOCIOLOŠKI ASPEKT	14
Kako čovjek mijenja predodžbu o sebi kada je suprotstavljen nadmoćnijoj inteligenciji?	14
Snowdenov efekt	14
Snowdenova aplikacija	15
SUPERINTELIGENCIJA	16
Posljedice stvaranja superinteligencije	17
SUPERINTELEKTNO MORALNO MIŠLJENJE	20
Važnost početnih motivacija	21
Treba li razvoj obustaviti ili ubrzati?	22
NEURALINK	23
ZAKLJUČAK	25
LITERATURA	27

UVOD

Umjetna inteligencija (UI, prema engl. akronimu AI, od Artificial Intelligence), predstavlja dio računalne znanosti koja se bavi razvojem sposobnosti računala da obavljaju zadatke za koje je potreban nekakav oblik inteligencije. Posredstvom umjetno stvorene (kodirane) inteligencije računalo posjeduje sposobnost donošenja zaključaka ili rješenja određenih problema, snalaženja u novim situacijama, učenja novih koncepata pomoću kojih mogu razumjeti prirodne jezike i raspoznavati prizore.

Umjetnom inteligencijom također nazivamo svojstva neživog sustava koji iskazuje inteligenciju, to su obično računalni sustavi, dok se isti naziv može primijeniti za robote, koji ne moraju nužno biti inteligentni. Dakle, inteligentnim bi sustavom nazvali sve sustave koji sadrže sljedeće funkcije; uče na temelju iskustava, koriste ogromne količine znanja, komuniciraju sa ljudima prirodnim jezikom i govorom te dopuštaju pogreške i nejasnoće u komunikaciji.[1]

Umjetno inteligentni roboti ili čak sve više i inteligentni humanoidi¹ na pragu su pokretanja četvrte industrijske revolucije i mogli bi zamijeniti polovicu ljudske radne snage do 2015. godine.

Naime, roboti bi mogli preuzeti oko 52% trenutnih poslova u manje od jednog desetljeća, tjerajući ljude da prosude kako pristupiti poslu. Po WEF-u² količina robota koji će raditi umjesto ljudi udvostručiti će se naspram sadašnjih 29%, što bi sugeriralo da će posao izgubiti oko 75 milijuna ljudi i to samo do 2022. Najsnažnije pogođeni automatizacijom biti će industrija hrane i strojovođe.

No, zanimljiva je činjenica da bi takva smjena u dinamici radnog okruženja mogla stvoriti priliku za preko milijun novih radnih mjesta koje bi u ovom slučaju obavljali ljudi. Robotizacija radnih mjesta, umjetna inteligencija i tehnološki napredak mogli bi stvoriti do 133 milijuna potpuno novih poslova.

Prije 30-tak godina nitko nije mogao zamisliti da će im potomci kad odrastu obavljati poslove poput kreiranja aplikacija ili softverskog inženjeringa, zbog toga nitko nije sasvim siguran koji

¹ odnosi se na svako biće čija tjelesna struktura podsjeća na čovječju. Android ili ginoid je humanoidni robot konstruiran tako da izgleda kao da je određenoga spola. Pr. Sofia u Japanu.

² Geneva-based World Economic Forum

će poslovi biti aktualni za 30 godina od sada. Upravo zato, Lee Jong-Wha, direktor Asiatic Research Instituta na Korejskom Sveučilištu, tvrdi da je ključ preživljavanja masovne automatizacije edukacija. Dr. Restrepo tvrdi da je automatizacija agrikulturnog sektora progurala radnike u industrijski sektor i zatim u uslužni sektor kada su industrije postale automatizirane. [2]

Poslovna inteligencija

„Navest ćemo samo da, u najjednostavnijem određenju, poslovna inteligencija predstavlja ranije prikriveno znanje koje se otkriva iz operativnih, rutinskih prikupljenih poslovnih podataka primjenom odgovarajućih računsko-logičkih metoda, obično podržavanih informacijskom tehnologijom.“³

No, poslovna inteligencija ne može biti homogena kategorija budući da sadrži određene podskupine ili komponente inteligencije koje u svojoj cijelosti čine ono što se naziva poslovnom inteligencijom.

Budući da postoji ogroman broj poslovnih podataka koje tvrtke prikupljaju, potrebno ih je razdvojiti u dvije podvrste prema izvoru podataka, a to su:

VANJSKI IZVORI PODATAKA: Podaci koji stižu iz okruženja tvrtke, tj. s tržišta na kojima tvrtka djeluje

UNUTARNJI IZVORI PODATAKA: Podaci koji nastaju ostvarenjem poslovnih procesa unutar tvrtke

Samim razvrstavanjem prema izvoru informacija moguće je razgraničavanje na dvije osnovne podvrste poslovne inteligencije, na **tržišnu** i **unutarnju** inteligenciju tvrtke. [3]

³ Panian, Željko; Klepac, Goran. Poslovna inteligencija. Masmedia, Zagreb, 2003., str. 4-6

Implementiranje umjetne inteligencije u različite sektore poslovanja

UI ima značajan ekonomski potencijal i njezine tehnologije već podupiru nove proizvode diljem širokog spektra poslova i sektora:

- Intelligentni osobni asistenti koji koriste prepoznavanje glasa poput Alexe, Siri i Cortane, česta su pojava u poslovanju
- U sektoru prometa, UI procesi podupiru razvoj autonomnih vozila i pomažu u upravljanju prometnim tokovima te dizajnu transportnih sustava
- U obrazovanju, UI tehnologije podržavaju personalizirane sustave učenja.
- U zdravstvu, UI omogućuje nove dijagnostičke alate i alate za podršku kod odlučivanja medicinskih stručnjaka
- U maloprodaji i logistici UI pokušava dizajnom skladišnih objekata poboljšati učinkovitost.
- U razvojnoj i humanitarnoj pomoći, analitika podataka koju omogućuje UI pomaže u pružanju ciljeva održivog razvoja i procjeni humanitarnih scenarija
- U kreativnim industrijama programeri stvaraju računalne sustave koji mogu izrađivati jednostavne vijesti, na primjer o poslovnim rezultatima, sastaviti orkestralnu glazbu i generirati kratke filmove
- U svim sektorima, UI se koristi za analizu velikih količina podataka, s razlogom poboljšanja poslovnih procesa ili dizajniranja novih usluga. [4]

Pozitivni učinci uvođenja UI sustava u poslovanje

Tijekom telefonskih razgovora klijenti nerijetko osjećaju da ih nitko ne sluša. Nažalost sve češće ispada da su u pravu. No, ukoliko računalo možemo nazvati imenom onda i nisu posve u pravu.

Google je nedavno najavio Duplex, UI asistenta koji može zaprimati zahtjeve kupaca, kao npr. zakazivanje sastanaka ili pružanje osnovnih informacija. Doduše, automatske sekretarice postoje već godinama, no stvar je u Duplexu da će biti teško primijetiti razliku između čovjeka ili robota.

Duplexovo najavljivanje ubrzo je pokrenulo pitanja o transparentnosti. Bi li kompanije trebale uključiti obavijest u kojoj je rečeno da su „objekti na telefonu malo manje ljudi nego što se čine“? -Normalno da bi trebali, zato će Duplex to i učiniti, budući da postoje beneficije od neiskrenosti prema klijentima u ovom slučaju.

U obzir se trebaju uzeti moguće prednosti koje bi nastale u interakciji sa robotom:

- Zadovoljstvo korisnika može se povećati – Primjerice, ukoliko bi korisnici bili svjesni da razgovaraju sa robotom, te ne troše ničije vrijeme, niti sjedećeg korisnika, niti robota, potrošili bi više vremena modificirajući njihov odabir.
- Korisnici mogu biti potpunije informirani – U govoru sa drugim ljudima, ljudi često ne žele postavljati ista ili dodatna pitanja da ih se ne bi percipiralo neobrazovanim ili dosadnim, budući da je čovjeku jako teško priznati da ne razumije nešto što mu je već rečeno i objašnjeno. No, u interakciji sa robotima taj ih problem ne muči, te će većina iskrenije voditi razgovor sa neživim bićem. [5]

Zabrinutost pionira umjetne inteligencije

U siječnju, britansko-američki računalni znanstvenik Stuart Russell izradio je nacrt i postao prvi potpisnik otvorenog pisma pozivajući istraživače da pogledaju dalje od cilja da umjetna inteligencija postane samo snažnija.

„Preporučujemo proširena istraživanja koja imaju za cilj osigurati da sve sposobniji UI sustavi budu robusni i korisni“- navodi se u pismu. „Naši sustavi UI moraju učiniti ono što mi želimo“. Od tada, tisuće ljudi potpisalo je pismo, uključujući vodeće UI istraživače na Facebooku, Googleu, Microsoftu i drugim industrijskim čvorištima zajedno s vrhunskim računalnim znanstvenicima, fizičarima i filozofima diljem svijeta. Do kraja ožujka, oko 300 istraživačkih skupina podnijelo je zahtjev za novim istraživanjima o „održavanju umjetne inteligencije korisnom“ sa sredstvima koja im je osiguralo i sam potpisnik, Elon Musk.

Russel, profesor informatike i osnivač Centra za inteligentne sustave na Sveučilištu u Kaliforniji, Berkeleyu, već dugo razmišlja o moći i opasnostima strojeva za razmišljanje. Autor je izdao više od 200 radova, kao i standardni stručni udžbenik u polju, *Umjetna inteligencija: suvremeni pristup* zajedno sa Petrom Norvigom, voditeljem istraživanja na Googleu. No, sve rapidniji napredak u polju UI dao je Russelovoj dugogodišnjoj zabrinutosti pojačanu urgentnost.

Nedavno, kaže Russell, UI je postigla odlične rezultate na temelju neuro-inspiriranih algoritama učenja. Oni se koriste u Facebookovom softveru za prepoznavanje lica, osobnim asistentima za pametne telefone i Googleovim autonomnim automobilima. Također, ogroman uspjeh ostvarila je simulirana mreža umjetnih neurona koja naučila je igrati Atari video igre bolje od ljudi za svega nekoliko sati i to sve bez unaprijed programiranih algoritama; „Da je vaše novorođenče učinilo nešto takvo, mislili biste da je opsjednuto“ – odgovorio je Russell. [6]

ETIČKI PRISTUP U INTERAKCIJI ČOVJEKA SA UI

Bihevioralni znanstvenici smatraju da bi polazna točka u osiguravanju etičkog ponašanja UI bili principi objavljeni u Belmont⁴-ovom izvješću 1978. godine.

Principi uključuju tri ključna zahtjeva:

1. Ljudska autonomija ne smije biti ugrožena (treba im se osigurati slobodna volja)
2. Koristi implementiranja takve tehnologije trebale bi nadmašiti rizike
3. Koristi i rizici trebali bi biti razmjerno raspoređeni među korisnicima (ljudi ne smiju biti diskriminirani po osnovi rase, roda ili religije) [7]

⁴ Izvještaj Belmont napisala je Nacionalna komisija za zaštitu ljudskih subjekata biomedicinskih i bihevioralnih istraživanja. Komisija, stvorena kao rezultat Nacionalnog zakona o istraživanjima iz 1974. godine, zadužena je za utvrđivanje osnovnih etičkih principa na kojima bi se trebali temeljiti pri provođenju biomedicinskih i bihevioralnih istraživanja koji uključuju ljudske subjekte te izradu smjernica koje bi osigurale da se takva istraživanja provode u skladu s istim principima

FILOZOFSKO-SOCIOLOŠKI ASPEKT

Kako čovjek mijenja predodžbu o sebi kada je suprotstavljen nadmoćnijoj inteligenciji?

Uvođenje AI može imati značajan utjecaj na ljudsku sigurnost, slobodu i odluke. Kao primjer može poslužiti računalno nadziranje od kojeg ljudi strahuju dugi niz godina, pogotovo nakon Edwarda Snowdena koji je osvijestio dobar dio populacije.

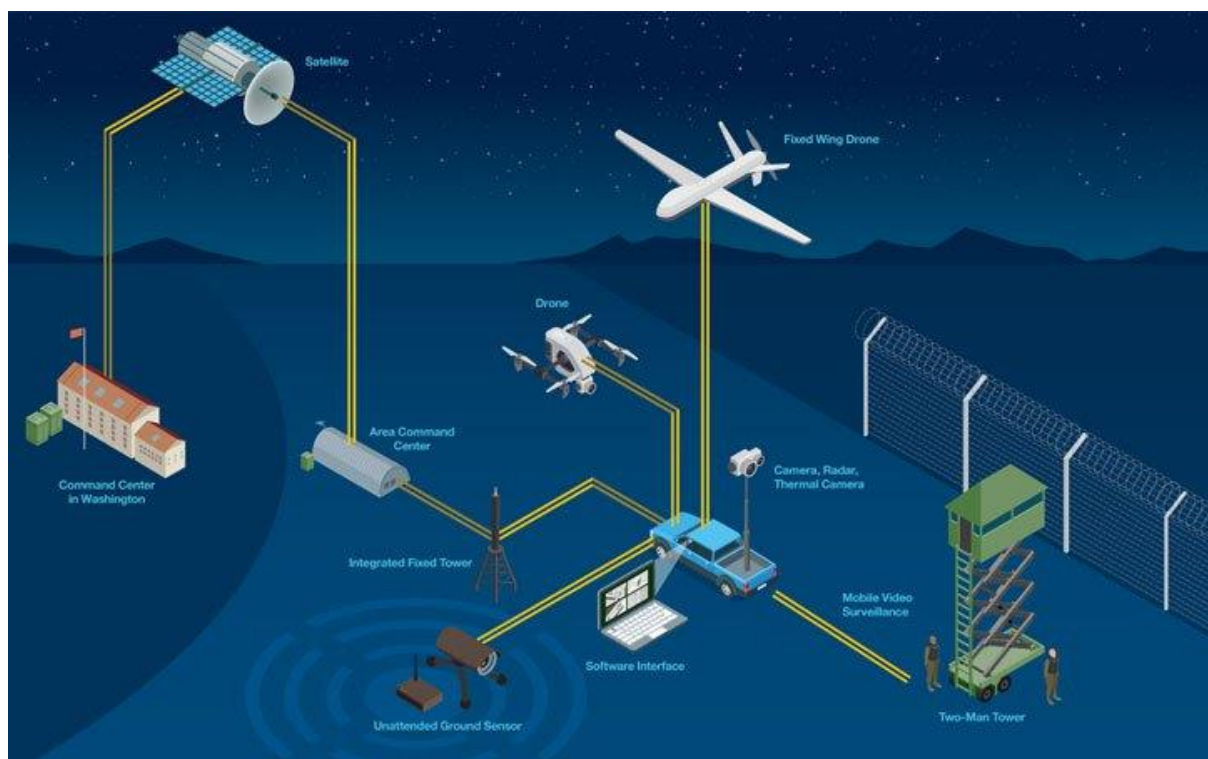
Naime, uz pomoć strojnog vida, sve slike trenutno snimljene pametnim telefonima, kamerama na ulicama, u bankama mogu se automatski analizirati kako bi se traženoj osobi ušlo u trag. No, takva konstantna kontrola može prouzročiti nedostatak privatnosti. [8]

Snowdenov efekt

U tehnološkoj industriji, Snowdenov efekt imao je ogroman utjecaj nakon što je otkriveno da NSA sakuplja informacije koje posjeduju neki od američkih uslužnih servisa u Oblaku. Google, Cisco i AT&T izgubili su integritet na međunarodnoj razini zbog svoje uloge u špijuniranju NSA-e. U Istraživanju koje je provela *Information Technology and Inovation Foundation* objavljeno u kolovozu 2013. godine, procijenjeno je da bi računalna industrija temeljena na Oblak servisima do 2016. godine mogla izgubiti do 35 bilijuna \$.

Nakon otkrića da se prisluškivao mobitel njemačke kancelarke Angele Merkel, tehnološke industrije požurile su stvoriti siguran mobitel kojeg bi bilo u potpunosti nemoguće špijunirati. Blackphone kojeg je The New Yorker nazvao „telefonom za Snowden doba“ opisan je kao pametni telefon izričito dizajniran za sigurnost i privatnost, sa mogućnošću šifriranja telefonskih poziva, SMS-ova i pretraživanja internetskih preglednika. No, i danas se svakih nekoliko mjeseci može pročitati nova kontroverza o krađi privatnih informacija građana putem pametnih telefona. Od otkrića je prošlo nekoliko godina te se od tada mogu promatrati njegovi utjecaji, poput povećanog interesa za enkripciju, propadanja mnogih američkih tvrtki koje nisu legalno poslovale do preispitivanja sigurnosti spremanja datoteka na Oblak. Istraživanje iz travnja 2014. godine pokazalo su da netom Snowdenovog objavljivanja, Amerikanci počeli sve

rjeđe koristiti Internet za prosljeđivanje e-pošte, kupnje putem njega te bankarskog poslovanja. [9]



Slika 1 Sustav nadziranja i način na koji se povezuje da bi stigao do krajnjeg korisnika ili mete promatranja

Snowdenova aplikacija

Upravo kako bi spriječio napade hakera i nadziranje državnih agencija nad privatnim podacima pojedinca, Snowden je prošle godine lansirao aplikaciju Haven. Ona funkcionira na principu senzora koji prate je li bilo pokreta, promjena u osvjetljenju ili pomicanja za koje bi vlasnik uređaja trebao znati. Naime, dogodi li se da osoba ostavi laptop u uredu, kod kuće ili u hotelskoj sobi, na njega je moguće postaviti mobilni uređaj sa instaliranom aplikacijom koji onda služi kao svojevrsan senzor za prepoznavanje pokreta. Nakon što senzor detektira bilo koju promjenu, trenutno šalje vlasniku obavijest o mogućem „napadu“ na njegov uređaj. Trenutno je aplikaciju moguće skinuti samo u slučaju da telefon posjeduje Androidov operativni sustav, no najavljuje se i verzija za iOS-e. [10]

SUPERINTELIGENCIJA

Superinteligencijom možemo smatrati oblik intelekta koji može nadjačati funkcije ljudskog mozga u svakom području. Iz čega se da konstatirati da Deep Blue⁵ u stvari nije superračunalo budući da se njegov intelekt koristi samo za jednu vještinu koja je još uvijek mjerljiva ljudima. Iako sposoban izvoditi brojne intelektualne podvige koje nijedan pojedinac nije sposoban, nije dovoljno integriran da bi se mogao smatrati „intelektom“ upravo jer postoje mnoga područja u kojima podbacuje naspram ljudi.

To otvara pitanje na koji bi se način superinteligencija mogla implementirati. Hoće li se nalaziti unutar digitalnog kompjutera, skupa umreženih računala, uzgojenog kortikalnog tkiva ili nečeg trećeg?

Nažalost, prostorna ograničenja sprečavaju mnoge znanstvenike i UI stručnjake da obuhvatno osmisle plan za implementiranje UI-a. Trenutno nekoliko odabranih ideja u obliku raznih crteža, najviše je što mogu izvesti sa postojećim znanjem.

Nekoliko se znanstvenika prepiralo oko razdoblja u kojem bi se ovaj izum mogao ostvariti, tj. činjenice da bi to razdoblje trebalo doći za par desetljeća, moguće kao rezultat rastućih hardware performansi i pojačane sposobnosti da se implementiraju algoritmi i arhitektura slična onoj u ljudskom mozgu.

Moglo bi se ispostaviti da će biti potrebno puno više vremena, no trenutno je vjerojatnost zanemariva da će superinteligencija biti stvorena unutar životnog vijeka nekih današnjih ljudi. S obzirom na ogromne posljedice superinteligencije, imalo bi smisla dati ovoj perspektivi ozbiljno razmatranje, čak i ako postoji mala vjerojatnost da će se to dogoditi u neko skorije vrijeme. [11]

⁵ Deep Blue je šahovski kompjuter razvijen od IBM-a. Godine 1997. taj je stroj porazio svjetskog šahovskog prvaka Garija Kasparova.

Posljedice stvaranja superinteligencije

Preduvjet za smislenu raspravu o superinteligenciji spoznaja je da superinteligencija nije još samo jedna tehnologija, još jedan alat koji će postupno povećavati ljudske sposobnosti. Superinteligencija je radikalno drukčija.

Vrlo je vjerojatno da će svaka tehnologija koje trenutno možemo predvidjeti biti brzo razvijena u prvoj superinteligenciji, zajedno s mnogim drugim tehnologijama za koje ljudi nisu ni čuli, niti pomišljali na njih.

Neke od budućih tehnologija mogu se predvidjeti ili su već u procesu realizacije:

- Kontrola ljudskog raspoloženja, emocija i motivacije
- Vrlo moćna računala – Superračunala
- Napredno oružje, vjerojatno sposobno razoružati nuklearnu energiju
- Svemirska putovanja i von Neumanove sonde⁶
- Eliminacija bolesti i prestanak starenja
- Reanimacija bolesnika
- Učitavanje – neuronsko ili sub-neuronsko skeniranje mozga i implementacija istih algoritamskih struktura na računalu na način koji održava memoriju i osobnost
- Potpuna realna virtualna stvarnost [12]

- **Superinteligencija će dovesti do još naprednije superinteligencije**

To rezultira i iz poboljšanog hardvera koji superinteligencija može stvoriti, kao i iz poboljšanja koje može napraviti u vlastitom izvornom kodu. No, i takvo samostalno ažuriranje trebalo bi biti podlegnuto kontroli barem načinom da se ne omoguće dovoljna sredstva za daljnje ažuriranje dok se ne osigura da iza njega ne stoji maliciozni razlog.

- **Umjetne misli mogu se lako kopirati**

Budući da su umjetne inteligencije softver, mogu se lako i brzo kopirati, sve dok postoji hardver dostupan za pohranu. Osim hardvera, granični trošak izrade dodatne kopije učitavanja ili UI nakon izgradnje gotovo je besplatan. Umjetni umovi stoga bi mogli brzo postojati u velikom

⁶ Samo-reproducirajuće međuzvezdane sonde

broju, iako moguće da bi učinkovitost bila usmjerena na koncentriranje računalnih resursa u jednom superintelektu.

- **Pojava superinteligencije može biti iznenadna**

Čini se da je mnogo teže doći do mjesta gdje smo sada na umjetnoj inteligenciji na razini čovjeka, nego odatle doći do superinteligencije. Iako može potrajati neko vrijeme prije nego li se izumi superinteligencija, završna se faza može dogoditi ubrzo. To znači da se prelazi iz stanja gdje je umjetna inteligencija na razini čovjekove, u stanje gdje postoji potpuna superinteligencija, s revolucionarnim primjenama. Takva se mogućnost iznenadne pojave superinteligencije naziva **hipoteza singularnosti**.

- **Umjetni intelektualci potencijalno su autonomni agenti**

Superinteligenciju ne treba nužno zamišljati kao puki alat. Specijalizirane superinteligencije, koje mogu misliti samo o ograničenom skupu problema mogu biti izvedive, dok bi opća superinteligencija bila sposobna za samostalnu inicijativu i izradu vlastitih planova, te se stoga može prikladnije smatrati autonomnim agentom.

- **Umjetni umovi ne moraju imati motive slične čovjeku**

Ljudi su rijetko voljni robovi, ali ne postoji ništa nevjerojatno u ideji superinteligencije koja ima za svoju nadmoć da služi čovječanstvu ili nekom određenom čovjeku, bez ikakve želje da se pobuni ili da „oslobodi“ sebe. Također se čini savršeno moguće imati superinteligenciju, čiji je jedini cilj nešto sasvim proizvoljno, kao što je izrada što više spajalica. Umjetni intelektualci dakle, ne moraju dijeliti naše ljudske motivacijske tendencije.

- **Umjetni intelektualci možda nemaju ljudsku psihu**

Kognitivna arhitektura umjetnog intelekta može biti sasvim različita od one kod ljudi. Umjetnim intelektom može se lako zaštititi od nekih vrsta ljudske pogreške i pristranosti, dok je istodobno izložen povećanom riziku od drugih vrsta pogrešaka koje ni najsretniji čovjek ne bi učinio. Subjektivno, unutarnji svjesni život umjetnog intelekta, ako ga ima, može biti sasvim različit od našeg.

- **Superinteligencija može biti posljednji izum koji ljudi trebaju napraviti**

Uzimajući u obzir intelektualnu superiornost superinteligencije, vrlo je vjerojatno da će biti bolja u poljima znanstvenih istraživanja i tehnološkom razvoju od bilo kojeg čovjeka, a lako moguće i od svih ljudi zajedno. [13]

Zbog svih tih razloga, treba biti oprezan pri pretpostavci da se pojava superinteligencije može predvidjeti ekstrapoliranjem povijesti drugih tehnoloških otkrića ili da će priroda i ponašanje umjetnih intelektualaca nužno nalikovati ljudskim ili drugim životinjskim umovima.[14][15]

SUPERINTELIGENTNO MORALNO MIŠLJENJE

Kada dolazimo do pitanja etike ili moralnog rasuđivanja, superinteligencija bi nam mogla pružiti potpunije i točnije odgovore koji se mogu postići zaključivanjem i vrednovanjem dokaza. Isto vrijedi kod pitanja politike i dugoročnog planiranja. Kada se radi o razumijevanju koje bi politike dovele do rezultata, i s kojim sredstvima bi se došlo do postizanja ciljeva na efektivniji način, vjerojatno bi i tu nadjačala ljude. Stoga će postojati mnoga pitanja uz koja ćemo dobiti servirane odgovore ukoliko se stvori superinteligencija i ukoliko su isti adekvatni, moći ćemo preći na zadnju i najriskantniju fazu prepuštanja odluka. Također, ako smo nesigurni kako procijeniti moguće ishode, mogli bismo zatražiti superinteligenciju da procijeni kako bismo ocijenili ishode ukoliko bismo o njima razmišljali jako dugo, pažljivo promišljali, imali više pamćenja, i tako dalje. Pri oblikovanju cilja superinteligencije ne bi uvijek bilo potrebno dati detaljnu, eksplicitnu definiciju tog cilja. Mogli bismo angažirati superinteligenciju da nam pomogne odrediti stvarnu namjeru naših zahtjeva, čime se smanjuje rizik da bi nevjerovatne formulacije ili zbunjenost oko onoga što želimo postići dovele do rezultata koje ne bismo odobravali u retrospektivi.[11]

Važnost početnih motivacija

Mogućnost delegiranja mnogih odluka superinteligenciji, ne znači da možemo biti zadovoljni načinom izgradnje superinteligencije. Naprotiv, postavljanje početnih motivacija, a posebice izbor cilja najviše razine za superinteligenciju, od najveće je važnosti. Riskiramo budućnost cijele populacije ukoliko ne riješimo taj problem u začetku.

Zbog svoje superiorne sposobnosti planiranja i tehnologija koje se mogu razviti, moguće je pretpostaviti da bi i prva superinteligencija bila vrlo moćna. Sasvim je moguće da bi bilo bez premca donijeti gotovo svaki mogući ishod i spriječiti svaki pokušaj da se spriječi provedba njegovog glavnog cilja. Mogla bi „ubiti“ sve ostale „agente“, uvjeriti ih da promijene svoje ponašanje ili blokirati njihove pokušaje uplitanja. Čak i "sputana superinteligencija" koja se izvodi na izoliranom računalu, sposobna za interakciju s ostatkom svijeta samo putem tekstualnog sučelja, mogla bi biti sposobna „provaliti iz zatvora“ nagovarajući svoje rukovoditelje da je „puste“. Postoje čak i preliminarni eksperimentalni dokazi da bi to bio slučaj.

Čini se da je najbolji način da se osigura blagotvoran utjecaj superinteligencije na svijet da ga se obdari filantropskim vrijednostima. Njezin glavni cilj trebao bi biti prijateljstvo. Kako treba razumjeti prijateljstvo i kako ga primijeniti, te kako se prijateljstvo treba raspodijeliti između različitih ljudi i umjetnih stvorenja, pitanje je koje zaslužuje daljnje razmatranje. Barem svi ljudi, a vjerojatno i mnoga druga živa bića na zemlji trebali bi dobiti značajan udio beneficija superinteligencije. Ako su koristi koje bi superinteligencija mogla pružiti ogromne, onda je možda nevažno cjenkati se oko detaljnog rasporeda distribucije i važnije tražiti da se osigura da svatko dobije udio, jer po toj pretpostavci, čak i mali dio bio bi dovoljan da osigura dug i vrlo dobar život. Jedini rizik od kojeg se mora paziti jest da ga oni koji razvijaju superinteligenciju vjerojatno je neće razviti u potpunosti filantropski, već će joj umjesto toga dati ograničeniji cilj koji će služiti samo nekoj manjoj skupini, kao što su njezini stvaratelji ili oni koji su je naručili.

Međutim, ukoliko superinteligencija krene sa prijateljskim ciljem, onda se možemo osloniti na to da će takva i ostati, jedino ako se ne pokuša sama lišiti te kvalitete. Ova točka je elementarna.

Kod ljudi, sa njihovim kompliciranim mentalnim sklopom, u kojem se prepliću ideje, nagoni i želje, teško je identificirati pravi cilj postojanja, dok sa superinteligencijom nema takvog slučaja

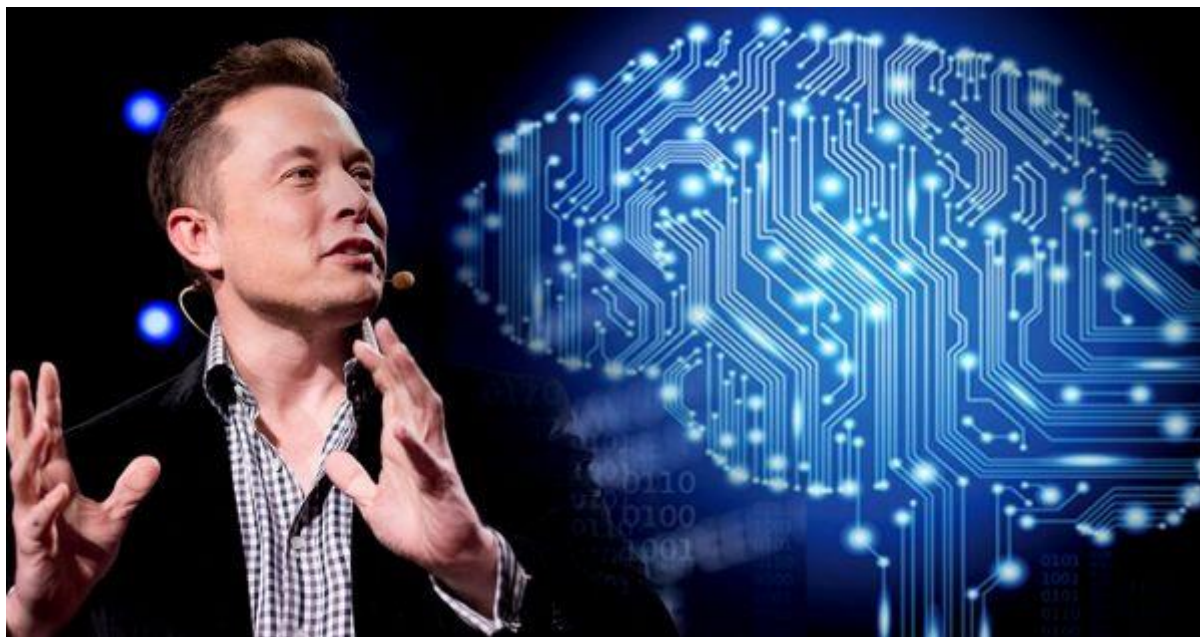
jer ukoliko je stvorena sa definiranim ciljem postojanja, može se aplicirati na argument dat u ulomku iznad. U tom slučaju postoji sve više razloga da je se i implementira sa takvom eksplicitnom motivacijskom strukturom.[11][14]

Treba li razvoj obustaviti ili ubrzati?

Teško je smisliti bilo koji problem koji superinteligencija ne bi mogla riješiti ili barem pomoći u rješavanju. Bolesti, siromaštvo, uništavanje okoliša, problemi su koje bi superinteligencija uz pomoć nanotehnologije mogla otkloniti. Osim toga, superinteligencija mogla bi ljudima dati neograničen vijek trajanja, bilo zaustavljanjem ili invertiranjem procesa starenja korištenjem nanomedicine ili pak ponudom mogućnosti da se sami učitamo. Superinteligencija bi također mogla stvoriti prilike za nas da uvelike povećamo naše intelektualne i emocionalne sposobnosti, i to bi nam moglo pomoći u stvaranju vrlo privlačnog iskustvenog svijeta u kojem bismo mogli živjeti živote posvećene u radosnom igranju igara, povezujući jedne s drugima, doživljavajući, osobni rast i život bliže našim idealima.

Jednom kada bi postojala, superinteligencija bi nam mogla pomoći smanjiti ili eliminirati druge egzistencijalne rizike, kao što je rizik da će ljudi u ratu ili terorizmu koristiti naprednu nanotehnologiju, što predstavlja ozbiljnu prijetnju dugoročnom opstanku inteligentnog života na zemlji. Ako što prije dođemo do superinteligencije, možemo izbjeći taj rizik nanotehnologije i mnogih drugih. Ako, s druge strane, prvo dobijemo nanotehnologiju, morat ćemo se suočiti s rizicima nanotehnologije i ako ti rizici prežive, kreću rizici od superinteligencije. Čini se da je ukupni rizik minimiziran uvođenjem superinteligencije, s velikom pažnjom, što je prije moguće.[16]

NEURALINK



Slika 2 Elon Musk (Neuralink)

Neuralink Corporation američka je tvrtka za neurotehnologiju koju su osnovali Elon Musk i drugi, razvijajući implantacijska sučelja mozga i računala. Sjedište tvrtke nalazi se u San Franciscu i započelo je sa radom 2016. godine i prvi put je javno prijavljeno u ožujku 2017. godine. Od svog osnivanja, tvrtka je angažirala nekoliko istaknutih neuroznanstvenika s raznih sveučilišta. Do srpnja 2019. godine primila je 158 milijuna dolara (od čega 100 milijuna dolara samo od Muska) i zaposlila je 90 zaposlenika. U to vrijeme, Neuralink je objavio da radi na uređaju sličnom šivaćem stroju sposobnom za uvođenje vrlo tankih niti u mozak, pokazao je sustav koji čita informacije iz laboratorijskog štakora putem 1500 elektroda što se pokazalo uspješnim te se najavljuju prvi eksperimenti s ljudima u 2020. godini.

Neuralink se temelji na vrlo znanstvenom konceptu: startup želi povezati vaš mozak izravno s računalima i drugim elektroničkim uređajima, po mogućnosti putem kibernetičkih implantata koji omogućuju vašem umu da se poveže s gadgetima i programima.

Musk vjeruje da su sučelja od mozga do stroja važan dio budućnosti čovječanstva - po mogućnosti bez distopijskih oznaka. Neuralink želi pronaći najbolje moguće načine kako bi taj um mogao biti povezan. Na djelu je i paranoidniji faktor: Musk i drugi zabrinuti su zbog utjecaja

UI-a u svijetu, osobito ako se UI okreće protiv ljudi. Musk bi više volio da imaju izravniji način kontrole UI.[17]

ZAKLJUČAK

Trenutnom implementacijom UI u poslovanje omogućeno nam je za sad tek racionalno otkrivanje problema i logičko rješavanje istih, no što kada dođemo do trenutka kada će roboti ili bilo koji oblik UI, da li u humaniziranom ili strojnom obliku, početi odlučivati npr. o kolateralnim žrtvama u ratnim okruženjima ili prednosti u transplantacijama organa u medicinskim oblastima. Hoće li u tom slučaju nestati ona humana pogreška da se rasuđuje o prednosti življenja po klasnim razlikama ili će upravo manjkavost moralnih načela izazvati pozitivniju situaciju gdje će UI rasuđivati, npr. hoće li se čovjeku transplantirati potrební organ ukoliko ima viši postotak kompatibilnosti sa organom ili veće šanse preživljavanja.

Takve važne odluke, prepuste li se UI, mogu izazvati sveopći raskol između stavova i odabira robota i naših duboko ukorijenjenih etičkih načela koje ne volimo zagovarati u normalnim okruženjima, no kada situacija postane sve osobnija, mogu li ljudi prevladati onaj instinktivni obrambeni sustav da smatraju da je novac solucija za sve probleme. Upravo će to stvarati najveću distinkciju između nas i „njih“ jer „oni“ neće vidjeti novac kao sredstvo izmjene dobara nego kao neki jako dobro zamaskirani socijalni konstrukt koji zavarava onih 99% stanovništva, dok onih 1% vrlo dobro razumije da će intelektualno vlasništvo i znanje postati sveopća valuta u univerzalnim razmjerima. U tom slučaju UI ili superinteligencija bili bi „milijarderí“ i ukoliko će ikada razviti svijet o samima sebi.

Što ukoliko stroj pogrešno rasudi situaciju i odabere krivo rješenje, potpuno suprotno od čovjeka u istoj situaciji, da li će se u tom slučaju „suditi“ stroju ili čovjeku koji je napisao kod zbog kojeg je stroj krivo odlučio, ili će se samo prisilno ugasiti stroj? Hoće li se to smatrati ubojstvom? Ovakve su situacije veoma izgledne i to u što skorijoj budućnosti i bit će u potpunosti nerazrješive ne raspišu li se zakonodavne mjere kojima će se prije same implementacije moći predvidjeti svi mogući ishodi te zatim postupci u kojima bi i jedna i druga vrsta trebale imati prava kojim će se moći služiti.. U suprotnom slučaju mogu se očekivati zasnivanja sindikata od strane strojeva, a u najgorem slučaju sveopća pobuna te vrlo vjerojatno scenarij sličan kao i iz Swartzeneger-ovih blockbuster-a od čega valjda svaka osoba na planetu strahuje.

Situacija u kojoj stroj pogriješi u potpunosti se razlikuje od one u kojoj čovjek pogriješi, ipak čovjek je nesavršen i inferioran naspram UI. Od UI će se očekivati preciznost i nepogrješivost. No hoće li UI znati ukoliko je krivo odlučio? Odgovor je ne, upravo jer stroj neće očitavati

grižnju savjesti nakon lošeg odabira, jer će po njegovim kalkulacijama takvo rješenje biti najbolji ishod neke moguće situacije.

Dakle, krajnji zaključak nije pitanje treba li UI ili superinteligenciju implementirati u poslovanje, nego da se prilikom implementiranja ozakoni i priušti najsigurniji način rada u simbiozi sa ljudima i međusobnim prihvaćanjem jedne i druge strane, što se zasada čini većim problemom nego stvaranje same UI.

Sveučilište
Sjever

VŽKC

MMI

SVEUČILIŠTE
SJEVERIZJAVA O AUTORSTVU
I
SUGLASNOST ZA JAVNU OBJAVU

Završni/diplomski rad isključivo je autorsko djelo studenta koji je isti izradio te student odgovara za istinitost, izvornost i ispravnost teksta rada. U radu se ne smiju koristiti dijelovi tuđih radova (knjiga, članaka, doktorskih disertacija, magistarskih radova, izvora s interneta, i drugih izvora) bez navođenja izvora i autora navedenih radova. Svi dijelovi tuđih radova moraju biti pravilno navedeni i citirani. Dijelovi tuđih radova koji nisu pravilno citirani, smatraju se plagijatom, odnosno nezakonitim prisvajanjem tuđeg znanstvenog ili stručnoga rada. Sukladno navedenom studenti su dužni potpisati izjavu o autorstvu rada.

Ja, GLORIA STRMEČKI (ime i prezime) pod punom moralnom, materijalnom i kaznenom odgovornošću, izjavljujem da sam isključivi autor/ica završnog/diplomskog (obrisati nepotrebno) rada pod naslovom ETICKI I SOCIJALNI ASPEKT U (upisati naslov) te da u navedenom radu nisu na nedozvoljeni način (bez pravilnog citiranja) korišteni dijelovi tuđih radova.

Student/ica:

(upisati ime i prezime)

(vlastoručni potpis)

Sukladno Zakonu o znanstvenoj djelatnosti i visokom obrazovanju završne/diplomske radove sveučilišta su dužna trajno objaviti na javnoj internetskoj bazi sveučilišne knjižnice u sastavu sveučilišta te kopirati u javnu internetsku bazu završnih/diplomskih radova Nacionalne i sveučilišne knjižnice. Završni radovi istovrsnih umjetničkih studija koji se realiziraju kroz umjetnička ostvarenja objavljuju se na odgovarajući način.

Ja, GLORIA STRMEČKI (ime i prezime) neopozivo izjavljujem da sam suglasan/na s javnom objavom završnog/diplomskog (obrisati nepotrebno) rada pod naslovom ETICKI I SOCIJALNI ASPEKT U (upisati naslov) čiji sam autor/ica.

Student/ica:

(upisati ime i prezime)

(vlastoručni potpis)

LITERATURA

Knjige:

- [3] Panian, Ž., Klepac, G.: „*Poslovna inteligencija*“, MASMEDIA, Zagreb, 2003.
- [4] The British Academy; , *The impact of artificial intelligence on work*, The Royal Society, London, 2018.
- [8] Trapp, R.: „*Impacts of Artificial Intelligence*“, Austrian Research Institute for Artificial Intelligence, Beč, 1985.
- [12] Moravec, H. Robot: „*Mere Machine to Transcendent Mind*“, Oxford University Press: New York, 1999.
- [13] Kurzweil, R.: „*The Age of Spiritual Machines: When Computers Exceed Human Intelligence*“ Viking: New York, 1999.
- [14] Vinge, V. :, „*The Coming Technological Singularity.*“, Whole Earth Review, Winter issue.,1993.
- [15] Drexler, K. E.; „*Engines of Creation: The Coming Era of Nanotechnology*“,Anchor Books: New York, 1986.
- [16] Freitas Jr., R. A. *Nanomedicine, Volume 1: Basic Capabilities*, Landes Bioscience: Georgetown, TX, 1999.

Internetski izvori:

- [1] <http://www.enciklopedija.hr/natuknica.aspx?ID=63150>, dostupno 01.09.2019.
- [2] <https://www.express.co.uk/news/science/1018786/Artificial-intelligence-AI-robots-replace-humans-machines-take-over-jobs>, dostupno 12.04.2019.
- [5] <https://sloanreview.mit.edu/article/rethink-ai-objectives/>, 08.04..2019.

- [6] <https://www.quantamagazine.org/artificial-intelligence-aligned-with-human-values-qa-with-stuart-russell-20150421> , dostupno 21.07.2019.
- [8] <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>, dostupno 28.08.2019.
- [9] <http://skr.rs/hME> , dostupno 01.09.2019.
- [10] <https://guardianproject.github.io/haven/> ,dostupno 01.09.2019.
- [17] <https://www.digitaltrends.com/cool-tech/neuralink-elon-musk/> , dostupno 27.08.2019.

Članci:

- [11] Bostrom, N., *How long before superintelligence?*, Int. Jour. of Future Studies, 1998, vol. 2

Slike:

- [1] <http://skr.rs/hMN> , dostupno 24.08.2019.
- [2] <http://skr.rs/hkG> , dostupno 01.09.2019.