

Umjetna inteligencija i društvo 21. stoljeća

Bošnjak, Mia

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences / Sveučilište Josipa Jurja Strossmayera u Osijeku, Filozofski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:142:840697>

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

Download date / Datum preuzimanja: **2024-12-22**



FILOZOFSKI FAKULTET
SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

Repository / Repozitorij:

[FFOS-repository - Repository of the Faculty of Humanities and Social Sciences Osijek](#)



Sveučilište J.J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Sveučilišni diplomski dvopredmetni studij Nakladništvo i Informatologija

Mia Bošnjak

Umjetna inteligencija i društvo 21. stoljeća

Diplomski rad

Mentorica: doc. dr. sc. Milijana Mičunović

Osijek, 2024.godine

Sveučilište J.J. Strossmayera u Osijeku
Filozofski fakultet Osijek
Odsjek za informacijske znanosti
Sveučilišni diplomski dvopredmetni studij Nakladništvo i Informatologija

Mia Bošnjak

Umjetna inteligencija i društvo 21. stoljeća

Diplomski rad

Društvene znanosti, Informacijske i komunikacijske znanosti, Informacijski
sustavi i informatologija

Mentorica: doc. dr. sc. Milijana Mičunović

Osijek, 2024.godine

IZJAVA

Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napravila te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citat s napisanim izvorom odakle su preneseni.

Svojim vlastoručnim potpisom potvrđujem da sam suglasna da Filozofski fakultet Osijek trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta Osijek, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

Mia Bošnjak , 3001000355035
Ime i prezime studenta i JMBAG

Sadržaj

Sažetak

1.Uvod.....	1
2.Kratka povijest umjetne inteligencije.....	3
2.1. Definicija umjetne inteligencije.....	3
2.2. Povijesni razvoj osnovnih pojmova i koncepata umjetne inteligencije kroz književnost i znanost.....	5
2.3. Razvoj discipline umjetne inteligencije.....	7
2.4.Početak odnosa umjetne inteligencije i društva.....	11
3.Umjetna inteligencija i društvo 21. stoljeća.....	13
3.1. Trendovi u razvoju umjetne inteligencije.....	14
3.1.1. Algoritmi umjetne inteligencije i pametni sustavi.....	15
3.1.2. Generativna umjetna inteligencija.....	16
3.1.3. <i>Deep fake</i> tehnologija.....	18
3.1.4. Umjetna inteligencija i robotika.....	20
3.2. Primjena umjetne inteligencije.....	22
3.2.1. Umjetna inteligencija, automatizacija i tržište rada.....	23
3.2.2.Umjetna inteligencija, urbanizacija i promet.....	26
3.2.3. Umjetna inteligencija i kreativni rad.....	28
3.3. Izazovi u razvoju umjetne inteligencije.....	29
3.3.1. Društveno odgovorni razvoj i primjena umjetne inteligencije.....	32
3.4. Pismenost u području umjetne inteligencije.....	35
3.5. Budućnost razvoja i primjene umjetne inteligencije.....	36
4.Zaključak.....	39
5.Literatura.....	41

Sažetak

Sadržaj rada bavi se pitanjem tehnologije umjetne inteligencije (UI) i njezinim utjecajem na društvo 21. stoljeća. Cilj je rada opisati tehnologiju UI, ukratko prikazati njezin razvoj te opisati njezinu trenutnu ulogu i utjecaj na društvo 21. stoljeća. Svrha rada je pružiti dublji uvid u trendove te ukazati na prednosti i mogućnosti, odnosno na izazove, rizike i opasnosti tehnologije UI za društvo 21. stoljeća. Poseban dio rada posvećen je konkretnoj primjeni UI i etičkim aspektima njezina razvoja i korištenja. Metoda korištena u radu jest kvalitativna metoda sinteze, odnosno sustavni pregled literature koja ponajprije obuhvaća knjige, znanstvene članke, pravilnike, smjernice i izvještaje, ali i relevantne mrežne izvore. Literatura i izvori odabrani su prema kriteriju sadržajne relevantnosti i aktualnosti, suvremenosti i autorske relevantnosti, a obuhvaćaju vremensko razdoblje od 1988. do 2023. godine. Opisujući i analizirajući trendove, primjenu i izazove u području tehnologije UI, rad ukazuje na nužnost izbjegavanja generaliziranja, pojednostavljivanja i polarizacije mišljenja i stavova spram iste te ju pokušava prikazati strateškom tehnologijom 21. stoljeća koja će svojim disruptivnim djelovanjem u značajnoj mjeri oblikovati društvene politike i prakse.

Ključne riječi: društvo 21. stoljeća, (etički) izazovi, primjena i razvoj, trendovi, UI

1.Uvod

Tema umjetne inteligencije (UI) predstavlja složeno pitanje koje se može analizirati iz različitih perspektiva, te postaje sve značajnija tema istraživanja među znanstvenicima i stručnjacima različitih područja. UI je tehnologija koja značajno utječe na naše radne i profesionalne sfere, ali također ima dubok utjecaj na naš svakodnevni život, potičući stoga sve veći interes znanstvene zajednice, industrije, vlada i šireg društva. Njezin utjecaj se očituje u industriji i tržištu rada, gdje mijenja prirodu posla, strukturu radnih mjesta te zahtijeva proširenje postojećih i razvoj novih vještina. Također, UI postavlja pitanja o ljudskim odnosima i komunikaciji, dovodeći u pitanje prirodu ljudske interakcije, kako s drugim ljudima i okolinom, tako i sa samom tehnologijom.

Zadatak ovog rada je analizirati i opisati ključne aspekte utjecaja tehnologije UI na suvremeno društvo 21. stoljeća ističući potencijalne prednosti i izazove koji proizlaze iz njezine sveprisutne primjene. Uz prvotni kratki prikaz povijesnog razvoja UI, opisuju se i analiziraju trendovi u razvoju UI, primjena u odabranim područjima ljudskog djelovanja te izazovi koje predstavlja za suvremeno društvo. Analizom odabranih dimenzija utjecaja UI na suvremeno društvo, rad nastoji pružiti sveobuhvatan uvid u dinamičan odnos između tehnologije UI i suvremenog društva, njegovih društvenih struktura, društvenih institucija i građana, a s ciljem poticanja daljnjeg promišljanja i razumijevanja disruptivnih, inovativnih i jedinstvenih učinaka i utjecaja ove tehnologije.

Temeljna istraživačka pitanja na koja će se nastojati dati odgovor i koja su usmjerila sadržajni koncept rada jesu:

1. Koja je važnost, odnosno koja je uloga tehnologije UI u društvu 21. stoljeća?
2. Koji su trendovi u razvoju i primjeni tehnologije UI?
3. Koji su izazovi u razvoju i primjeni tehnologije UI?

Uz aktualne trendove, posebno je značajna etička dimenzija UI koja otvara brojna pitanja poput pitanja o transparentnosti, pravednosti, privatnosti, sigurnosti i odgovornosti. Ta su pitanja važna u svim kontekstima, a posebice u kontekstu automatizacije sustava i procesa, kao što su algoritamski sustavi za preporuku sadržaja, algoritamsko profiliranje ili algoritamsko odlučivanje.

Sadržajno rad je organiziran u četiri glavna tematska poglavlja. U uvodnom tematskom poglavlju definirat će se temeljni pojmovi vezano za UI te će se ukratko opisati povijesni razvoj temeljnih koncepata i pojmova UI, ali i same discipline. Središnji i glavni dio rada organiziran je u pet tematskih potpoglavlja, a posvećen je pitanju trendova u razvoju i primjeni UI, prikazani su odabrani primjeri konkretne primjene tehnologije UI u društvu, te su analizirani glavni izazovi u razvoju i primjeni UI. S obzirom na složenost pojma i dinamičnost i inovativnost područja UI, za svako su potpoglavlje odabran tek određeni primjeri. Tako su u kontekstu trendova analizirani algoritmizacija i pametni sustavi, generativna UI, tzv. *deep fake* tehnologija te pitanje suradnje discipline UI s disciplinom robotike. Kada su u pitanju primjeri primjene UI, rad će istražiti primjenu UI u ekonomiji i poslovanju te njezin utjecaj na tržište rada, primjenu i utjecaj UI na urbanizaciju i prometni sustav te primjenu i utjecaj UI u kreativnoj industriji. U kontekstu izazova bit će opisani rizici i prijetnje koje donosi brzi razvoj UI, uključujući pitanja transparentnosti, privatnosti, sigurnosti, pristranosti, odgovornosti i slično. Poseban naglasak bit će stavljen na odgovoran razvoj UI i nužnost uravnotežene regulacije. U konačnici, središnji dio rada problematizira pitanje pismenosti za UI te budućnost njezina razvoja i primjene. U zaključnom dijelu ističu se glavne ideje i ključna problemska pitanja vezana uz odnos između UI i društva 21. stoljeća, u skladu s prvotno postavljenim istraživačkim pitanjima.

2. Kratka povijest umjetne inteligencije

Tehnologija UI jedan je od najvećih tehnoloških trendova društva 21. stoljeća, a pojam „umjetna inteligencija” poznat je među svim disciplinama, profesijama i generacijama. No što predstavlja pojam UI, kako ga možemo definirati i kako ga možemo razlikovati u odnosu na ljudsku inteligenciju? Sljedeća potpoglavlja donose definiciju UI, opisuju povijesni razvoj glavnih pojmova i ideja u području UI i razvoj same discipline UI te analiziraju početak odnosa UI i društva.

2.1. Definicija umjetne inteligencije

Izraz „umjetna inteligencija” prvi put je upotrijebio John McCarthy 1956. godine tijekom prve akademske konferencije na tu temu. No, pokušaj razumijevanja UI pojavio se prije toga, kada je Vannevar Bush u svom djelu *As We May Think* UI opisao kao sustav koji pojačava ljudsko znanje i razumijevanje (Smith, et al., 2006).

UI, kao dio računalne znanosti, bavi se kreiranjem inteligentnih strojeva, posebno inteligentnih računalnih programa i sustava. UI se često definira kroz prizmu, tj. u usporedbi s ljudskom inteligencijom koja se može višestruko definirati, obuhvaća različite sposobnosti poput sposobnosti učenja iz iskustva, apstraktnog mišljenja i zaključivanja, razumijevanja uzročno-posljedičnih odnosa, rješavanja problema, i dr., te podrazumijeva različite vrste inteligencije poput socijalne, emocionalne, logičke, tjelesne, itd. (Leksikografski zavod Miroslav Krleža, 1992). Ona je još uvijek odlika koja se uglavnom pripisuje čovjeku, iako životinje također ispoljavaju određenu razinu i vrstu inteligencije. Disciplina UI inteligenciju je pripisala strojevima, no otvorila je pitanje njezine prirode – o kakvoj se vrsti inteligencije radi, mora li ona biti nužno biološki određena, odnosno koji sustav ili proces možemo smatrati inteligentnim? Primjerice, neki sustavi ili programi UI koji imaju impresivne izvedbe, posebno u inteligentnom ponašanju, smatraju se donekle „inteligentnima”. UI ponekad, ali ne uvijek, simulira ljudsku inteligenciju; na primjer promatranjem i proučavanjem metoda koje koriste ljudi, odnosno ljudskog ponašanja. No, znanstvenici koji se bave područjem UI mogu koristiti metode koje ne postoje kod ljudi i koje uključuju radnje koje ljudi ne mogu učiniti. Uz pojam (ljudske) inteligencije često se vezuje pojam kvocijenta inteligencije. Postupnim normalnim razvojem čovjeka povećava se njegov kvocijent inteligencije, no kod računala i UI kvocijent inteligencije se ne koristi kao koncept koji opisuje razinu njihovih intelektualnih sposobnosti. Konačni cilj discipline UI je kod strojeva postići razinu ljudske inteligencije, odnosno stvoriti

računalne programe koji rješavaju probleme i ostvaruju ciljeve jednako kao i ljudi. Hoće li i kada UI dostići razinu ljudske inteligencije još nije poznato i teško je predvidjeti.

Disciplina, odnosno područje UI obuhvaća određene grane i vrste primjene. Neke od grana UI obuhvaćaju UI pretraživanje, logičku UI, prepoznavanje obrazaca, reprezentacije znanja, rješavanje problema, planiranje i djelovanje, zaključivanje, znanje temeljeno na zajedničkom razumu i razmišljanju, učenja iz iskustva, planiranje, epistemologiju, ontologiju, heuristiku te genetsko programiranje (McCarthy, 2007). Uz navedene grane, postoje i razne primjene UI. UI se primjenjuje u video igricama, prepoznavanju govora, razumijevanju prirodnog jezika, računalnom vidu, ekspertnim sustavima, kod heurističkih klasifikacija u analizama i odlučivanju, i dr. Mladi se tako danas najčešće susreću s tehnologijom UI kroz video igre. Naime, postoje inteligentni strojevi koji mogu igrati šah poput profesionalca, no, za razliku od čovjeka, koriste računalne operacije i izračune koji pregledavaju sto tisuća pozicija u dijelu vremena. Da bi takav stroj pobijedio svjetskog prvaka mora pregledati 200 milijuna pozicija svake sekunde. Uz primjenu UI u igricama, zanimljiva je i primjena UI u računalnom vidu koji je važan za sustave navigacije, sustave prepoznavanja obrazaca i lica, i sl. Svijet u kojem živimo stvoren je od trodimenzionalnih objekata, dok su 'ulazi' u ljudsko oko i kamere računala prilagođeni dvodimenzionalnoj slici. Postoji nekolicina programa koji omogućuju djelomičnu trodimenzionalnost u računalnom vidu, no ona još uvijek nije posve razvijena kao kod ljudi (McCarthy, 2007).

Kada je u pitanju istraživanje UI, kao i svako područje istraživanja, ono ima teorijsku i eksperimentalnu stranu. Dva su pristupa istraživanjima u području UI. Prvi je biološki i odnosi se na ideju da UI trebala promatrati i oponašati ljudsku psihologiju i fiziologiju, tj. ljudsko inteligentno ponašanje i komunikaciju. Drugi pristup je fenomenološki i temelji se na zdravorazumskom proučavanju i zaključivanju o općim činjenicama i problemima. Oba pristupa imaju jednaku mogućnost uspjeha, no njihov je napredak relativno spor (McCarthy, 2007, 12).

Temeljna definicija UI, prema Hrvatskoj enciklopediji, navodi da je UI "dio računalstva koji se bavi razvojem sposobnosti računala da obavljaju zadatke za koje je potreban neki oblik inteligencije; također označava svojstvo neživog sustava koji pokazuje inteligenciju" (Leksikografski zavod Miroslav Krleža, 2021, 1). Enciklopedija dalje tumači kako se razvoj UI temelji na strojnom učenju, što omogućuje računalima da usvoje značajnu količinu znanja te da komuniciraju sa živim bićima ili drugim neživim sustavima. Prema stupnju inteligencije, UI se dijeli na jaku (engl. *strong AI*) i slabu (engl. *weak AI*). Slaba UI obuhvaća postignuća u području

UI do danas, dok se jaka UI odnosi na onu koja još nije ostvarena. Jaka inteligencija uključuje razinu inteligencije jednaku ljudskoj, poznatu i kao opća UI (eng. *general AI*), ili čak nadmoćniju od ljudske, što se naziva umjetna superinteligencija (eng. *artificial superintelligence*), a podrazumijeva da je UI postigla određeno stanje svijesti (Leksikografski zavod Miroslav Krleža, 2021). Umjetno inteligentnu tehnologiju, prema Davenportu (2021), još nazivaju i kognitivnom tehnologijom, koja posjeduje sposobnosti koje su prije pripadale samo ljudima, kao što su znanje, percepcija i sposobnost rješavanja problema. Postoje rasprave o tome je li UI bliska oponašanju ljudskog mozga i njegovog funkcioniranja. Međutim, Davenport smatra da UI još uvijek nije ni približno dostigla razinu funkcioniranja ljudskog mozga.

2.2. Povijesni razvoj osnovnih pojmova i koncepata umjetne inteligencije kroz književnost i znanost

Počeci UI sežu u 1940-e godine. Naime, 1942. godine, američki pisac Isaac Asimov objavio je priču pod nazivom *Runaround*, u kojoj opisuje razvoj robota od strane inženjera Gregoryja Powella i Mikea Donovana, i u kojoj su definirana tri zakona robotike koja su se smatrala polazištem za kreiranje etičkog okvira za razvoj UI. Prvi zakon nalaže da robot ne smije ozlijediti ljudsko biće ili ga ugroziti, drugi zakon kaže da robot mora slušati naredbe ljudi osim ako su u sukobu s prvim zakonom, a treći zakon nalaže da robot mora štiti vlastiti integritet sve dok ta zaštita nije u sukobu s prvim ili drugim zakonom. Ova priča inspirirala je mnoge znanstvenike da počnu promišljati o (etičnom) ponašanju robota i UI, uključujući Marvinu Minskog koji je kasnije osnovao Laboratorij umjetne inteligencije na Institutu tehnologije u Massachusettsu (Hanlein i Kaplan, 2019). Asimov je putem fiktivne priče postavio temelje za raspravu o stvarnim političkim i etičkim problemima (Morioka, 2023). Otprilike 20-ak godina prije toga, Karel Čapek napisao je djelo *Rossumovi Univerzalni Roboti* (1920), dramu u kojoj se problematizira „oslobođenje” i „humanizaciju” robota koja na kraju prijete nestankom čovječanstva. Radi se o djelu u kojem se po prvi puta spominje riječ „robot” i u kojem se dovodi u pitanje pravo robota na osjećaje, misli, ljubav i „dušu”. Također, jedna od prvih ideja o kreiranju umjetnog života koja bi se mogla povezati s idejom umjetne inteligencije javlja se u djelu Mary Shelley *Frankenstein ili Moderni Prometej* (1818) koje upozorava na opasnost moderne znanosti i eksperiment kreiranja života. Navedena problemska pitanja i danas su aktualna u kontekstu razvoja UI i robotike, posebno u odnosu na pitanja etičkog odnosa između čovjeka i UI/robota te pozicije UI/robota kao moralnog agenta, odnosno moralnog pacijenta.

Za razliku od „robota”, prva uporaba pojma „umjetna inteligencija” bilježi se 1955. godine kada taj pojam McCarthyjevu koristi na radionici u Dartmouthu, no više o tome slijedi u sljedećem potpoglavlju.

Uz razvoj umjetne inteligencije često se vezuju funkcionalizam i teorija uma. Funkcionalizam kao teorija pretpostavlja uzročni odnos između materijalnog (mozga) i mentalnog (uma), između hardvera i softvera, između *inputa* i *outputa*. Glavna ideja teorije funkcionalizma jest da se stanja uma mogu ostvariti u različitim medijima, kako u biološkom mozgu i neuronima, tako i u nebiološkom mediju. Nije važan medij, već očuvanje funkcije, tj. procesa. Slijedom toga, inteligenciju i mentalna stanja i procese moguće je ostvariti i u umjetnom mediju – UI i robotu. Teorija uma podrazumijeva kognitivni mehanizam i složenu funkciju koja se razvija odrastanjem i socijalizacijom. Ona je važna ne samo za samorazumijevanje (pripisivanje mentalnih stanja samome sebi), već i za razumijevanje drugih (pripisivanje mentalnih stanja drugima). Kada se radi o kognitivnim arhitekturama UI, tj. modeliranju kognicije UI uglavnom se primjenjuju dva pristupa – simbolizam i konekcionizam/konektivizam. Simbolizam pretpostavlja postojanje baze eksplicitnog znanja uređenog na određeni način i pravila koja se primjenjuju za pretraživanje baze, zaključivanje, i dr. Konekcionizam podrazumijeva paralelnu i distribuiranu komputaciju apstraktnih neurona koji su povezani u jednu mrežu i među kojima se odvijaju jednostavne interakcije. Za razliku od simbolizma, konekcionizam pretpostavlja uporabu implicitnog znanja karakterističnog za određeni model (History of Artificial Intelligence, n.d.; Smith i sur., 2006.; Valerjev, 2006).

Razvoj UI može se tumačiti s različitih stajališta, pa tako i filozofskog. Jedan od filozofskih problema je i problem okvira. Problem okvira predložili su John McCarthy i Patrick J. Hayes 1969. godine, a to je zapravo problem s kojim se UI suočava zbog toga što ona autonomno ne može razlikovati važno od nevažnog te razumjeti situaciju i posljedice akcije u kontekstu nepredviđenih okolnosti i neočekivanih situacija. Taj problem mogao bi nastati kad bi se dopustilo da umjetno inteligentni strojevi samostalno djeluju i rješavaju probleme u stvarnom, fizičkom svijetu. Godine 2016., čak godinama nakon predlaganja problema okvira, Margaret T. Boden piše kako on i dalje nije riješen, te da postoje kontradiktorne tvrdnje koje problem okvira čine još složenijim. Filozof Hubert Dreyfus, jedan od kritičara simboličke UI i nekontekstualizma, tvrdi da UI nailazi na problem okvira jer, prije svega, ne razumije općenito koje znanje je važno u kojoj situaciji, odnosno da objekt ima značenje samo u kontekstu konkretne situacije (eng. *highly context bound*). UI, prema njegovom mišljenju, nije prava inteligencija ako do razumijevanje i spoznaje dolazi tek manipuliranjem unutarnjih simbola na

temelju određenih internih pravila, a ne na temelju (složenog) situacijskog konteksta. Dreyfus se također poziva na fenomenologiju i hermeneutiku Martina Heideggera te se osvrće na njegov izraz priručnost (eng. *ready-to-hand*) koji spominje u svojoj knjizi *Sein und Zeit*, (hrv. prij. *Bitak i vrijeme*) kako bi opisao bitak bića kao pribora. Taj izraz opisuje priručno odnošenje spram svijeta. Dreyfus zaključuje da tradicionalnoj UI nedostaje dimenzija priručnosti te da UI može riješiti problem okvira i postati istinska UI ukoliko postane takozvana heidegerska UI (eng. *Heideggerian AI*), tj. implementira dimenziju priručnosti. U konačnici, UI nedostaje dimenzija tjelesnosti, tj. utjelovljenja koje pruža dodatni kontekst i nadilazi puko bivanje u obliku bazičnog mišljenja i rješavanje problema. Proučavajući istraživanja i napore u području UI Dreyfus je zaključio da do sada nitko nije realizirao takvu vrstu UI (Morioka, 2023).

Utjelovljena UI također je jedan od temeljnih pojmova u području UI, a odnosi se na ideju da je inteligencija utjelovljena, tj. da je inteligencija određenog entiteta rezultat njegove interakcije s okolinom. Ta je interakcija moguća samo ako entitet, uz mentalni, ima i tjelesni aspekt. Kroz (složenu) interakciju s okolinom entitet stječe motornu inteligenciju tzv. niske razine i spoznaju tzv. visoke razine. Krajni cilj discipline UI jest postići u budućnosti punu utjelovljenost umjetnih neuronskih mreža koja bi u konačnici omogućila samo-organizaciju UI. U tom pogledu, važna je filozofija uma koja se bavi pitanjem odnosa uma, mentalnih svojstava i svjesnosti s jedne strane i tijela s druge strane.

Na kraju, važno je spomenuti pojam ekspertnih sustava. „Ekspertni sustavi (ES) su računalni sustavi koji simuliraju ljudsko donošenje odluka. Mogu se integrirati s informacijskim sustavima kako bi poboljšali njihovu točnost i učinkovitost“ (Singh i sur., 1996; citirano prema Asemi i sur., 2020, 4). Ekspertni sustavi su zasnovani na znanju (bazama znanja) i oni predstavljaju načine na koje se može riješiti određeni problem. Radi se o računalnim programima koji su namijenjeni rješavanju složenih problema iz specijaliziranih područja. Za razliku od konvencionalnih programa, ekspertni sustavi biraju pravila automatski i slijede postupak logičke dedukcije. Glavni čimbenik u izvedbi ekspertnog sustava, uz specijaliziranu znanstvenu bazu, je znanje čovjeka čime sustav simulira ljudski način rješavanja problema (Leksikografski zavod Miroslav Krleža, 2013-2024; Mainzer, 2006).

2.3. Razvoj discipline umjetne inteligencije

Razdoblje između 1940. i 1960. predstavlja doba početka razvoja UI i kibernetike. Tada se razvija prvi računalni i matematički model biološkog neurona. Razvoj UI bio je podržan

tehnološkim razvojem koji je bio ubrzan povijesno-društveno-političkim kontekstom Drugog svjetskog rata.

U vrijeme kada je priča *Runaround* bila aktualna u Americi, u Engleskoj je matematičar Alan Turing radio na razvoju stroja za rješavanje kodova, nazvanog *bomba* (eng. *the Bombe*), koji je izradio za britansku vladu s ciljem dešifriranja Enigma koda korištenog od strane njemačke vojske tijekom Drugog svjetskog rata. *Bomba* je smatran prvim uspješnim elektromehaničkim strojem, odnosno računalom, jer je postigao ono što ni najbolji ljudski matematičari nisu mogli - dešifrirati Enigma kod pronalaskom odgovarajućih postavki rotora. Taj događaj potaknuo je Turinga na pisanje rada objavljenog 1950. godine pod nazivom *Computing Machinery and Intelligence*. U tom radu opisao je proces kreiranja i testiranja inteligentnih strojeva (Hanlein i Kaplan, 2019). Taj rad je objavljen nekoliko godina prije nego što je termin "umjetna inteligencija" prihvatio John McCarthy. Turing je postavio pitanje "Mogu li strojevi razmišljati?" te predložio test koji bi trebao dati odgovor na to pitanje pod nazivom imitacijska igra koji je postao poznat kao Turingov test. Test je pristupao pragmatički i pretpostavljao da strojevi mogu razmišljati. Prvotni cilj testa bio je da ispitivač, prepozna razliku između žene i muškarca. Kasnije bi se muškarac ili žena zamijenili sustavom UI pa bi se postavilo pitanje može li ispitivač prepoznati komunicira li s drugim čovjekom ili sa strojem. Stroj bi položio Turingov test ako bi mogao voditi opušten razgovor s ispitivačem s razumijevanjem. Ako bi to postigao, Turing bi vjerovao da strojevi mogu razmišljati. Dva potencijalna problema s Turingovim testom uključuju pitanje može li se smatrati da je imitiranje ljudskog ponašanja inteligencija te je li inteligencija moguća bez položenog Turingovog testa. Prema Turingu, ako računalo položi Turingov test, smatra se inteligentnim, no postoje mnogi aspekti koje treba razmotriti prije nego što se računalo odredi inteligentnim. Pretpostavka Alana Turinga da će test biti položen do 2000. godine nije se ostvarila (Smith i sur., 2006).

Nekoliko godina kasnije, točnije 1956. izraz "umjetna inteligencija" je službeno uveden u uporabu kada su Marvin Minsky i John McCarthy održali osmotjedni kongres na fakultetu Dartmouth u New Hampshire-u pod nazivom *Dartmouth Summer Research Project on Artificial Intelligence* (DSRPAI). Taj projekt označava početak takozvanog proljeća UI te su u njemu sudjelovali brojni znanstvenici koji se kasnije smatraju utemeljiteljima UI. Glavni suradnici bili su Nathaniel Rochester, koji je kasnije dizajnirao IBM 701, prvo komercijalno znanstveno računalo i matematičar Claude Shannon, koji se smatra tvorcem teorije informacije. Glavni cilj projekta bio je okupiti znanstvenike iz različitih područja znanosti i s njima kreirati novo područje istraživanja koje se bavi izgradnjom strojeva koji mogu simulirati ljudsku

inteligenciju (Hanlein i Kaplan, 2019). Originalni strojopis projekta sastojao se od 17 stranica i naslovne stranice. Kopije se i dalje čuvaju u arhivima Dartmouth fakulteta i Sveučilišta Stanford. Prvih pet stranica donosi prijedlog projekta, dok preostali dio opisuje kvalifikacije i interese organizatora. U prijedlogu projekta navedena su neka problemska pitanja u području UI: automatska računala, uporaba prirodnog jezika, neuronske mreže, teorija veličine izračuna, samopoboljšanje, apstrakcije te slučajnost i kreativnost (McCarthy i sur., 2006). Nakon spomenutog proljeća UI, slijede ljeto i zima, odnosno razdoblja uspona i padova. Nakon DSRPAI kongresa, došlo je do značajnih uspjeha u području UI. Istaknuo se razvoj računalnog programa, odnosno prvog tzv. *chatbota* ELIZA, koji je kreirao Joseph Weizenbaum na Institutu Tehnologije u Massachusettsu između 1964. i 1966. godine. ELIZA program bio je alat koji je omogućavao simulaciju razgovora prirodnim jezikom s čovjekom i jedan od prvih programa koji je bio u mogućnosti pokušati proći Turingov test. Uz taj postojalo je još nekoliko uspješnih primjera napretka UI, a uz tako uspješne projekte, krenula su proricanja uspjeha za općeniti napredak inteligencije kod strojeva. Tako je Marvin Minsky, 1970. godine, u jednom intervjuu izjavio kako će strojevi s inteligencijom prosječnog čovjeka biti razvijeni kroz 3 do 8 godina. Ali kako se to nije dogodilo započinju kritike na razvoj UI (Hanlein i Kaplan, 2019). 1973. godine Kongres Sjedinjenih Američkih Država kritizira visoke troškove istraživanja UI. Te iste godine britanski matematičar James Lighthill objavio je izvješće u kojem propituje optimistična gledišta znanstvenika u području UI. Lighthill je tvrdio kako će strojevi dosegnuti razinu iskusnog amatera u nekim igricama poput šaha, a da će im sve ostalo biti nedostižno. U tom razdoblju započela je takozvana zima UI koja je trajala od 1972. do 1986. godine. Nakon razdoblja zime, došlo je do razdoblja takozvane jeseni, odnosno žetve UI kada se u fokus vraća strojno učenje. 1997. godine IBM-ov šahovski program Deep Blue uspio je pobijediti svjetskog prvaka u šahu, Garyja Kasparova te je time djelomično opravdao Lighthillovu izjavu. Deep Blue bio je sposoban obraditi 200 milijuna poteza u sekundi i pobijediti majstora šaha Kasparova.

Metode za postizanje prave UI počele su se razvijati 1940-ih, kada je kanadski psiholog Donald Hebb razvio teoriju učenja pod nazivom Hebbijansko učenje, koja replicira proces neurona u ljudskom mozgu. Ta je teorija dovela do napretka umjetnih neuronskih mreža (Hanlein i Kaplan, 2019). Neuronske mreže, sustavi umreženih umjetnih neurona koji oponašaju rad ljudskog mozga, široko se koriste za identifikacijske i klasifikacijske zadatke. Danas u brojnim istraživanjima, uz neuronske mreže spominje se i rudarenje podacima koje primjenom algoritama procesira i izdvaja uzorke iz podataka (Botto-Tobar i sur., 2022).

Neuronske mreže kao i UI fokusirane su na pitanje ponašanja, odnosno pretpostavljaju da inteligencija leži u ponašanju koje program ili neuronska mreža proizvodi, a zapravo njezina najvažnija odlika je dati ispravan i željeni iskaz (Hawkins i Blakeslee, 2005). 1969. godine Marvin Minsky i Seymour Papert dokazali su da računala nemaju dovoljnu snagu za obradu i obavljanje zadataka umjetnim neuronskim mrežama. No koncept umjetnih neuronskih mreža doživio je povratak u obliku grane strojnog učenja poznate kao duboko učenje. 2015. godine AlphaGo program koji je razvio Google postigao je velike rezultate koristeći umjetnu neuronsku mrežu dubokog učenja. Taj program uspio je pobijediti svjetskog prvaka u igri Go koja je kompleksnija od šaha, a vjerovalo se da računalo nikad neće moći pobijediti ljude u toj igri. Duboko učenje temelji se na dubokim neuronskim mrežama koje imaju puno slojeva i koriste veliki broj parametara. Duboko učenje danas ima široku primjenu, od računalnog vida do razumijevanja govora i obrade prirodnog jezika jer koristi naučene nelinearne transformacije za kreiranje složenih podatkovnih reprezentacija (Hanlein i Kaplan, 2019). Na primjer, provedeno je istraživanje o bolestima listova kukuruza i u tom istraživanju korišteno je duboko učenje kako bi se klasificirale i identificirale slike listova. Dakle, algoritmi dubokog učenja su kategorizirali ulazne slike kukuruznih listova u unaprijed definiranu klasu, odnosno sustavi dubokog učenja podučeni su slikama kukuruznih listova iz baza podataka. Nakon obuke, sustavi dubokog učenja mogu vršiti binarne ili višerazredne klasifikacije kukuruznih listova (Bottotobar i sur., 2022).

U 21. stoljeću UI napreduje, pojavljuju se razni programi i sustavi koji olakšavaju čovjeku svakodnevicu. 2011. godine pojavljuju se Appleova Siri i Microsoftova Cortana koje upravljaju aplikacijama na temelju glasovnih naredbi čovjeka putem mikrofona u mobilnom uređaju. Razvija se Googleov prevoditelj koji olakšava prevođenje teksta sa svih jezika svijeta, a krajem 2022. godine dolazi do razvoja generativne UI, velikog jezičnog modela ChatGPT, generatora slika i fotografija iz teksta DALL-E i Midjourney. UI neprestano napreduje (Leksikografski zavod Miroslav Krleža, 2021). Primjerice, Siri kao inteligentni asistent koji funkcionira putem korisničkog sučelja prirodnog jezika, ima sposobnost prirodnog razgovora s čovjekom. Softver koji je temelj za Siri kreiran je na kalifornijskom istraživačkom institutu SRI International, a tvrtka Apple kupila ga je 2010. godine te ga prvi put predstavila na mobilnom uređaju iPhone 4S. Ponekad bizarni i frustrirajući odgovori koje Siri daje vrlo su poznati, ali moć tehnologije i pomoć koju može pružiti je neupitna (Brynjolfsson i McAfee, n.d.).

Uz razvoj UI, spominje se i izraz inteligentni sustavi. „Inteligentni sustavi (IS) su definirani kao bilo koji formalni ili neformalni sustav koji je u mogućnosti prikupljati i

obrađivati podatke, interpretirati podatke primjenom tehnologija UI i poslovne inteligencije te pružati obrazložene prosudbe donositeljima odluka kao osnovu za njihovo djelovanje“ (Sharda i sur., 2017; citirano prema Asemi i sur., 2020, 3-4). Uz inteligentne sustave pojavljuju se već spomenuti ekspertni sustavi.

Strojno učenje jedan je od temelja tehnologije UI i podatkovne znanosti. Ono predstavlja softver koji može „učiti“. Neki znanstvenici i stručnjaci smatraju da strojno učenje nije uistinu učenje jer nema prave kognitivnosti te da jedino ljudi mogu učiti. Shodno tome, strojno učenje ima „malo ili nimalo sličnosti s onim što se vjerojatno događa u ljudskim glavama“ (Boden, 2016, 16; citirano prema Coeckelbergh, 2020, 96). Strojno učenje temelji se na statistici, to je statistički proces. Koristi se za razne zadatke, prepoznaje uzorke apstrakcijom značajki. Algoritmi identificiraju uzorke i te uzorke koriste za objašnjavanje podataka i davanje prognoza za buduće podatke (Coeckelbergh, 2020). Strojno učenje koje se oslanja na velike podatke zanimljivo je zbog mogućnosti prikupljanja i analiziranja velikih količina podataka. Veliki podaci mogu značiti da su podaci različitih organizacija kombinirani i spojeni. Dovedi su do utemeljenja znanosti o podacima, interdisciplinarnе znanosti koja koristi statistiku i algoritme za nalaženje korisnih skupova podataka (Coeckelbergh, 2020).

2.4. Početak odnosa umjetne inteligencije i društva

Pojava i razvoj UI utjecali su na gotovo sve aspekte i segmente društvenog života, između ostalog rad, obrazovanje, znanost i istraživanje, komunikaciju, odnose, zabavu i slobodno vrijeme.

Tehnologija općenito, pa tako i UI, povezana je s područjima ljudskog rada na više načina. Od početka industrijske revolucije pa i prije nje postoji imperativ uvođenja nove tehnologije kako bi se povećala produktivnost, učinkovitost i konkurentnost. No, postavlja se pitanje dovodi li to do ujedno i do gubitka radnih mjesta? Nilsson je 1984. sumnjao u budućnost gdje inteligentni strojevi pružaju usluge koje su nekad pružali ljudi. Razni autori su se slagali s tom tvrdnjom. Stvarna situacija je da je uvođenje automatizacije već imalo i nastavit će imati utjecaj na radnike i tržište rada. Problemi koji tako nastaju obuhvaćaju smanjenje razine stručnosti, ugrožavanje ljudskog dostojanstva i samopoštovanja, povećani stres, ograničen broj poslova. Pojavljuje se i prijetnja ljudskoj autonomiji, onome što nas čini ljudima, odnosno našoj sposobnosti upravljanja i razmišljanja. Jednako tako i neki budući projekti mogu predstavljati prijetnju za čovjeka (Rosenberg 1988, 15). Ipak, UI može dovesti i društvene koristi. Jedna od koristi jest da UI može obavljati zadatke i poslove koji su visokorizični i opasni za čovjeka,

poput rada s opasnim otpadom ili otrovnim supstancama. Takvi se poslovi često nazivaju 3D poslovima (eng. *3D jobs (Dirty, Dull and Dangerous)*). Inteligentni sustavi mogu se koristiti u uredima, tvornicama, u terenskom radu i na drugim mjestima.

Uz utjecaj UI na rad ljudi, značajan je i njezin utjecaj na pitanje privatnosti. Uporabom baza podataka koje mogu biti javne ili privatne, stvara se mogućnost zloupotrebe podataka zbog lakoće pristupa istima. Pojavljuju se izrazi računalno profiliranje i računalno usklađivanje. Računalno profiliranje (engl. *computer profiling*) predstavlja pokušaj predviđanja kriminalnog ponašanja određivanjem profila na temelju dostupnih podataka i uspoređivanjem sa sličnim definiranim profilima koji se mogu pretraživati u postojećim bazama. Računalno usklađivanje (engl. *computer matching*) koristi se kao križanje datoteka kako bi se utvrdilo jesu li pojedinci počinili određeni zločin, a temelji se na otkrivanju nepravilnosti u datotečnim zapisima (Rosenberg, 1988, 16). Primjer može biti osoba koja dobije značajan prihod za prodaju zemljišta, a u isto vrijeme prima naknadu za nezaposlenost. Vlade širom svijeta donijele su zakone protiv zloupotrebe privatnih podataka, ali napretkom UI stvaraju se novi izazovi za građanske slobode i privatnost. Donošenje odluka (engl. *decision-making*) još je jedno područje na koje je UI utječe. Radi se o procesu u kojem pojedinci, organizacije, tvrtke i institucije korištenjem prikupljenih i organiziranih informacija i podataka, a na temelju iskustva i znanja donose odluke i izbore (Rosenberg, 1988: 17). Autori Jack Williamson (1963, 1981; citirano prema Rosenberg, 1988, 17) i Jack Chalker (1986; citirano prema Rosenberg, 1988, 17) pretpostavljaju da računala preuzimaju moć kako bi spriječili ljude od činjenja velikih grešaka i mogućeg samouništenja. Povećanom upotrebom računala opaža se smanjenje ljudskog donošenja odluka, posebno u kritičnim situacijama. Područje računalnog donošenja odluka započinje sve prisutnijim oslanjanjem na UI.

Sfera društvene organizacije također je pod utjecajem UI. Svjetska mreža (World Wide Web) donijela je ogromnu količinu strojno čitljivih podataka i informacija, što omogućuje da milijarde slika, tekstualnih stranica i raznih podataka postanu dostupnima u obliku digitalnih resursa. Bez dijela tih podataka, napredak u području UI kakvog poznajemo danas bio bi nezamisliv. IBM je 2011. godine najavio eru kognitivnog računarstva s Watsonom, sustavom sposobnim pobijediti ljudske igrače u zadatku općeg znanja. Budući da je sustav pokazao iznimnu sposobnost, odnosno da je inteligencija mogla biti primijenjena i na druge domene, osmišljeni su UI asistenti koji obavljaju rutinske zadatke zajedno s liječnicima u sklopu IBM-ovog sustava Watson Health (Shadbolt, 2022).

3. Umjetna inteligencija i društvo 21. stoljeća

U odnosu na dosadašnji povijesni pregled pojma i područja UI, odnos UI i suvremenog društva obilježen je nešto drukčijim trendovima. Prvotni pristupi razumijevanju problematike UI i njezina utjecaja na društvo bili su usmjereni uglavnom na pitanje hoće li i u kojoj mjeri UI dostići razinu ljudske inteligencije te u kojim je segmentima UI bolja od čovjeka. Sada već dokazano bolja, barem po pitanju brzine i količine obrade podataka, UI je obilježena nekim novim trendovima od kojih je danas možda najaktualniji razvoj generativne UI. U posljednje vrijeme novi se trendovi predviđaju na početku svake kalendarske godine, a zbog emergentne prirode tehnologije UI i sve češće nepredvidivog razvoja ta se predviđanja nerijetko

'modificiraju' tijekom godine. Tako su, primjerice, neki od trendova koji će obilježiti 2024. godinu UI agenti, multimodalna UI, *open source* UI, kastomizirana UI, i dr. (Craig, 2024).

Transformativni učinak UI na područje poslovanja i financija, obrazovanja, medicine i zdravstva, prometa i urbanizacije te odnosa čovjeka i tehnologije, ukazao je na brojne prednosti primjene UI, poput povećanja učinkovitosti i uspješnosti i smanjenja troškova (Sen, 2023; West i Allen, 2018). Iako su trendovi, prije svega, najprisutniji u industriji i poslovanju, za očekivati je da će imati i sve veći utjecaj na društvo u cjelini zbog čega je posebno važno uzeti u obzir etičke aspekte razvoja i primjene UI, mogućnosti odgovornog razvoja i upravljanja te širi društveni utjecaj. U sljedećim poglavljima analizirat će se neki od trendova u području UI, odabrani primjeri njezine primjene u društvu, kao i izazovi i prijetnje s kojima se društvo pritom suočava. Na kraju, pojasnit će se potreba za kreiranjem etičkih i zakonskih okvira koji će osigurati odgovoran i održiv razvoj UI. Naime, dinamičan razvoj i integracija UI nije samo pitanje tehnološke infrastrukture i mogućnosti, već i šireg utjecaja na društvo i njegove građane pri čemu su jednako važni aspekti istraživanja i razvoja, prakse i primjene, regulacije i upravljanja.

3.1. Trendovi u razvoju umjetne inteligencije

Tehnologije UI eksponencijalno se razvijaju i utječu na mnoga područja djelatnosti. Na primjer, strojno učenje se primjenjuje u marketingu i digitalnom marketingu gdje se kreiraju oglasi usmjereni na ciljane skupine (tzv. ciljano oglašavanje). Trgovine privlače kupce personaliziranim aplikacijama i mrežnim stranicama. Primjerice, trgovina Macy's surađuje s IBM-ovim Watsonom i Cognitive Scale-om, pri čemu Cognitive Scale stvara osobni profil kupca i odjeće koju voli, a Watsonova aplikacija, Macy's OnCall, odgovara na pitanja kupaca i usmjerava ih na željene odjele u fizičkim trgovinama. Trendovi su raznoliki i sveprisutni. Još jedan primjer čestog trenda UI u modnoj industriji su virtualni stilisti, čiji je cilj pomoći kupcu u odabiru odjeće i poticati ga na internet kupovinu, kao što to radi trgovkinja u fizičkoj trgovini. Konkretni primjer je brand Levi's koji koristi takav sustav, koji je zapravo kombinacija razgovornog i vizualnog sučelja. Slično tomu, sustav za odgovaranje na upite, Amelia, pomaže američkoj banci koja zaprima dvije milijarde poziva svake godine. Uspijeva odgovoriti barem na jednostavnija pitanja vezana za osobne račune i transakcije. Slične tehnologije koriste se i u organizacijskim jedinicama za komunikaciju sa zaposlenima, u lancu nabave i slično. Trend virtualnog asistenta ili pomoćnika sve je češća pojava na tržištu rada. Integracija UI tehnologije u poslovanje cilj je velikog broja tvrtki. Industrijski strojevi, vozila, kućanski aparati, neke su

od platformi koje mogu biti podloga za korištenje UI u industriji, tvrtkama i kućanstvima (Davenport, 2021).

3.1.1. Algoritmi umjetne inteligencije i pametni sustavi

Smatra se da je optimalno zaustavljanje, poznato i kao problem optimalnog zaustavljanja (engl. *optimal stopping*), ključno kako bi se pronašla najbolja prilika za postizanje željenog cilja. Prema ovom pristupu, preporučuje se ulaganje otprilike 37% vremena ili resursa u istraživanje prije nego što se donese konačna odluka. Na primjer, ako tražimo stan i imamo 29 dana za potragu, trebali bismo izdvojiti oko 37% toga vremena, odnosno 11 dana, kako bismo pronašli najbolji mogući stan koji je dostupan. Algoritmi su ključni u rješavanju problema optimalnog zaustavljanja. Oni predstavljaju niz koraka ili pravila koji omogućuju sustavno rješavanje problema i donošenje odluka. Važno je napomenuti da su algoritmi stariji od samih računala te su se koristili i prije nego što su računala izumljena. Danas se algoritmi povezuju s različitim disciplinama, uključujući kognitivne znanosti, psihologiju, ekonomiju i druge. U digitalnom svijetu, tvrtke poput Amazona i Googlea koriste, primjerice, A/B testiranje kako bi poboljšale svoje usluge. Ovaj postupak uključuje stvaranje različitih inačica iste mrežne stranice s različitim varijablama, poput drugih boja ili slika na zaslonu. Korisnicima se nasumično prikazuju mrežne stranice, a zatim se analizira koja inačica ostvaruje bolji rezultat kod korisnika na temelju statističkih podataka i rada algoritama. Glavni cilj tvrtki je optimizirati korisničko iskustvo i povećati prihode, a algoritmi im pomažu kako bi bolje razumjeli preferencije i ponašanje korisnika.

Cijeli internet postao je kontrolirani eksperiment. Unatoč svim prednostima korištenja algoritama, ljudi se još uvijek suočavaju s izazovnim situacijama u kojima često donose odluke u stanju velike nesigurnosti, nepredvidivosti ili vremenske ograničenosti. U takvim situacijama, čak ni najnapredniji algoritmi ne mogu postići u potpunosti uspješan rezultat, a upitno je i hoće li takvi savršeni algoritmi ikada biti razvijeni (Christian i Griffiths, 2016). Napretkom tehnologije, algoritmi, veliki podaci i pametni telefoni postaju sve prisutniji polako kolonizirajući ljudsko društvo. Ekonomija i tržište sve više bivaju podložni upravljanju algoritama i strojeva. Raznovrsni stalno prisutni oglasi oslabljuju našu svijest i sposobnost pažnje, maštanja i koncentracije, konstantno smo izloženi sadržajima koji nam se nameću. Osim koloniziranja društva, svijesti i pažnje, algoritmi infosfere ugrožavaju privatnost i slobodu. Mrežne stranice preuzimaju korisnikove podatke i potencijalno mogu profilirati korisnika i predvidjeti njegovo ponašanje.

Benasayag navodi da smo ušli u razdoblje algoritamske vlasti, u kojem je donošenje odluka sve više prepušteno UI (Benasayag, 2021). Algoritmi UI sve se češće koriste u razne svrhe, stoga je važno razvijati transparentne algoritme čije se funkcioniranje može provjeriti i razumjeti. Važno je osigurati otpornost algoritama na manipulaciju. Primjerice, pri skeniranju prtljage u zračnoj luci, algoritmi moraju biti otporni na namjerne pokušaje pronalaska slabosti u sustavu, kao što je neutraliziranje prepoznavanja oružja raznim metoda koje 'ometaju' rad algoritama. Ključni kriterij informacijske sigurnosti jest otpornost na manipulaciju. Transparentnost, provjerljivost, nepristranost, predvidljivost, odgovornost i sprječavanje stvaranja problema i zapreka nevini osobama, kriteriji su koje algoritmi moraju zadovoljiti. Navedeni kriteriji također trebaju biti zadovoljeni kod algoritama koji se koriste u svrhu prosudbe i zaključivanja. Kako bi se izbjegli određeni rizici i opasnosti, algoritmi UI s performansama sličnim ljudskima namjerno su programirani za kompetencije samo u određenoj, ograničenoj domeni. Na primjer, algoritam Deep Blue bio je svjetski prvak u šahu, ali ne posjeduje sposobnost vožnje automobila ili provođenja znanstvenih istraživanja (Bostrom i Yudkowsky, 2014).

Konkretni primjer inteligentnog sustava koji koristi algoritme jest Netflix, servis za streaming na pretplatu koji broji preko 75 milijuna pretplatnika diljem svijeta. Netflix koristi softverske alate poput MySQL-a i Java. Način na koji interagira s pretplatnicima je putem prikupljanja njihovih podataka s uređaja na kojima koriste Netflix te im tako pruža sadržaj prilagođen njihovim preferencijama. Netflix koristi sustav preporuka filmova i serija, koji se zapravo sastoji od skupa algoritama koji povezuju pretplatnike sa sadržajem koji ih zanima. Taj aspekt predstavlja najpoznatiji aspekt rada inteligentnog sustava na Netflixu. Što se pretplatnici duže zadržavaju na platformi, sustav više uči o njima i predlaže im najprikladniji sadržaj. Takozvano skladište podataka Netflix-a posjeduje 10 petabajta informacija, pri čemu 1 petabajt odgovara 13,3 godine snimljenog video sadržaja. Algoritmi Netflix-a obrađuju velike količine filmskih podataka koje dobivaju od obučeniha označitelja filmova (Frank i sur., 2017).

3.1.2. Generativna umjetna inteligencija

Nakon godina istraživanja, UI dolazi do prekretnice. UI počinje, u manjoj ili većoj mjeri, utjecati na sve generacije u gotovo svim sferama ljudskog djelovanja. To se posebno odnosi na razvoj generativne UI. Generativna UI odnosi se na kategoriju algoritama UI koji generiraju, odnosno stvaraju nove *outpute* (učinke) koji su utemeljeni na podacima pomoću kojih su algoritmi obučeni. Tradicionalni sustavi UI dizajnirani su za prepoznavanje uzoraka i predviđanja, a generativna UI stvara novi sadržaj u obliku zvuka, teksta, slike i video zapisa.

Kako bi stvorila novi sadržaj generativna UI koristi duboko učenje temeljeno na generativnoj kontradiktornoj mreži (eng. *generative adversarial network*, (GAN)). GAN je vrsta strojnog učenja koje se sastoji od dvije neuronske mreže: generatora koji kreira nove podatke i diskriminatora koji ocjenjuje podatke. Generator poboljšava svoje učinke na temelju povratnih informacija koje dobije od diskriminatora (Routley, 6.2.2023, World Economic Forum).

Generativna UI može poslužiti za razne primjene, kao što su kreiranje slike, teksta, videa i zvuka. Može kreirati slike pomoću već postojećih, kao što je kreiranje portreta pomoću slike nečijeg lica. Također kreira tekst u obliku poezije, kratkih priča, članaka te prevodi tekst s jednog jezika na drugi. Može generirati zvučne efekte pa čak i glasove. Mnogi strahuju da će generativna UI i automatizacija dovesti do povećanja nezaposlenosti zbog toga što strojevi sada mogu obavljati neke funkcije koje su nekada bile karakteristične samo za ljudsku aktivnost i rad. Strahuje se da će veća prisutnost UI dovesti do smanjenja broja poslova, posebno u industrijama koje se vezuju za proizvodnju, unos podataka, korisničku podršku i kreativnu industriju. Industrije koje se nalaze na tzv. udaru generativne UI su oglašavanje, umjetnost i grafički dizajn te zabavna industrija, primjerice kreiranje novih video igrica i drugih video sadržaja. Jedan od pozitivnih aspekata alata generativne UI jest da mogu olakšati i ubrzati izvršenje određenih zadataka te tako osloboditi vrijeme ljudima kako bi se mogli usredotočiti na dodanu vrijednost posla (Routley, 2023). Generativna UI tako ima i negativan i pozitivan utjecaj. Može doprinijeti širenju dezinformacija, a otvara i etička pitanja vezano za sadržaje koje kreira. Važno je uspostaviti načelo ravnoteže, uvijek pristupiti korištenju generativne UI s oprezom, biti svjestan potencijalnih rizika i koristiti ju na odgovoran i etički pravilan način te za opće dobro društva (Desautels Faculty of Management, McGill University, 2023).

ChatGPT je trenutno najpoznatiji alat generativne UI. Proizveden je od strane američke tvrtke Open AI iz San Francisca (Mandić, 2023). ChatGPT objavljen je u studenom 2022. godine te u nazivu sadrži riječ *chat*, što opisuje njegovu funkcionalnost u obliku *chatbota*. GPT je kratica za Generative Pre-trained Transformer, vrstu velikog jezičnog modela. Na početku je nosio oznaku verzije 3.5, a 2023. godine prešao je na verziju 4.0. ChatGPT piše tekst samostalno i semantički točno i time je prikladan za razne konverzijske aplikacije kao što su *chatbotovi*, prijevodi, virtualni pomoćnici (Hajdarović, 2023). Lingvist Noam Chomsky izjavio je kako „ChatGPT potkopava obrazovanje jer se zapravo radi o visoko tehnološkom plagiranju“ (Stewart, 2023; citirano prema Hajdarović, 2023). Kako bi se ljudi naviknuli na ovakav tip UI potrebno je učiti o prednostima i opasnostima koje isti donosi i koristiti ga s oprezom. U tome mogu pomoći različiti oblici pismenosti, ponajprije medijska, digitalna, algoritamska i

računala, ali i pismenost u području UI o kojoj će biti riječi nešto kasnije. ChatGPT zbog lakoće korištenja predstavlja određene izazove u obrazovanju jer, primjerice, omogućuje lakše plagiranje. Postoje razni načini kako to spriječiti, a jedan od njih je blokiranje korištenja ChatGPT-ja na školskim računalima, kao što to čine neke škole u New Yorku, ili provoditi ispitivanja na tradicionalan način, na papiru, umjesto na računalu, kao što čine neke škole u Australiji. Alati poput ChatGPT-ja uče uz pomoć algoritama i neprestanog analiziranja podataka pomoću kojih donose vlastite zaključke i time oponašaju rad ljudskog mozga (Mandić, 2023). ChatGPT su testirali brojni znanstvenici i stručnjaci u industriji. Tako je podatkovni znanstvenik, Ivan Židov testirao ChatGPT u kontekstu polaganja državne mature iz hrvatskog jezika na kojemu je model postigao rezultat od 65% točnosti u dijelu književnosti i jezika te 62,5% iz pisanja eseja (Pavelić, 2023; parafrazirano prema Hajdarović, 2023). Kako je engleski jezik najzastupljeniji, neki veći problemi pojavljuju se pri korištenju alata na drugim jezicima. Recimo, ako mu se postavi pitanje o posljedicama Drugog svjetskog rata u mjestu Špičkovina, ChatGPT odgovara da nema podatke o postavljenom pitanju jer ne pronalazi informacije o traženom mjestu budući da je manje poznato ili je o njemu manje zabilježenih informacija. Dok na pitanja koja su općenitija, a vezana su za Drugi svjetski rat, uspješno odgovara (Hajdarović, 2023). Preciznost ChatGPT-ja ovisi o podacima koji su „nahranili“ algoritam, odnosno o setu podataka koji su korišteni za treniranje i učenje algoritama (Hitchins, 2023 parafrazirano prema Hajdarović, 2023). Uz ChatGPT, aplikacije Duolingo, Quizlet i Grammarly također koriste sustav UI. Aplikacija Grammarly ispravlja englesku gramatiku te integrira internetski preglednik i Office paket. Tako ima mogućnost pomaganja u raznim svakodnevnim situacijama, poput odgovaranja na elektroničku poštu umjesto samog korisnika (Hajdarović, 2023). Predviđanja za ChatGPT uključuju korištenje veće memorije, multimodalnost i veću računalnu moć što bi ga učinilo još bližim konceptu jake ili opće UI (Jarić Dauenhauer, 2023; parafrazirano prema Hajdarović, 2023).

3.1.3. *Deep fake* tehnologija

Izraz *deep fake* obuhvaća koncept je koji se pojavio u prosincu 2017. godine. Naziv je nastao kao složenica pojmova „deep learning” zbog korištenja tehnologije koja se temelji na dubokom učenju i „fake” zbog upućivanja na lažan, neautentičan sadržaj (Risse, 2023). To je lažni medijski sadržaj koji se stvara pomoću tehnologije UI, odnosno zamjenjuje sliku ili videozapis jedne osobe s videozapisom ili slikom druge osobe i time stvara neautentičan sadržaj. *Deep fake* tehnologija koristi se za zamjenu ljudskih lica u lažnim vijestima, financijskim prevarama,

lažnim prikazima slavnih osoba i slično. Navedene situacije često navode vlade da zabranjuju korištenje takve tehnologije.

Tri najrizičnija načina primjene algoritama za zamjenu lica su: 1) *face-swap*, zamjena, odnosno preklapanje jednog lica drugim, 2) *lipsync*, tehnika u kojoj se mijenja dio lica osobe kako bi se činilo da izgovara riječi koje u stvarnosti ne izgovara, 3) *puppet-master*, tehnika u kojoj se u realnom vremenu lice jedne osobe animira preklopljenim licem druge osobe, to jest osobe koja je ispred kamere.

Prvi pokušaj primjene *deep fake* tehnologije bila je aplikacija „FakeApp“. Provedena su razna istraživanja o takvim pokušajima prevare zbog brzog širenja i manipulacije prouzrokovane lažnim videozapisima. Predložena su brojna rješenja za ovaj problem. Pokrenuto je natjecanje za pronalazak *deep fake* videozapisa u suradnji s META-om, Microsoft-om i AWS-om u okviru Odbora za vođenje integriteta medija za umjetnu inteligenciju i akademsku zajednicu (AI's Media Integrity Steering Committee and academics (DFDC)). Cilj te suradnje bio je uvjeriti znanstvenike širom svijeta u nužnost kreiranja tehnika i alata za uspješno otkrivanje *deep fake* videozapisa te alate za kontroliranje i provjeru medijskih sadržaja. Provedena je eksperimentalna analiza u kojoj se koristila hibridna konvolucijska neuronska mreža. Identificirale su se *deep fake* slike koristeći znak izvora kako bi se istaknule nepravilnosti u generiranim slikama. Cjelokupna analiza temeljila se na teoriji da slike koje imaju izvorne karakteristike mogu biti zaštićene. Model se pokazao dobrim te se dokazalo da se primjenom hibridne konvolucijske neuronske mreže može prepoznati je li isječak istinit ili lažan. Pokazano je da je moguće prepoznati lažiranje videozapisa te da daljnjim napretkom tehnologije predloženi model može biti još bolji i detaljniji. Nažalost, širenje *deep fake* videozapisa postalo je sve češće. Duboke laži koje se nalaze u takvom sadržaju mogu biti zastrašujuće za opću populaciju pa se otkrivanje takvih video sadržaja smatra iznimno važnim. Važno je razviti tehnologije koje mogu prepoznati krivotvorene sadržaje i spriječiti njihovo širenje (Ikram i sur., 2023). *Deep fake* videozapisi postaju sve stvarniji, a *deep fake* aplikacije su sve dostupnije. Profesorica Claire Wardle, pak, smatra da se stvara panika koja je nepotrebna i pretjerana te navodi kako su situacije poput ovih postojale i prije. Odnosno da su i prije postojali programi za obradu fotografija i videozapisa kojima su se mogli kreirati neautentični sadržaji. Claire Wardle isto tako objašnjava kako se manipulirati ljudima i širiti laži može na puno načina, ne samo uz pomoć UI (Dokler, 2019). Uz sve naprednije aplikacije i stručnjaci teško procjenjuju radi li se o stvarnom ili lažnom sadržaju. Primjerice, *deep fake* videozapis može reproducirati Hitlera kako pjeva „Druže Tito mi ti se kunemo“ pri čemu je sve usklađeno

vremenom i prostorom kada su snimani Hitlerovi govori, te usklađeno s Hitlerovim izvornim glasom jer alati UI za reprodukciju i generiranje audio sadržaja mogu lako kopirati i prilagoditi bilo čiji glas u drugom situacijskom kontekstu (Hajdarović, 2023). Jedan primjer pozitivnog korištenja *deep fake* tehnologije jest korištenje iste u sklopu muzeja Dali na Floridi, gdje su kao dio izložbe „Dali Lives“ kreirali *deep fake* umjetnika Dalija uz pomoć stroja koji je učio o njemu gledajući njegove intervjuje tisuću sati. Ponekad *deep fake* tehnologija može pomoći prenijeti snažne i važne poruke publici s većim i boljim učinkom. Tako je 2019. godine britanska dobrotvorna zdravstvena organizacija pomoću *deep fake* tehnologije kreirala poruku u kojoj nogometaš Davida Beckham govori o važnosti borbe protiv malarije na čak 9 jezika. Manje sofisticirani koncepti *deep fake* tehnologije nazivaju se *cheap fake*. Takvi videozapisi uređeni su bez strojnog učenja, često kreirani u Photoshopu. Bave se audiovizualnim manipulacijama, korištenjem dvojnika, usporavanjem ili ubrzavanjem video materijala i drugim sličnim manipulacijama video sadržaja (Risse, 2023).

Pronalaženje i otkrivanje *deep fake* sadržaja može se metaforički opisati kao igra mačke i miša. Ta igra odvija se između generatora ili kreatora *deep fake*-ova i detektora koji su stvoreni da ih prepoznaju i otkriju. Najčešći čin koji otkrije da je video sadržaj lažan jest treptanje, odnosno proces sklapanja očnih kapaka. *Deep fake* videozapisi ne mogu reproducirati prirodno treptanje. *Deep fake* tehnologija donosi promjene koje će imati i pozitivne i negativne posljedice za društvo. Potrebno je puno promišljanja i uspostavljanje regulatornih mehanizama i etičkih okvira kako se epistemološka prava i pravda ne bi prekršila. Na makro razini postoji velik rizik opasnosti za demokraciju, a na mikro razini postoji rizik nanošenja moralne i materijalne štete i povrede dostojanstva i ugleda pojedinca. Radi se, dakle, o vrlo izazovnoj tehnologiji UI pa je potrebno učiniti sve moguće kako bi prednosti *deep fake* tehnologije nadjačale, pa možda čak i nadomjestile njezine nedostatke (Risse, 2023).

3.1.4. Umjetna inteligencija i robotika

Riječ robot dolazi od češkog izraza „robota“, što znači ponavljajući i naporan rad. Prvi put izraz se pojavljuje u već spomenutoj noveli Karela Čapeka *Rossumovi univerzalni roboti* iz 1920. godine (Frankish i Ramsey, 2014). Stare civilizacije poput Rimljana sumnjale su u mogućnost razvoja umjetnih ili mehaničkih ljudi (Wilks, 2019). Grey Walter, neurolog iz Bristol, 1949. godine razvio je nekoliko strojeva koje je nazvao kornjače. To su bili prvi ikad mobilni autonomni roboti s kotačima. Walter je htio pokazati kako iz jednostavnog (umjetnog) živčanog sustava, a imali su dva umjetna neurona, može nastati interakcija. Kasnije ih je dalje razvijao te je uz pomoć svog tehničara Mr. W. J. „Bunny“ Warren-a dizajnirao šest novih kornjača koje

su imale sposobnost fototaksije, uočavanja svjetlosnog izvora koja im je služila kako bi se vratile na mjesto gdje bi se punile kada bi im oslabila baterija. Nazvao ih je „Machina speculatrix“. 1950-ih godina u fokusu robotike bila je praktičnost. Jedan od glavnih zadataka bio je postići da robot pomakne ruku na predodređenu točku, što se postiglo usavršavanjem motora koji je kontrolirao pokretanje zglobova robotske ruke. Industrijski roboti bili su programirani da ponavljaju zadanu radnju, odnosno da vrše unaprijed programiranu aktivnost. Roboti se nisu mogli slobodno autonomno kretati i djelovati. Stoga se kasnije razvija novi pristup koji će im to omogućiti poznat kao inteligentna mobilna robotika, u koju se integrira i tehnologija UI. Od 1966. do 1972. godine, SRI International proveo je istraživanje na mobilnom robotskom sustavu pod nazivom Shakey. Taj robot imao je mogućnost vida, što mu je dalo mogućnost percepcije okoline. Mogao je tražiti rute i razmjestiti jednostavne objekte u neposrednoj okolini. Shakey je bio primjer rane robotske tehnologije koju je pokretala UI (Franksih i Ramsey, 2014).

Jedna od prvih poveznica UI i robotike jesu već spomenuta tri zakona robotike Isaaca Asimova. Asimov je napisao povijesni serijal pod nazivom *The Foundation* (1951) u kojemu ljudi koloniziraju svemir i ondje grade carstvo. Kasnije je pisao priče u kojima uključuje robote kao (glavne) likove. Ljudi u tim pričama postaju slabiji, dok roboti postaju jačima i imaju sposobnost učenja. Tek naknadno, roboti shvaćaju da inicijativa ipak treba ostati na ljudima i povlače se iz društva kako bi ljudi u potpunosti mogli doživjeti napredak. Asimov tumači kako je neočekivano ponašanje robota zapravo greška čovjeka koji ga je proizveo, odnosno da bi roboti trebali uvijek raditi prema instrukcijama čovjeka. Asimov je ljude percipirao kao bespomoćne, ali kreativne i spremne na promjenu i rast (Morioka, 2023).

Postoje različite definicije robota. Bekey navodi kako su roboti strojevi koji se nalaze u svijetu koji osjeća, razmišlja i radi. Murphy robote definira više u odnosu na UI te ih opisuje kao mehanička stvorenja koja funkcioniraju autonomno (Bhaumik, 2018). Umjetno inteligentni roboti namijenjeni su prvenstveno istraživanju okoline uz poseban zadatak ili ulogu. Primjer su istražiteljski roboti, spasiteljski roboti i slično.

Razdoblja čovječanstva metaforički su podijeljena na četiri doba. Prvo doba bilo je doba pismenosti i vatre, zatim Drugo doba, doba poljoprivrede i gradova, Treće doba kao doba pisma i kotača te Četvrto doba, doba robota i UI. Kroz razvoj UI i moć robotike, strojevima dajemo mogućnost interakcije s fizičkim svijetom. Sve se više koristimo računalima i robotima kako bismo izrazili naše vlastite ljudske misli, osjećaje i kako bismo djelovali. Pitanje koje se u tom kontekstu pojavljuje je "Mogu li strojevi misliti?" (Reese, 2018). Roboti često prikazujemo

antropomorfizirano, odnosno nalik ljudima (tzv. roboti humanoidi – androidi i geminoidi – koji oponašaju ljudsku anatomiju i estetiku), no roboti mogu biti raznih oblika. Recimo, kada je američka vojska naručila robote koji nose ruksake vojnika, najuspješniji su bili oni koji su se sastojali od šest nogu i nalikovali insektima, podsjećajući malo na magarca ili konja koji nose prtljagu. Današnji najpoznatiji robot toga tipa je Spot, robot-pas tvrtke Boston Dynamics. Naravno, postoje i roboti humanoidi poput Atlasa kojeg je također kreirala tvrtka Boston Dynamics, a može se penjati po stepenicama, svladavati različite prepreke i kretati kao ljudi (Wilks, 2019).

Posebna grana robotike je društvena robotika, tj. interdisciplinarno područje koje se bavi razvojem robota koji uz fizičko utjelovljenje (tijelo robota) i autonomno kretanje i djelovanje mogu izazvati emocionalnu reakciju kod čovjeka. Radi se o robotima koji su tehnološki razvijeni, ali imaju razvijene i međuljudske vještine. U tome im je pomogala tehnologija UI kao i različiti senzori. Cilj je bio koristiti ih u uslužnim djelatnostima, obrazovanju, u domaćinstvima (pomoć u kući) i za brigu i skrb o starijima i nemoćnima (Breazeal, 2002). Iako je područje društvene robotike puno obećavalo, trenutno robotika u kombinaciji s tehnologijama UI najviše transformira industriju. U kontekstu Četvrte industrijske revolucije i trendova automatizacije i robotizacije radni roboti (eng. *work robots*) koriste se u industriji, posebno u području logistike, te u uslužnim djelatnostima (WEF, 2023). Amazon tako koristi robote u distribuciji u skladištima. Razvili su 55 000 Kiva robota. Prednost toga je da roboti mogu nositi teže predmete, locirati proizvode i pakirati predmete u kutije bez pomoći ljudi. Zanimljivost je da za vrijeme učenja, znanje pohranjuju u virtualni oblak i time automatski pomažu drugim strojevima, odnosno dijele naučeno s ostalim robotima (West, 2018).

3.2. Primjena umjetne inteligencije

Primjena UI sve je češća i sveobuhvatnija. Globalni forum politike (Global Policy Forum) koji promiče opće dobro društva objavio je 2021. godine priručnik s analizom stanja primjene UI te načelima i politikama UI. Priručnik se bavi primjenom UI na nacionalnoj razini. Opisuje ulaganje u UI, istraživanje i razvoj UI, stvaranje digitalnog ekosustava za UI, oblikovanje i omogućavanje politika za UI te izgradnju ljudskog kapaciteta i pripremu za tržišnu transformaciju (OECD, 2021). UI će kratkoročno pružiti evolucijsku korist, a dugoročno ona je i revolucionarna tehnologija. Iako pruža ekonomsku i gospodarsku vrijednost, većina njezine vrijednosti trenutno nije izravno uočljiva. Naime, unatoč tomu što postoji puno mogućnosti za primjenu UI na tržištu rada, javlja se pitanje njezine koristi za poslovni sektor i industriju.

Produktivnost Europe godišnje raste tek za 0.5% što je vrlo malo u odnosu na trenutne mogućnosti i dostupnu suvremenu tehnologiju. No, povećana produktivnost ujedno znači i da manje radnika obavlja istu količinu posla, stoga povećanje produktivnosti može donijeti gubitak radnih mjesta. Primjena UI u poslovanju nosi i prednosti i nedostatke, primjera je mnogo, a u idućim poglavljima opisat će se odabrani primjeri primjene UI u društvu 21. stoljeća (Davenport, 2021).

3.2.1. Umjetna inteligencija, automatizacija i tržište rada

Društvo danas ima sve više povjerenja u UI, dodjeljujući joj sve više funkcija na individualnoj i kolektivnoj razini. GPS sustav s integriranom UI upravlja automobilima, burze ovise o UI kako bi donijele odluke koje utječu na nacionalnu i međunarodnu ekonomiju, industrija i razvoj sve više ovise o integraciji sustava UI. Zašto se UI čini tako pouzdanom? Za razliku od ljudskog uma koji je podložan promjenama pod utjecajem impulsa, refleksa, emocija i koji se neprestano razmješta između prošlosti (refleksija) i budućnosti (planiranje), UI je uvijek 'prisutna' i objektivna. Benasayag (2021) objašnjava kako su nepredvidljivost prošlosti i budućnosti te slučajna priroda sadašnjosti karakteristični za čovjeka i njegovo djelovanje. Dok je UI, u usporedbi s njim, u tom smislu potpuno pouzdana. Zbog toga je čovječanstvo delegiralo svoje funkcije moćnijim strojevima. Tako danas alati i sustavi UI analiziraju veliki broj podataka i vrše predviđanja u različitim industrijama.

Pitanje utjecaja UI na tržište rada posebno je aktualno u kontekstu rasprava o Četvrtoj industrijskoj revoluciji i trendova automatizacije i robotizacije. Tehnološke inovacije poput UI, interneta stvari, robotike i dr. transformativno su i disruptivno utjecale na tržište rada, poslove, proizvodnju, ali i na procese upravljanja i organizacije. Uz automatizaciju, tehnologija UI, smatra se jednom od četiri ključna trenda koja utječu na sadašnjost i oblikuju budućnost rada (Cribb i Glover, 2018; WEF, 2016). West (2018) smatra da će UI, uz robotiku i internet stvari, redefinirati koncept rada što će utjecati ne samo na tržište rada i poslove, već i na nužnost ekonomske, političke, obrazovne i društvene reforme koja, između ostalog, podrazumijeva poticanje cjeloživotnog učenja, kreiranje novih društvenih ugovora i redefiniranje poreznog sustava i sustava socijalne zaštite. Pritom, treba imati na umu da utjecaj UI i automatizacije nije ravnomjerno raspoređen, kako zemljopisno, tako i u pogledu industrijskih područja. Također, utjecaj može biti izravan, posebno za razvijene zemlje, i neizravan, ponajviše za zemlje u razvoju. Iako mnogi trenutno predviđaju porast nezaposlenosti, neki znanstvenici i ekonomisti smatraju da će, gledajući dugoročno, navedeni trendovi ipak kreirati nove poslove (WEF, 2016).

Prvi primjer primjene UI koji će se opisati dolazi iz područja industrije video igara. U pitanju je duboka neuronska mreža koja predviđa neplanirane obrasce ljudskog ponašanja, nazvana Libratus. Razvilo ju je sveučilište Carnegie Mellon, a koristi se za igranje pokera. Poznato je da poker uključuje statističko predviđanje i element blefiranja ponašanja kao strategiju. Upravo to ponašanje, uključeno je u analizu neuronske mreže Libratus te se uspješno koristilo za predviđanje ljudskog ponašanja u igri i blefiranja. Algoritmi kojima se mreža služi nisu specificirani za poker ni igre. UI nije naučena određenim strategijama, već samostalno smišlja najbolji način za igranje na temelju prikupljenih podataka, u ovom slučaju pravila za igranje pokera. Libratus se tako može koristiti u bilo kojoj situaciji koja zahtjeva odgovor na temelju nepotpunih informacija. Koristeći duboko učenje Libratus je uložio otprilike 15 milijuna sati računalnih podataka za uvježbavanje kako bi izoštrio svoje strategije. Duboka neuronska mreža izračunala je popis strategija koje su se mogle koristiti u igri. U obzir su uzete i moguće greške koje protivnici mogu napraviti te slabosti strategije UI, kako bi se uspješno kreiralo rješenje za svaki mogući problem. Libratus je uspio izazvati četiri najbolja svjetska igrača pokera u kasinu u Pennsylvaniji. Pobjedio je s prednošću od više od 1,7 milijuna žetona i pobjedom na 20-dnevnom poker turniru.

Idući primjer primjene UI je virtualni asistent koji odgovara na upite ljudi. NTT Resonant tvrtka kreirala je sustav nazvan Oshi-el koji odgovara na upite ljudi, poput virtualnog savjetnika. Duboka neuronska mreža trenirala je svoj algoritam koristeći 190 000 pitanja i 770 000 odgovora s foruma tvrtke Oshiete goo. Na temelju toga, razvili su generičku strukturu odgovora uz mogućnost izražavanja suosjećanja, predlaganja rješenja problema i pokoji komentar potpore. Oshi-el UI bira rečenice iz baze podataka temeljene na riječima koje su korištene u postavljenom pitanju. Iako Oshi-el daje opširan odgovor, uporaba je ograničena jer korišteni sustav UI ima samo površno razumijevanje, tj. nema potpuno razumijevanje (Skilton i Hovsepian, 2018).

U razdoblju od 15. prosinca 2022. do 30. siječnja 2023. provedeno je istraživanje o primjeni UI u sveučilišnim knjižnicama u Ujedinjenom Kraljevstvu i Kini. Uzorak je obuhvatio 25 sveučilišta iz Ujedinjenog Kraljevstva i 25 sveučilišta iz Kine. Istraživanje je provedeno putem tražilica Bing.com i Baidu.com. Autori su prikupili dokumente knjižnica koji opisuju buduće razvojne planove, strateške planove, vizije i druge planove te su analizirali njihov sadržaj kako bi utvrdili u kojoj se mjeri u navedenim dokumentima spominje tehnologija UI. Utvrđeno je da je UI spomenuta u strategijama 24 sveučilišta, od kojih su 3 iz Ujedinjenog Kraljevstva, a 21 iz Kine. 5 sveučilišta imalo je cilj izgraditi pametnu knjižnicu u sklopu vizije

ustanove. 18 knjižnica integriralo je UI u svoje usluge, 2 u Ujedinjenom Kraljevstvu i 16 u Kini. Razlika između Kine i Ujedinjenog Kraljevstva je da Kina koristi inteligentne sustave i usluge poput *chatbotova* i virtualnih asistenata, dok ih Ujedinjeno Kraljevstvo još uvijek ne koristi. Primjena UI u knjižnicama većinom je jednostavne prirode, u obliku virtualnih asistenata. Sigurno je da knjižnice proteklih godina istražuju načine primjene UI u svom adu (Huang i sur., 2023). 2018. godine Sveučilište Rhode Island otvorilo je prvi na svijetu knjižnični laboratorij temeljen na tehnologiji UI. Zaposlenike čine upravitelj laboratorija i stručnjak za strojno učenje. Studenti imaju priliku raditi u laboratoriju i razvijati strojeve uz pomoć robotike i programiranja. Uz razvoj vještina studenata, laboratorij je namijenjen i istraživanju društvenih i kulturnih pitanja vezanih za razvoj UI. Knjižnica organizira različite događaje kako bi se nastavnici i osoblje upoznali s tehnologijom UI iz različitih perspektiva. Kreiraju nastavne programe i organiziraju javne programe. Sudjeluju u javnim raspravama o prednostima i izazovima UI. Financijsku podršku laboratoriju, uz Sveučilište, pruža i fond Champlin Foundation (Hervieux i Wheatley, 2022). Što se tiče utjecaja primjene UI u sveučilišnim i drugim knjižnicama, moguće ih je identificirati nekoliko. Prvi značajan utjecaj je poboljšano pretraživanje informacija. Algoritmi osnaženi umjetnom inteligencijom poboljšavaju učinkovitost i točnost pretraživanja informacija. UI sustavi, analizirajući podatke i ponašanje korisnika, generiraju relevantne rezultate pretraživanja, time pomažu studentima i znanstvenicima pronaći literaturu brzo i učinkovito. Uz to, još jedan od utjecaja je mogućnost kreiranja personaliziranih preporuka. Odnosno, UI algoritmi analiziraju preferencije korisnika pomoću povijesti pretraživanja i korištenja i posudbe građe te tako kreiraju personalizirane preporuke knjiga, članaka, radova i slično. Time korisnici mogu pronaći literaturu, koju možda samostalno ne bi uspjeli naći. Zatim, tu je inteligentno upravljanje izvorima pomoću obrade prirodnog jezika što omogućuje lakši pronalazak digitalizirane građe. Katalozi omogućuju pristup literaturi i široj javnosti. Već spomenuti virtualni asistenti i *chatbotovi* su također primjer primjene UI u knjižnicama. Oni pružaju pomoć korisnicima knjižnice, odgovaraju na upite korisnika, pomažu im u pronalasku literature, upućuju korisnike na izvore i brinu da korisnici dobiju svu potrebnu pomoć kada im se obrate. Još jedna primjena UI u knjižnicama je analiza podataka za donošenje odluka. Analiza uzoraka korištenja građe dopušta algoritmima UI da pomognu knjižnicama optimizirati zbirke, organizirati proračun ili rasporede usluga na učinkovitiji način kako bi se zadovoljile potrebe korisnika. UI također olakšava razmjenu znanja i suradnju među knjižnicama i grupama korisnika. Primjena tehnologije UI u knjižnicama otvara brojne mogućnosti za njihov razvoj, kreiranje novih i unaprijeđenje postojećih službi i usluga, te joj pomaže učvrstiti ulogu koju ima u akademskoj zajednici (Herrlich, 2023).

3.2.2. Umjetna inteligencija, urbanizacija i promet

Računalna znanstvenica Marti Hearst u poglavlju knjige *What to Think About Machines that Think* opisuje automatizaciju svijeta i predviđa njegovu budućnost koju naziva eGaia. U tom imaginarnom svijetu Hearst predočava kako će izgleda potpuna automatizacija društva, urbanizacija, i promjene koje donosi primjena tehnologije UI. Odnosno prikazuje sliku svijeta koji je moguć nastaviti se dosadašnji razvoj i primjena UI. Hearst objašnjava kako taj svijet uz elektroničke senzore i inteligentne sustave organizira individualne potrebe, automatizira čišćenje prostorija, promet, nadzor i proizvodnju. Urbani prostori su već danas ispunjeni sensorima, a njihov će broj u budućnosti samo rasti. Primjerice, u Kopenhagenu svjetla na cestama uključuju se samo kad detektiraju da netko vozi bicikl. Sustav se može nadograditi tako da obavještava odgovorne kada treba posoliti zaleđenu cestu ili ukloniti smeće, prijavljuje opasnosti na ulici i slično. eGaia djelomično je prisutna u svim razvijenijim dijelovima svijeta (Brockman, 2015).

Kao što spomenuta knjiga kaže, Četvrta industrijska revolucija tek počinje. Napredna urbanizacija u sprezi s tehnologijama UI sve se više širi. Jedan od primjera primjene UI u urbanom planiranju može biti projektiranje zgrade. Za projektiranje se koristi algoritam strojnog učenja koji integrira ideje iz prirode i matematike te stvara inovativan dizajn koji čovjek ne bi mogao osmisliti. Takav rezultat naziva se generativnom komponentom. Za takav način projektiranja koristi se inženjerski softver za dizajn Autodesk Dreamcatcher. Takozvani generativni dizajn uspoređuje se sa strojnim učenjem, jer njegovu srž čini sustav koji ima mogućnost dizajniranja zgrada primjenom složenih algoritamskih izračuna. Primjer građevine, izgrađene uz pomoć generativnog dizajna je Nacionalni stadion u Pekingu.

Uz primjer UI u arhitekturi, postoje razni primjeri primjene UI u prometu čiji je cilj povećati sigurnost u prometu i unaprijediti prometni sustav. Prvobitna asocijacija na UI i promet su pametna vozila. Pametna vozila koriste inteligentne sustave koji pomažu vozaču u vožnji ili ga zamjenjuju u vožnji, ovisno o stupnju autonomnosti. Ključni čimbenik je inteligentni transportni sustav (eng. *Intelligent Transport System*, (ITS)). Postoje automobili koji su povezani telekomunikacijama i mogu međusobno komunicirati. Primjer takvog vozila je Tesla 6.0 gdje GPS uređaj pamti rutu vožnje kako bi idući put na istoj ruti poboljšao iskustvo vožnje, a prikupljene podatke može podijeliti i s drugim vozilima. Također, Tesla 8.1 uspjela je integrirati autopilot funkcije koje novim značajkama poboljšavaju postojeće funkcije samostalne vožnje (Skilton i Hovsepian, 2018). Za realizaciju pametnih autonomnih vozila prva

je bila zaslužna vojna industrija, dok danas veliki broj uspješnih automobilskih tvrtki razvija autonomna vozila različite kategorije.

Kada je riječ o automatizaciji, bitno je spomenuti izraz internet stvari (eng. *Internet of Things* (IoT)), novu tehnologiju koja umrežuje i povezuje uređaje putem interneta. Radi se o mreži elektroničkih uređaja koji su putem senzora i odgovarajućeg softvera povezani s internetom i čine mrežnu infrastrukturu u kojoj povezani uređaji komuniciraju međusobno ili s odabranim aplikacijama. U sprezi s tehnologijama UI, internet stvari omogućuje napredne usluge upravljanja i kontrole, ali i uspostavljanja sigurnosti. Takvi se sustavi najčešće koriste u pametnim domovima te za ekonomično upravljanje resursima (Prister, 2019).

Video nadzor koristi se od davnih 1970-ih. Kamere su postojale na javnim mjestima i zgradama. Video analitika krovni je pojam za tehnike koje analiziraju i prikupljaju podatke iz određenog video zapisa. Glavni cilj tih tehnika je identificirati relevantnu informaciju. Video analitika često uključuje tehnologiju prepoznavanja lica, Tehnologija, odnosno sustavi za prepoznavanje lica služe za identifikaciju ljudskih lica prikazanih u video ili slikovnim okvirima te uspoređivanje istih s onima iz baze podataka lica. Takav softver očitava geometriju određenog lica i istu uspoređuje sa značajkama lica iz baze podataka. Ukoliko se dogodi podudaranje promatranog lica s primjerom iz baze znači da je sustav prepoznao lice. Postoji mogućnost da lica u bazi podataka sadrže i dodatni zapis s osobnim informacijama pojedine osobe. Te informacije svakako mogu pomoći u daljnjoj analizi ukoliko je potrebna. Dvije glavne primjene prepoznavanja lica su ovjera ili autentifikacija te nadzor. Ovjera je utvrđivanje identiteta osobe, a koristi se za pristup nekom prostoru ili usluzi. Može biti u obliku snimanja lica kamerom. Ovjera lica koristi se i u imigraciji, glasovanju, zahtjevima za vozačku dozvolu i slično. Može se koristiti i u privatne svrhe, primjerice za otključavanje pametnog telefona. Za razliku od postupka ovjere odnosno autentifikacije, koji je namijenjen prepoznavanju pojedinca, postupak nadzora služi za skeniranje grupe ljudi, iz daljine, pomoću javnih kamera, gdje postoji mogućnost da ljudi nisu uvijek svjesni da ih se nadzire. Nadzor se provodi iz nekoliko razloga. Može se provoditi kako bi se onemogućio pristup pojedincima koji predstavljaju potencijalni sigurnosni rizik. Primjerice, trgovine mogu zabraniti pristup nekim osobama zbog prethodnih krađa. Još jedan razlog za provođenje nadzora je pronalazak pojedinaca od strane policije, a u pitanju mogu biti kriminalci, osumnjičeni teroristi i slično. Također, nadzor služi i za identifikaciju pojedinaca i praćenja njihova ponašanja kako bi se na osnovu dobivenih podataka i procjene rizika odredilo odgovarajuće postupanje. Korištenje nadzora provodi se kada je u pitanju masovni nadzor, recimo nadziranje građana u kineskom

sustavu društvenog kreditiranja, gdje se promatra i ocjenjuje ponašanje građana u javnom prostoru, što se kasnije odražava na njihov društveni status. Sustavi za autentifikaciju koriste kamere na relativnu blizinu, dok sustavi za nadzor koriste kamere i na daljinu pa samim time sustavi za nadzor mogu biti manje pouzdani (Waelen i Brey, n.d.).

3.2.3. Umjetna inteligencija i kreativni rad

Veza UI i umjetnosti seže u 20. stoljeće, kada je umjetnik Nam June Paik eksperimentirao s različitim tehnologijama, surađivao s televizijama, stvarao umjetnički sadržaj emitiranjem uživo putem satelita, promicao međukulturalni kapacitet. Suvremeni umjetnici nastavili su koristiti tehnologiju UI, internet i digitalnu tehnologiju u svom radu.

Razni suvremeni umjetnici svojim radom doprinose raspravama i promišljanjima o UI, posebice u kontekstu digitalne umjetnosti i korištenja digitalnih alata i alata UI za kreativni umjetnički rad. Različite su razine i načini povezivanja UI i umjetnosti. Umjetnica Hito Steyerl koja se bavi filmom izdvaja kako su očekivanja UI precijenjena i kako ona ima zavaravajući učinak. Upotrebljava izraz umjetna glupost (engl. *artificial stupidity*), te kaže kako programeri mogu stvarati slike uz pomoć UI, no da bi se one uistinu interpretirale i doživjele potrebno je mišljenje stručnjaka iz tog područja umjetnosti (Hito Steyerl; citirano prema Brockman, 2019, 228).

Umjetnik Paul Klee često je govorio kako je umjetnost stvaranje nevidljivoga vidljivim. U računalnim znanostima poznato je kako algoritmi rade nevidljivi dio posla, odnosno pozadinski posao. No, postoje primjeri vidljivih algoritama strojnog učenja koji procesuiraju podatke i vidljivi su u aplikacijama poput Google's DeepDream aplikacije, koja vizualizira proces računalnog prepoznavanja uzoraka u stvarnom vremenu. Aplikacija pokazuje kako algoritam usklađuje oblike životinja sa zadanim unosom. Steyerl tvrdi kako mogu postojati apstraktni računalni uzorci koji podsjećaju na slike poznatih umjetnika, no da oni ne predstavljaju stvarno umjetničko djelo. Radi se suvremenom znanstvenom i tehničkom viđenju umjetnosti, no sa stajališta povijesti umjetnosti računalno generirane slike lišene su estetske tradicije i prave umjetničke vrijednosti (Steyerl; citirano prema Brockman, 2019, 229-230).

Umjetnica Rachel Rose, bavi se stvaranjem videozapisa i u sklopu toga koristi tehnologiju UI. Pomoću UI slojevito manipulira zvukom i slikom i time proces montaže videozapisa postaje najvažniji aspekt njezinog rada. Ipak, umjetničko donošenje odluka je za nju proces koji je povezan s empatijom ljudskog bića, a računala to ne mogu zamijeniti (Brockman, 2019).

DALL-E je danas možda najpoznatiji primjer sustava UI koji se povezuje s kreativnim radom i utječe na rad umjetnika i onih u kreativnoj industriji. Sustav koji omogućuje pretvaranje ideja u slike. Radi se o sustavu koji uz pomoć tekstualne naredbe može generirati različite slikovne, tj. vizualne prikaze. Trenutno je aktivna verzija DALL-E 3 koja je stvorena na temelju ChatGPT-a i modela CLIP (Contrastive Language Image Pre-training). Za razliku od prošlih verzija, ova verzija omogućuje da stvorene slike budu u punom vlasništvu osobe koja ih je kreirala, zabranjuje se bilo kakav zahtjev za kreiranje negativnog sadržaja, a sigurnost je prioritet kako bi se smanjio rizik širenja mržnje. Pomoću alata za klasifikaciju, moguće je razaznati koja je slika kreirana od strane UI, a koja od strane čovjeka. Vršiti se i tzv. kontrola kreativnosti, u kojoj se odbijaju zahtjevi za stvaranjem slike u stilu živućeg umjetnika (OpenAI, 2023). Slični sustavi za kreiranje slika pomoću predloženog teksta su Adobe Firefly i Midjourney. Adobe Firefly je alat generativne UI koji je dio Photoshopa i omogućuje visokokvalitetnu manipulaciju fotografijama. Osim toga, nudi mogućnost tzv. generativnog popunjavanja slike, odnosno mijenjanja, proširivanja i dodavanja novih elemenata fotografiji. Midjourney je također alat UI koji koristi kanal Discord i kreira slike na temelju tekstualnih naredbi, a ima i mogućnost generiranja tekstualnog opisa iz ponuđene slike. Alat Midjourney je revolucionarizirao područje digitalne fotografije u sprezi s tehnologijom UI. Očekuje se da će navedeni i slični alati imati značajan disruptivni učinak na kreativnu industriju, posebno na područje fotografije, ilustracije i grafičkog dizajna, ali i da će potaknuti stvaranje novih oblika umjetničkog i kreativnog izražaja.

3.3. Izazovi u razvoju umjetne inteligencije

UNESCO je 2021. godine predložio dokument *Preporuka o etici umjetne inteligencije* (eng. *Recommendation on the Ethics of Artificial Intelligence*) koji je usvojen 23. studenog 2021. godine, a 2022. godine objavljena je publikacija koja opisuje ciljeve i načela, donosi preporuke za primjenu etičkih načela i preporuke za praćenje i evaluaciju, upućuje na i promovira korištenje trenutnih preporuka te donosi konačne, tj. zaključne odredbe. UNESCO prepoznaje dinamične pozitivne i negativne utjecaje UI na društvo, okoliš, ekosustav, ljudske živote i umove, ljudske odnose i interakciju, proces donošenja odluka, obrazovanje, humanističke, društvene i prirodne znanosti, kulturu, komunikaciju i informacije. Svjestan je da se sve države suočavaju sa širenjem informacijskih i komunikacijskih tehnologija i tehnologija UI. Također prepoznaje potrebu za informacijskom pismenošću kao i različite društvene izazove razvoja i primjene UI, poput njezine etičke primjene i upravljanja, poštivanja i zaštite kulturne

raznolikosti. Također, svjestan je da pojedini izazovi mogu narušiti lokalne i regionalne etičke norme i vrijednosti.

Sustavi UI otvaraju nova etička pitanja koja uključuju pitanje donošenja odluka, zapošljavanje, pristup informacijama, zaštitu osobnih podataka, zaštitu potrošača, zakonska prava, sigurnost, ljudska prava, temeljne slobode, privatnost i tako dalje. Etički izazovi sve su prisutniji zbog algoritama UI koji mogu reproducirati postojeće pristranosti i time pogoršati postojeće oblike diskriminacije, stereotipa i predrasuda. Neki izazovi povezuju se sa sposobnošću sustava UI da obavljaju zadatke koje su prije obavljali samo ljudi. Zbog toga mogu ugroziti ljudski osjećaj iskustva i agencije, samospoznaje, autonomije, vrijednosti, dostojanstva i interakcije. Transparentnost je iznimno važna pri korištenju UI, a manjak iste negativno utječe na izazovne odluke koje donosi UI sustav te uzrokuje moguće kršenje prava. Zbog toga je potrebno ograničiti područja primjene takvih sustava (UNESCO, 2022). Organization for Economic Cooperation and Development (OECD) nalaže načela UI za odgovorno upravljanje pouzdanim UI sustavima, a to su: 1) uključivi rast, održivi razvoj i dobrobit, 2) humanocentrične vrijednosti i pravednost, 3) transparentnost i objašnjivost, 4) postojanost, zaštita i sigurnost te 5) odgovornost (OECD, 2021, 6). Humanocentrične vrijednosti i pravednost važni su za pitanje ljudskih prava, demokraciju i zakonska pravila koja trebaju biti uključena u sustave UI, dozvoljavajući ljudsku intervenciju kroz sigurnosne mehanizme. Kada su u pitanju sigurnost i zaštita, sustavi UI moraju podržavati i primjenjivati sistematske pristupe u upravljanju rizicima kako bi se umanjili sigurnosti rizici (OECD, 2021).

UI ponekad ima tendenciju pojačati već postojeću pristranost utemeljenu na interesima i pretpostavkama njezinih tvoraca. Strojno učenje kao takvo radi na temelju obrazaca, odnosno podacima za učenje, ali ti podaci mogu izostaviti određene društvene skupine ili biti ograničeni na jedan kulturni kontekst. Stoga, stvarni podaci za učenje trebaju poštivati načela uključivosti, jednakosti, nepristranosti i pravednosti. Uklanjanje informacija o rasi ili spolu ne rješava problem nejednakosti. UI sustavi ponekad izostavljaju društvene skupine, pojačavaju pristranost svojih podataka za učenje i to ponekad rezultira rasizmom ili diskriminacijom. Stoga su znanstvenici odlučili matematički mjeriti pravednost, ali shvatili su da je često nemoguće na taj način obuhvatiti baš sve aspekte pravednosti. Uloženo je puno napora da se stvori istinski pravedna UI, no puno je prepreka na tom putu. Prije tzv. automatizacije pravednosti trebali bi se procijeniti sustavi i definirati vrste pravednosti te identificirati što točno dovodi do društvene nepravde. Slijedom toga kreirali bi se alternativni pristupi i pravedniji, inovativniji sustavi (Desautels Faculty of Management, McGill University, 2023).

Razvojem UI sustava dolazi do izazova i po pitanju kibersigurnosti. Taj se problem pogoršava na globalnoj razini, posebno za velike tvrtke i vladine organizacije. Glavno rješenje takvih problema je osiguranje pomoću sustava koji se bave zaštitom privatnosti kao što su IDS (Instruction Detection System) i IPS (Instruction Prevention System). S druge strane, postoje sigurnosni mehanizmi kao što su vatrozidi, antivirusni sustavi te operativne kontrole i sigurnosni postupci koje tvrtke provode. Svi navedeni alati mogu integrirati tehnologiju UI i tako omogućiti uspješniju identifikaciju ekstremnih situacija koje mogu značiti rizik za organizaciju (Botto-Tobar i sur., 2022).

Velik broj etičkih izazova dolazi iz područja privatnosti i zaštite podataka. Sustavi UI rade s velikim podacima, što automatski uključuje korištenje osobnih podataka. Sustavi UI se koriste i za spomenuti nadzor na javnim mjestima, a građani često nisu svjesni da se podaci prikupljaju. Etičko korištenje UI zahtjeva da se podaci prikupljaju i dijele na način da se poštuje privatnost pojedinca te da pojedinci znaju što se radi s prikupljenim podacima i u koje svrhe se koriste. Brojne tvrtke osjećaju potrebu za razvojem i implementacijom etičke UI. Na primjer, Microsoft predlaže načela za bolju UI: pravednost, pouzdanost i sigurnost, privatnost i zaštita, uključenost, transparentnost i odgovornost. Program Digitalna Europa (2018) koji simbolizira digitalnu industriju u Europi, tvrdi da je pravni okvir spreman za izazove UI, uključujući pristranost i diskriminaciju te da je važno osigurati transparentnost, objašnjivost i interpretaciju. Ljudi trebaju razumjeti kada i kako se algoritmi koriste u donošenju odluka. Obje strane (vlade i tvrtke) se zalažu za slične pristupe, a konačan cilj im je održavati UI društveno korisnom i poštivati etička načela i vrijednosti kao što su sigurnost, transparentnost, odgovornost, usklađivanje vrijednosti, privatnost i ljudska kontrola (Coeckelbergh, 2020). Kada je riječ o pristupima razvoju tehnologije UI, često se spominje ljudski usmjereni pristup ili humanocentrični pristup. U pitanju je perspektiva koja podržava ljudsku samoučinkovitost, odgovornost i društvene veze. Što je veći naglasak na ljudski usmjerenoj UI, smanjit će se rizik egzistencijalne prijetnje UI i povećati korist društva u poslovanju, obrazovanju, zdravstvu, sigurnosti zajednice i zaštiti okoliša. Humanocentrična UI pokazuje kako stvarati uspješne tehnologije koje osnažuju ljudsku aktivnost i djelovanje. Stoga je važno osigurati kontrolu i nadzor nad autonomnim sustavima (Schneiderman, 2022).

Vezano za pitanje etike UI i etičkog postupanja prema UI, često se spominju moralno pravo i moralni status UI. Filozof Kenneth Einar Himma tumači kako je ideja o moralnoj agenciji povezana sa sposobnosti činjenja i djelovanja. Odnosno, X je agent samo ako je X sposoban djelovati. On kaže kako je djelo čin, ali svaki čin nije djelovanje. Recimo disanje je

čin, ali ne predstavlja djelovanje. Zbog toga, svi koji obavljaju djelovanje jesu moralni agenti. Moralna agencija podrazumijeva neku namjeru (Gunkel, 2012). Pitanje moralnog prava i moralnog statusa UI načešće se promatra u kontekstu umjetne superinteligencije koja može zauzeti položaj moralnog agenta ili položaj moralnog pacijenta. Naime, s jedne se strane postavlja pitanje hoće li razvoj umjetne superinteligencije i porast autonomnosti UI dovesti do jačanje njezine pozicije kao moralnog agenta (moralnog djelatelja), odnosno hoće li ona moći moralno razmišljati, prosuđivati i odlučivati. U tom pogledu postoje različita stajališta, ali i problem definiranja samog pojma moralnosti. S druge strane, u slučaju da UI zauzme položaj moralnog pacijenta (moralnog podnositelja) postavlja se pitanje trebamo li se onda prema UI odnositi ipak samo kao prema stroju ili ona zaslužuje neka prava. Određena istraživanja su pokazala da ljudi već danas empatiziraju s robotima (Suzuki i sur. 2015; Darling, Nandy i Breazeal, 2015), no postoje i indirektni argumenti u obranu moralnog ponašanja prema UI kao moralnom pacijentu temeljeni na etičkom argumentu vrline, odnosno Kantovom argumentu protiv ubijanja pasa koji kaže bi loše ponašanje spram UI štetilo našem moralnom karakteru. Osim toga, argument uzima u obzir i mogućnost da UI u budućnosti razvije određenu intrinzičnu vrijednost i svijest. Ipak, pitanje moralnog prava i moralnog statusa UI trenutno se ne smatra toliko važnim u odnosu na aktualne etičke izazove odgovornosti, transparentnosti, pristranosti, i dr. (Coeckelbergh, 2020).

Zaključno, što se tiče etičkih izazova UI, kao korisnici različitih tehnologija danas, svjesni smo brzog razvoja UI i problema koje može uzrokovati. UI može filtrirati situacijsku svijest ili kontekst što dodatno utječe na percepciju, razumijevanje, odlučivanje i djelovanje. Kao društvo, važno je da naučimo kako se nositi s različitim izazovnim situacijama, kako moralno, etično i odgovorno upravljati sustavima UI i kako donositi relevantne zakone i propise kako bismo se uspješno suočili s potencijalnim rizicima i opasnostima (Liu, 2023).

3.3.1. Društveno odgovorni razvoj i primjena umjetne inteligencije

Trenutno ne postoji jedan, univerzalan zakon ili pravilnik koji regulira razvoj i korištenje UI. Unatoč svim prednostima i mogućnostima koje UI nudi, postoje brojni razlozi za uspostavu zakonskih okvira i mehanizama koji će regulirati razvoj i integraciju tehnologije UI u društvo na odgovoran način. Jednako tako, postoje i razlozi protiv regulacije UI.

Prvi razlog protiv regulacije UI je da bi time došlo do gušenja inovacija i napretka, odnosno usporio bi se razvoj UI, usporilo bi se ostvarivanje financijske dobiti, ograničio bi se ulazak novih tvrtki za razvoj UI na postojeće tržište, a ojačao bi se položaj postojećih tvrtki

čime bi one mogle ostvariti svojevrsni monopol. Drugi razlog protiv regulacije je njezina kompleksna i izazovna implementacija. Smatra se da bi regulacija te vrste bila previše nejasna ili općenita da bi bila sustavno primjenjiva i provedena., Bila bi nužna međunarodna suradnja i koordinacija, a to je prilično složeno i teško za postići s obzirom na kulturne razlike i razlike u pravnim sustavima. Jedinствена regulacija na međunarodnoj razini je gotovo nemoguća, stoga je jasno da će Sjedinjene Američke Države imati vlastite zakone, kao i Europska unija (EU), Kina i ostatak svijeta. Treći razlog zašto ne regulirati UI je pretpostavka da bi došlo do prekomjerne regulacije i neželjenih posljedica. Tehnološki napredak je uvijek brz i neprestano dolazi do rizika te je regulacija novih tehnologija izazov za vlasti, posebno kada se u obzir uzme činjenica da se tehnologija razvija puno brže nego što pravni sustav svojim zakonima i okvirima može obuhvatiti te promjene. Potrebno je razjasniti opseg postojećih regulacija i osigurati da nove industrije ne budu njima ograničene.

U odnosu na navedene rizike, postoji mnogo koristi od regulacije tehnologije UI (Book, 2023). Prvi razlog je osiguravanje etičke upotrebe UI. Prvenstveno se misli na zaštitu privatnosti korisnika i sigurnost. Regulacija može pomoći razviti povjerenje, transparentnost i odgovornost među korisnicima generativne UI. Drugi razlog je čuvanje ljudskih prava i sigurnosti. Rizici su olakšano širenje dezinformacija i laži (Book, 2023). Potrebno je koncentrirati se na privatnost podataka jer suprotno može dovesti do ozbiljnih izazova za pojedince i društvo u cjelini. Zbog toga je potrebno kreirati sustave UI koji su usklađeni s ljudskim vrijednostima i moralom. Treći razlog zašto je potrebno regulirati generativnu UI je ublažavanje društvenog i ekonomskog utjecaja UI. Na primjer, UI utječe na veću automatizaciju, a to dovodi do veće nezaposlenosti. Predviđa se da bi do 2030. godine UI mogla zamijeniti 800 milijuna radnih mjesta što je 30% globalne radne snage. Važno je štititi svjetske ekonomije od monopola potaknutih tehnologijama UI. U budućnosti rad na regulaciji UI bit će zajednički i interdisciplinaran, te će biti neophodno redovito pratiti, vrednovati, mijenjati i prilagođavati sustav (Book, 2023).

Najbliže zakonu i regulaciji UI na području Europske unije jest *Akt Europske Unije o umjetnoj inteligenciji* (2021), kao prvi službeni dokument koji regulira područje UI na području EU i koji je Europska komisija prvi put objavila 21. travnja 2021. godine. Akt je dostupan na poveznici <https://artificialintelligenceact.com/>. U pitanju je još uvijek radna verzija akta, a posljednja inačica dokumenta (Final draft 2024) dostupna je na poveznici <https://artificialintelligenceact.eu/the-act/>. EU želi regulirati UI i omogućiti bolje uvjete za razvoj inovativne tehnologije. Uz to, predloženo je analiziranje i klasifikacija sustava UI zbog

rizika koji postoje za korisnike. Nakon što se pravila odobre, to će biti prva pravila o UI. Ciljevi Parlamenta u zakonodavstvu o umjetnoj inteligenciji nastoje osigurati da svi sustavi UI u EU budu transparentni i nediskriminirajući. To se posebno odnosi na sustave za nadzor i praćenje. Uz navedeno, želja Parlamenta je postaviti tehnološki neutralnu definiciju UI koja može biti primjenjiva na buduće sustave UI. Akt o umjetnoj inteligenciji donosit će pravila za različite razine rizika. Utvrdit će se obveze korisnika ovisno o razini rizika. Sustavi UI donose minimalan rizik, no potrebno ih je dobro procijeniti. Svaki rizik koji može biti prijetnja ljudima je neprihvatljiv i svi rizični sustavi bit će zabranjeni. Takav rizik može biti u obliku kognitivno-bihevioralnog manipuliranja osobama, kao što su igračke koje glasovno potiču djecu na opasno ponašanje, zatim bodovanja i klasifikacije ljudi na temelju ponašanja ili njihovog socioekonomskog statusa, te u obliku sustava biometrijske identifikacije poput prepoznavanja lica. Kod ovih rizika mogu postojati iznimke, ako je u pitanju na primjer odobrenje suda za kazneni progon počinitelja nekog kaznenog djela. U tom slučaju dopustila bi se biometrijska identifikacija sa zakašnjenjem. Uz neprihvatljive rizike, postoje i visoki rizici (Akt, 2023). Visokorizični sustavi su oni koji negativno utječu na sigurnost i temeljna prava. Takvi sustavi bit će podijeljeni u dvije kategorije: sustavi koji se upotrebljavaju u proizvodima koje obuhvaća zakonodavstvo EU o sigurnosti proizvoda, kao što su igračke, automobili ili medicinski uređaji, te sustavi iz osam posebnih područja koji će morati biti registrirani u bazi podataka EU. Tih osam posebnih područja su: 1) biometrijska identifikacija i kategorizacija fizičkih osoba, 2) upravljanje kritičnom infrastrukturom i njezin rad, 3) obrazovanje i strukovno osposobljavanje, 4) zapošljavanje, upravljanje radnicima i pristup samozapošljavanju, 5) pristup i korištenje osnovnih privatnih usluga i javnih usluga, 6) upravljanje migracijama, azilom i nadzor granica, 7) provedba zakona te 8) pomoć u pravnom tumačenju i primjeni zakona (Akt, 2023). Nabrojani visokorizični sustavi UI zahtijevaju procjenu prije stavljanja na tržište te tijekom životnog ciklusa na tržištu. Kada je u pitanju generativna UI, potrebno je ispuniti nekoliko zahtjeva kako bi ona ostala transparentna. Primjerice, označiti da je sadržaj kreiran uz pomoć tehnologije UI, navesti koji su podaci zaštićeni autorskim pravom te koristiti modele koji sprječavaju stvaranje nezakonitog sadržaja. Uz prethodno navedene rizike, postoji i ograničeni rizik, odnosno sustavi s ograničenim rizikom. Kako bi se ispunila spomenuta transparentnost, potrebno je omogućiti korisnicima donošenje informiranih odluka. Korisnici moraju biti svjesni da surađuju s UI i moraju imati pravo odlučiti hoće li nastaviti interakciju s pojedinim sustavom ili ne. Primjer takvih generativnih sustava je ChatGPT. Cilj je što prije dovršiti dogovor na Vijeću EU i realizirati ovaj Akt (Akt, 2023).

Društveno odgovorni razvoj UI uključuje kreiranje odgovorne, pouzdane i objašnjive UI. Objašnjiva UI zapravo znači da je moguće razumjeti UI sustave i pozadinske procese, odnosno kako takvi sustavu funkcioniraju. Primjer metode za razumijevanje UI je LIME (Local Interpretable Model-agnostic Explanations) u kojoj se predviđa algoritam crne kutije tako što se imitira uz pomoć drugog transparentnog modela. Transparentni model reproducira UI sustav i identificira njegove značajke (Schmidpeter i Altenburger, 2023). Manjak objašnjive UI povećava potrebu za razvojem odgovorne UI. Kako bi se postigla odgovorna UI potrebno je zadovoljiti nekoliko uvjeta. Naime, potrebno je uspostaviti načela UI, definirati način i proces korištenja UI, razumjeti kako ona utječe na društvene vrijednosti, uspostaviti odgovoran okvir upravljanja alatima i sustavima UI, operacionalizirati okvir upravljanja odnosno imati tim podrške koji snosi odgovornost za UI neke organizacije, podučavati o etičkim načelima UI unutar organizacije na konzistentan način te provoditi ispitivanja i evaluacije putem upitnika kako bi se omogućilo pravilno razumijevanje (Firth-Butterfield i Vogel, 2022). Pouzdana UI je UI kojoj ljudi mogu vjerovati. Ona podržava etička načela kao što su dostojanstvo i poštivanje ljudskih prava (Coeckelbergh, 2020).

Kada se govori o odgovornom razvoju UI važno je spomenuti i *Bijelu knjigu o umjetnoj inteligenciji*. UI se u *Bijeloj knjizi* definira kao skup tehnologija koje se koriste podacima, algoritmima i računalnom snagom (Europska komisija, 2020, 2). Europa ima potencijal za postati globalnim predvodnikom kada su u pitanju podatkovna gospodarstva i njihove promjene, odnosno razviti ekosustav za UI koji može služiti na dobrobit europskog društva i gospodarstva. Napredak poslovanja, zdravstvene skrbi, prometnih sustava i ostalo, samo su neke od prednosti koje takav ekosustav može donijeti. Europska UI nužno se mora temeljiti na vrijednostima građana te temeljnim ljudskim pravima, kao što su ljudsko dostojanstvo i zaštita privatnosti, vrijednosti kojima EU pridodaje puno pažnje. Europska komisija odabrala je sedam zahtjeva utvrđenih u Smjernicama stručne skupine na visokoj razini: 1) ljudsko djelovanje i nadzor, 2) tehnička stabilnost i sigurnost, 3) privatnost i upravljanje podacima, 4) transparentnost, 5) raznolikost, nediskriminacija i pravednost, 6) dobrobit društva i okoliša te 7) odgovornost. (Europska komisija, 2020, 9-10). Uz smjernice stručne skupine, izgradnji povjerenja potrošača UI pridonio bi regulatorni okvir koji bi bio usklađen s drugim mjerama za promicanje kapaciteta Europe u tom području (Europska komisija, 2020, 10).

3.4. Pismenost u području umjetne inteligencije

Društvo 21. stoljeća razlikuje različite oblike, odnosno vrste funkcionalne pismenosti – od pismenosti kao općenitog pojma koji podrazumijeva komunikaciju odabranim simboličkim

strukturama, preko informacijske, medijske i digitalne pismenosti, do računalne, algoritamske, podatkovne i pismenosti u području privatnosti (engl. *privacy literacy*). Poseban značaj danas pridaje se pojmu pismenosti u području UI koja se vezuje uz informacijsku i medijsku pismenost, odnosno računalnu i algoritamsku pismenost.

Temelj koncepta pismenosti za UI je tradicionalna pismenost, a uključuje razvoj i stjecanje sposobnosti potrebnih za doba UI utemeljeno na funkcionalnoj, društvenoj i tehničkoj pismenosti. Long i Magerko definiraju pismenost za UI kao individualnu sposobnost za kritičku evaluaciju, korištenje i komuniciranje s UI. Aoun (2017; citirano prema Yi, 2021) definira pismenost za UI kao sposobnost prepoznavanja i upotrebe UI razumijevanjem njezinih koncepta.

Konkretnije, pismenost u području UI odnosi se na vještine, znanje i kritičko razumijevanje koje omogućuje informiranu primjenu tehnologije i sustava UI. Također, podrazumijeva pravilno razumijevanje konteksta, dizajna, razvoja i primjene UI, kao i mogućnosti, prilika, rizika i štete koju njezino korištenje može izazvati. Radi se o složenom i multidisciplinarnom pojmu koji nadilazi granice računalne znanosti te uključuje tri glavne komponente:

1. tehničku – razumijevanje osnovnih načela funkcioniranja UI tehnologije i sustava,
2. praktičnu – uspješno korištenje UI tehnologije i sustava i razumijevanje njihove praktične primjene,
3. etičku – prepoznavanje širih društvenih implikacija primjene UI tehnologije i sustava i razumijevanje etičkih pitanja koja se mogu pojaviti (Crabtree, 2023; Europski parlament, 2023; Klein, 2023)

Cilj pismenosti o UI je educirati sve građane u suvremenom društvu kako bi objektivno percipirali i razumjeli tehnologiju koja postaje dio njihove svakodnevice. Svrha pismenosti za UI je anticipacija budućnosti društva na temelju osnovnih društvenih praksi (Yi, 2021). Primjerice, trenutno još uvijek postoji nedostatak kritičke javne svijesti o algoritamskom oblikovanju komunikacije zbog čega je potrebno opismenjavanje za UI koje će potaknuti razvoj kolektivne inteligencije (Vajzović i sur., 2021).

3.5. Budućnost razvoja i primjene umjetne inteligencije

Kada je riječ o budućnosti UI, predviđanja su mnoga i raznolika. Jedno od predviđanja odnosi se na izum kvantnih računala i njihov utjecaj na eksponencijalni razvoj UI. Prve ideje o

kvantnim računalima sežu u 1980. godinu kada je matematičar J. I. Manin došao do prve zamisli, a znanstvenik Richard Feynman godinu dana nakon započeo izradu računala koja se zasnivaju na zakonima kvantne mehanike. Današnja računala koriste bitove 1 ili 0, a kvantni bitovi su 1 ili 0, ali mogu biti i istovremeno i 1 i 0. Zbog toga, kvantna računala funkcije obavljaju puno brže. Google i NASA su na kvantnom računalu D-Wave 2X postigli brzinu veću 100 milijuna puta nego kod običnog računala. Kvantna računala mogu, primjerice, omogućiti puno jeftiniju proizvodnju lijekova ili prognozirati i spriječiti razvoj bolesti jer imaju mogućnost praćenja čestica u posebnim uvjetima (danas se to još uvijek odvija samo u posebnim akceleratorima) (Prister, 2019). Tu su i drugi tehnološki trendovi koji će oblikovati blisku budućnost UI, poput razvoja multimodalne UI, inteligentnih agenata i demokratizacije i *opensource*-anja UI (Craig, 2024).

Mnoga predviđanja odnose se na samu primjenu UI pa se tako često govori o budućoj primjeni tehnologije, tj. sustava UI u medicini i zdravstvu, posebice u kontekstu dijagnostike, liječenja i skrbi o pacijentima. Također, govori se o daljnjem unaprjeđivanju i širenju automatizacije, konvergenciji tehnologije UI i interneta stvari te trendu uvođenja autonomnih agenata u različite industrije i sektore (npr. robota i autonomnih vozila i letjelica). Mnogi vjeruju i da će tehnologija UI utjecati na pitanje okoliša, klimatskih promjena, bioraznolikosti i prirodne održivosti. U konačnici, neka predviđanja širi utjecaj UI vide u politici i očuvanju demokracije.

Kada je u pitanju predviđanje nešto dalje budućnosti UI, velik broj znanstvenika i stručnjaka je skeptičan i upozorava kako razvoj tehnologije može voditi do samouništenja čovjeka. Predviđaju kako će tijekom 2040-ih zaživjeti opća UI te neki znanstvenici izjavljuju bojazan zbog njezinog utjecaja na čovječanstvo. Neka predviđanja govore kako će već 30-ih godina ovog stoljeća neka računala biti u stanju simulirati ljudski mozak i njegove funkcije i imati veću brzinu obrade nego biološki mozak (Putica, 2018). Brojni autori predviđaju i komentiraju budućnost UI. Harari (2018; citirano prema Vertovšek i Knežević, 2020, 2517-2519) navodi kako treba biti oprezan s UI, posebno jer će ubrzano razvijati, a moguće opasnosti proizlaze ne toliko iz same UI, koliko iz zanemarivanja ljudske inteligencije (tzv. *upgrading machines and downgrading humans* fenomen¹). Aristeia Papadimitriou (2016; citirano prema Vertovšek i Knežević, 2020, 2517-2519) također komentira buduću svijesti izvan ljudskog uma. Predviđa pojavu novog intelekta o kojemu bi trebalo kritički raspravljati. S

¹Fenomen na koji je prije nekoliko godina upozoravao Tristan Harris (Thompson, 2019).

obzirom na to da je UI interdisciplinarno područje, ona predlaže interdisciplinarnu suradnju kako bi se stvorio etički kodeks i preduvjeti u kojima će UI biti na opću korist društva. Također ističe kako je jedan od najvećih izazova upravo pitanje "Što znači biti čovjek?". UI će pomoći čovjeku dosegnuti besmrtnost tijela i uma te razviti post-ljudsko društvo. Na razini super inteligencije potencijalno će se stvoriti preduvjeti za život izvan bioloških okvira. Inteligencija veća od ljudske podrazumijevala bi poboljšanu verziju bioloških mozgova. No, svaki pokušaj pokretanja programa eugenike suočio bi se s političkim i moralnim preprekama. Biomedicinska poboljšanja mogu pružiti veće prednosti, već sad postoje lijekovi koji poboljšavaju koncentraciju i energiju. Predviđa se kloniranje ljudi, nadarenih pojedinaca. Pristup tehnologiji kloniranja bi bio ograničen, ali značajan jer i mali porast nadarenih ljudi može imati značajan učinak na društvo i cjelini (Bostrom, 2014). Mnogo znanstvenika, pa tako i Eliezer Yudkowsky zalaže se za to da ako ikada ostvarimo super inteligenciju, trebamo nastojati kreirati prijateljsku UI (engl. *friendly AI*) onu čiji su ciljevi usklađeni s ljudskima (Yudkowsky, 2004; citirano prema Tegmark, 2017, 334). Ekonomist Erik Brynjolfsson ima optimističnu viziju tržišta rada i naziva ga „Digitalna Atena“ zbog toga što su građani antičke Atene imali lagodne živote jer su imali robove koji su obavljali sve za njih, pa tako on predlaže da ulogu nekadašnjih robova sada preuzmu roboti koje pokreće UI. Time bi se kod čovjeka eliminirao svakodnevni stres, a ujedno bi se potaknulo kreiranje mnoštva novih proizvoda i usluga. Kada su u pitanju poslovi i zapošljavanje, postoje optimisti i pesimisti u vidu budućnosti UI i njezina utjecaja na tržište rada. Neki optimisti smatraju da će fizičke i kognitivne poslove zamijeniti više kreativni poslovi, a pesimisti smatraju da je kreativnost isto mentalni, tj. kognitivni proces te da će i takve poslove potencijalno zamijeniti UI (Tegmark, 2017). Byron Reese (2018, 256-257), autor knjige *The Fourth Age* spominje izraz utopija koji opisuje nemoguće svjetove, svjetove koji su nam nedostižni. Tijekom 19. stoljeća postojala je literatura koja je prikazivala brojne primjere tadašnjih utopijskih svjetova i ideja, poput univerzalnog obrazovanja i pravne jednakosti muškaraca i žena. Te ideje se danas polako ostvaruju, a kada se one u potpunosti ostvare, društvo će kreirati nove ideje za još napredniji svijet sve dok ne shvatimo da ne možemo zamisliti bolji svijet od onoga u kojem živimo. Reese stoga predlaže da za opis takvog društva ne koristimo riječ „utopija“, koja više označuje mjesto koje ne postoji i koje je odraz puke želje društva za boljom budućnosti, već da uvedemo riječi i izraze koji ponajprije odražavaju aspiracije društva i njegovu ambicioznost i samouvjerenost da izgradi novi (bolji) svijet. Reese predlaže izraz „verutopija“ (eng. *verutopia*) koje opisuje mjesto budućnosti koje je stvarno i koje možemo zajednički izgraditi kao dobro i pravedno društvo koje uistinu predstavlja peto doba, tj. svijet pun mogućnosti za sve.

4. Zaključak

Rad se bavi pitanjem tehnologije UI u suvremenom društvu te, uz kratki prikaz povijesnog razvoja, problematizira trenutni razvoj i primjenu UI, kao i prednosti i izazove njezine implementacije. Temeljno problemsko pitanje rada jest uloga, odnosno utjecaj tehnologije UI na društvo 21. stoljeća. Pregledom odabrane literature i izvora nastojalo se odgovoriti na postavljena istraživačka pitanja koja su usmjerila koncept ovoga rada.

Prije svega, utjecaj tehnologije UI vidljiv je, posredno ili neposredno, u različitim sektorima i područjima čovjekova djelovanja, od gospodarstva, industrije, ekonomije i poslovanja, preko prometa, trgovine i uslužnih djelatnosti, do znanosti, zdravstva, obrazovanja, kulture, umjetnosti i zabave. U mnogim područjima UI je donijela brojne ekonomske i društvene dobrobiti te se nastoji smisleno integrirati kako bi optimizirala sustave i procese vodeći računa o pitanjima privatnosti, sigurnosti, transparentnosti, pristranosti, odgovornosti, i dr. Njezin se učinak najčešće opisuje disruptivnim, a pojedini trendovi, poput razvoja generativne UI, pametnih (inteligentnih) sustava, tzv. *deep fake* tehnologije i robotike imaju pozitivan učinak na društvo i građane kreirajući mnoge nove mogućnosti kao što su automatizacija brojnih procesa i aktivnosti i povećanje njihove učinkovitosti. No jednako tako imaju i rizičan učinak koji otvara brojna etička, pravna, ali i praktična pitanja kao što su pitanje privatnosti i sigurnosti te pitanje utjecaja UI na tržište rada i gubitak poslova. To svakako ukazuje na nužnost kritičkog pristupa i primjene uravnoteženih regulatornih mehanizama koji će omogućiti uspješno iskorištavanje svih prednosti i mogućnosti tehnologije UI, ali i smanjiti,

ograničiti ili ukloniti potencijalne prijetnje i rizike. Potrebno je naglasiti da su u radu prikazani tek odabrani primjeri primjene i utjecaja UI. Na kraju, u radu se naglašava važnost odgovornog razvoja i primjene UI te nužnost strateškog planiranja i razvoja etičkih (meki zakon) i zakonskih okvira (tvrđi zakon) koji će usmjeriti i regulirati odnos društva i UI. Uz primjenu (humanocentričnih) načela, smjernica i zakona, jednako je važno i opismenjavanje u području UI koje građane čini osviještenim i informiranim pojedincima koji znaju kritički i uspješno koristiti tehnologiju UI. Sve navedeno podrazumijeva interdisciplinarni i multidisciplinarni pristup, međunarodnu suradnju, ali i suradnju industrije, znanstvene zajednice, vlade i društvenih institucija i civilnog sektora. Jer je jedino tako moguće osigurati pozitivan, odnosno smanjiti, ograničiti i, ako je moguće, ukloniti negativan društveni utjecaj tehnologije UI, tj. na održiv i sigurna način maksimalno iskoristiti transformativni učinak UI.

U posljednje se vrijeme razvoj UI često opisuje nepredvidivim što svakako otežava praćenje i prije spomenuto strateško planiranje za odgovornu i humanocentričnu UI, ali i ukazuje na neophodnost daljnjeg kontinuiranog istraživanja, javnih rasprava i inicijativa koje adresiraju izazove i rizike UI, a koji će zajedno osigurati kako fleksibilan i demokratski pristup navedenoj problematici, tako i uravnoteženu regulaciju koja će nastaviti poticati kulturu inovacije bez ugrožavanja humanosti i ljudskih prava. Kako bi društvo moglo imati puno povjerenje u tehnologiju UI, posebno je važno graditi i koristiti transparentne i odgovorne sustave UI koji donose odluke i 'usmjeravaju' ljudsko ponašanje.

U konačnici, ovim se radom nastojalo ukazati na nužnost odmaka od generaliziranja i pojednostavlivanja pitanja tehnologije UI koji najčešće uzrokuju puku polarizaciju mišljenja, umjesto pravilno razumijevanje i kritički pristup, posebice u kontekstu društvene koristi. Stoga je važno proaktivno adresirati sva pitanja kroz javne rasprave, pronaći rješenja ulaganjem u istraživanje i razvoj te uspješno implementirati odabrana rješenja kreiranjem etičkih i zakonskih okvira, slijedenjem primjera dobre prakse i obrazovanjem za kritičko i odgovorno korištenje tehnologije UI. Jer napredak bilo koje tehnologije, pa tako i UI, ne bi trebao biti određen isključivo tehničkim uvjetima, pukom znanstvenom znatiželjom ili tržišnom konkurencijom i utrkom, već ponajprije potrebama i dobrobiti društva u cijelosti.

5.Literatura

1. *Akt EU-a o umjetnoj inteligenciji: prva regulacija tog područja.*(2023)
<https://www.europarl.europa.eu/news/hr/headlines/society/20230601STO93804/akt-eu-a-o-umjetnoj-inteligenciji-prva-regulacija-tog-podrucja>
2. Asemi, A., Ko, A., & Nowkarizi, M. (2020). *Intelligent libraries: a review on expert systems, artificial intelligence, and robot.* Emerald Publishing Limited.
3. Beg, M.J., & Verma, M. K. (2023). *The Mind and Machine: An Introduction to Artificial Intelligence: Exploring the Intersection of Artificial Intelligence, Human Cognition, Psychology and Robotics.*
4. Benasayag, M. (2021). *The Tyranny of Algorithms.*Europa Editions. (S. Rendall trans.)
5. Bhaumik, A. (2018). *From AI to Robotics: Mobile, Social, and Sentient Robots.* CRC Press.
6. Book, A. (2023). *Should AI be Regulated? The Arguments For and Against.*
<https://www.wearedevelopers.com/magazine/eu-ai-regulation-artificial-intelligence-regulations>
7. Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press.
8. Bostrom, N. i Yudkowsky, E. (2014). The Ethics of Artificial Intelligence. In K. Frankish i W. Ramsey (Ur.), *Cambridge Handbook of Artificial Intelligence* (str. 316-334). Cambridge University Press. <https://intelligence.org/files/EthicsofAI.pdf>
9. Botto-Tobar, M., Gomez, O. S., Rosero Miranda, R., Diaz Cadena, A., & Luna-Encalada, W.(2023). *Trends in Artificial Intelligence and Computer Engineering: Proceedings of ICAETT 2022.* Springer.
10. Breazeal, C. (2002). *Designing Sociable Robots.* MIT Press.

11. Brockman, J. (2019). *Possible Minds: Twenty-Five Ways of Looking at AI*. Penguin Press.
12. Brockman, J. (2015). *What to Think About Machines That Think: Today's Leading Thinkers on the Age of Machine Intelligence*. Harper Perennial.
13. Brynjolfsson, E., & McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company
14. Coeckelbergh, M. (2020). *AI Ethics*. The MIT Press.
15. Christian, B., & Griffiths, T. (2016). *Algorithms to Live By: The Computer Science of Human Decisions*. Henry Holt and Company.
16. Crabtree, M. (kolovoz 2023.). *What is AI Literacy? A Comprehensive Guide for Beginners*. Radar: The Analytics Edition. <https://www.datacamp.com/blog/what-is-ai-literacy-a-comprehensive-guide-for-beginners>
17. Craig, L. (4. siječnja 2024.). *10 top AI and machine learning trends for 2024*. TechTarget. <https://www.techtarget.com/searchenterpriseai/tip/9-top-AI-and-machine-learning-trends>
18. Cribb, J. i Glover, D. (2018). *Don't worry about the robots: how to survive and thrive in the world of work*. Allen & Unwin.
19. Darling, K.; Nandy, P.; i Breazeal, C. (2015.) Empathic Concern and the Effect of Stories in Human-Robot Interaction. U *2015 24th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE: 770–775.
20. Davenport, T. H. (2021). Prednost UI: kako iskoristiti revoluciju UI. Mate.
21. Desautels Faculty of Management, McGill University (2023). *Artificial Intelligence: Generative AI*. World Economic Forum. <https://intelligence.weforum.org/topics/a1Gb0000000pTDREA2/key-issues/a1G680000000Ne9EAE>
22. Desautels Faculty of Management, McGill University (2023). *Artificial Intelligence: Bias and Fairness in AI Algorithms*. World Economic Forum. <https://intelligence.weforum.org/topics/a1Gb0000000pTDREA2/key-issues/a1G0X0000057IWXUA2>
23. Dokler, A. (2019, 10. rujna). *Deepfake videozapisi: što kad ne možemo vjerovati onome što vidimo i čujemo?*. Medijska pismenost. <https://www.medijskapismenost.hr/deepfake-videozapisi-sto-kad-ne-mozemo-vjerovati-onome-sto-vidimo-i-cujemo/>
24. Leksikografski zavod Miroslav Krleža. (2013-2024). Ekspertni sustavi. Hrvatska enciklopedija. <https://www.enciklopedija.hr/clanak/ekspertni-sustavi>

25. Europska komisija. (2020). *Bijela knjiga o umjetnoj inteligenciji: Europski pristup izvrsnosti i izgradnji povjerenja*. URL: <https://op.europa.eu/hr/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1>
26. Europski parlament. (2023). *Amandmani koje je donio Europski parlament 14. lipnja 2023. o Prijedlogu uredbe Europskog parlamenta i Vijeća o utvrđivanju usklađenih pravila o umjetnoj inteligenciji (Akt o umjetnoj inteligenciji) i izmjeni određenih zakonodavnih akata Unije*. Europski parlament. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_HR.html
27. Firth-Butterfield, K. i Vogel, M. (2022, 5. srpnja). *5 ways to avoid artificial intelligence bias with 'responsible AI'*. World Economic Forum. <https://www.weforum.org/agenda/2022/07/5-governancetips-for-responsible-ai/>
28. Frank, M., Roehrig, P., i Pring, B. (2017). *What to Do When Machines Do Everything: How to Get Ahead in World of AI, Algorithms, Bots, and Big Data*. Wiley.
29. Frankish, K., i Ramsey, W. M. (2014). *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press.
30. Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. The MIT Press.
31. Haenlein, M., i Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present and Future of Artificial Intelligence. *California Management Review*, 61(4), 5-14. <https://doi.org/10.1177/0008125619864925>
32. Hajdarović, M. (2023). UI, ChatGPT i poučavanje Povijesti. *Poučavanje povijesti*, 2(1), 52-65. Preuzeto s <https://hrcak.srce.hr/303411>
33. Hawkins, J., i Blakeslee, S.(2005). *On Intelligence*. St. Martin's Griffin.
34. Herrlich, H. (2023, 23. svibnja). *The Future of Libraries: AI and Machine Learning*. Fordham Library News. <https://librarynews.blog.fordham.edu/2023/05/23/the-future-of-libraries-ai-and-machine-learning/>
35. Hervieux, S., & Wheatley, A. (2022). *The Rise of AI: Implications and Applications of Artificial Intelligence in Academic Libraries*. Association of College and Research Libraries.
36. History of Artificial Intelligence. (n.d.). Council of Europe. <https://www.coe.int/en/web/artificial-intelligence/history-of-ai>
37. Huang, Y., Cox, A. M. i Cox, Y. (2023). Artificial Intelligence in academic library strategy in the United Kingdom and the Mainland of China. *The Journal of Academic Librarianship*, 49(6):102772. <https://doi.org/10.1016/j.acalib.2023.102772>
38. Ikram, S.T., Chambial, S., Sood, D., V, P. i V, A. (2023). A Performance Enhancement of Deepfake Video Detection through the use of a Hybrid CNN Deep

- Learning Model. *International journal of electrical and computer englineering systems*, 14(2), 169-178. <https://doi.org/10.32985/ijeces.14.2.6>
39. Klein, A. (10. svibnja 2023). *AI Literacy, Explained*. EducationWeek. <https://www.edweek.org/technology/ai-literacy-explained/2023/05>
40. Leksikografski zavod Miroslav Krleža. (1992). *Inteligencija*. Medicinski leksikon: mrežno izdanje. <https://medicinski.lzmk.hr/clanak/6249>
41. Leksikografski zavod Miroslav Krleža. (2021). *UI*. Hrvatska enciklopedija. <https://www.enciklopedija.hr/natuknica.aspx?ID=63150>
42. Liu, W.(2023). *Integrated Human- Machine Intelligence: Beyond Artificial Intelligence*. Tsinghua University Press Limited.
43. Mainzer, K. (2006). Utjelovljeni um. O komputacijskim, evolucijskim i filozofskim tumačenjima poznaje. *Filozofska istraživanja*, 26(2), 405-421. <https://hrcak.srce.hr/5906>
44. Mandić, V. (2023, 19. siječnja). *Kako razvoj UI utječe na medije. Medijska pismenost – abeceda za 21. stoljeće*. Medijska pismenosr. <https://www.medijskapismenost.hr/kako-razvoj-umjetne-inteligencije-utjece-na-medije/>
45. McCarthy, J. (2007). *What is artificial intelligence?*. <https://www.diochnos.com/about/McCarthyWhatisAI.pdf>
46. McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, 27(4), 1.
47. Morioka, M. (2023). *Artificial Intelligence, Robots, and Philosophy*. Journal of Philosophy of Life.
48. OECD. (2021). *From principles to practice: tools for trustworthy AI: A Framework to Compare Implementation Tools for Trustworthy AI Systems*. OECD Digital Economy Papers, 312. OECD Publishing. <https://oecd.ai/en/tools-repor>
49. OECD. (2021). *State of implementation of the OECD AI Principles: Insights from national AI policies*. OECD Digital Economy Papers, 311. OECD Publishing. <https://doi.org/10.1787/1cd40c44-en>
50. OpenAI. (2023). Dall-e 3. <https://openai.com/dall-e-3>
51. Possati, L. M. (2023). *Unconscious Networks: Philosophy, Psychoanalysis, and Artificial Intelligence*. Routledge.
52. Prister, V. (2019). UI. *Mediji, kultura i odnosi s javnostima*, 10(1), 67-72. <https://doi.org/10.32914/mcpr.10.1.7>

53. Putica, M. (2018). UI: dvojbe suvremenog razvoja. *Hum*, 13(20), 198-213. Preuzeto s <https://hrcak.srce.hr/219733>
54. Reese, B. (2018). *The Fourth Age: Smart Robots, Conscious Computers and the Future of Humanity*. Atria Books.
55. Risse, M. (2023). *Political Theory of the Digital Age: Where Artificial Intelligence Might Take Us*. Cambridge University Press.
56. Rosenberg, R. S. (1988). *The Impact of Artificial Intelligence on Society*. University of British Columbia. <https://www.cs.ubc.ca/sites/default/files/tr/1988/TR-88-20.pdf>
57. Routley, N. (2023, 6. veljače). *What is generative AI? An AI explains*. World Economic Forum. <https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/>
58. Schmidpeter, R., i Altenburger, R. (2023). *Responsible Artificial Intelligence: Challenges for Sustainable Management*. Springer.
59. Sen, A. (28. ožujka 2023.). *The Impact of Artificial Intelligence on Society: Opportunities, Challenges, and Ethical Considerations*. LinkedIn. <https://www.linkedin.com/pulse/impact-artificial-intelligence-society-opportunities-challenges-sen>
60. Shadbolt, N. (2022). „From So Simple a Beginning”: Species of Artificial Intelligence. U Manyika, J. (ur.), *Daedalus: Journal of the American Academy of Arts & Sciences*. American Academy of Arts & Sciences. file:///home/mikica/Desktop/Daedalus_Sp22_AI-and-Society.pdf
61. Skilton, M., i Hovsepian, F. (2018). *The 4th Industrial Revolution: Responding to the Impact of Artificial Intelligence on Business*. Palgrave Macmillan
62. Shneiderman, B. (2022). *Human-Centered AI*. Oxford University Press.
63. Smith, C., McGuire, B., Huang, T., i Yang, G. (2006). The History of Artificial Intelligence. URL: <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>
64. Suzuki, Y.; Galli, L.; Ikeda, A.; Itakura, S. i Kitazaki, M. (2015.) Measuring Empathy for Human and Robot Hand Pain Using Electroencephalography. *Scientific Reports*, 5, 15924. <https://www.nature.com/articles/srep15924>
65. Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf.
66. Thompson, N. (23. travnja 2019). *Tristan Harris: Tech is 'Downgrading Humans.' It's Time to Fight Back*. Wired. <https://www.wired.com/story/tristan-harris-tech-is-downgrading-humans-time-to-fight-back/>
67. UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO.

68. Vajzović, E., Hilbert, M., Turčilo, L., Vučetić, V., i Silajdžić, L. (2021). *Medijska i informacijska pismenost: dizajn učenja za digitalno doba*. Fakultet političkih nauka Univerziteta u Sarajevu.
69. Valerjev, P. (2006). Povijest i perspektiva razvoja umjetne inteligencije u istraživanju uma. U Žebec, M.-S., Sabol, G., Šakić, M. i Kotrla Topić, M. *Mozak i um – trajni izazov čovjeku*. Institut društvenih znanosti Ivo Pilar, 107-121.
70. Vertovšek, N. i Knežević, I. G. (2020). Filozofija i svjesnost u budućnosti: kiborzi i UI u iščekivanju besmrtnosti. *In medias res*, 9,(16), 2511-2522.
<https://doi.org/10.46640/imr.9.16.3>
71. Waelen, R., i Brey, P. (2022). Ethical Dimensions of Facial Recognition and Video Analytics in Surveillance. In Michael Boylan i Wanda Teays (ur.). *Ethics in the AI, Technology, and Information Age*. Rowman& Littlefield, 217-231.
72. Wendell, W., i Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.
73. West, D. M. (2018). *The Future of Work: Robots, AI, and Automation*. Brookings Institution Press.
74. West, D. M. I Allen, J. R. (24. travnja 2018.). How artificial intelligence is transforming the world: research. Brookings. <https://www.brookings.edu/articles/how-artificial-intelligence-is-transforming-the-world/>
75. Wilks, Y. (2019). *Artificial Intelligence: Modern Magic or Dangerous Future?*. Hot Science.
76. World Economic Forum. (2016). *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution: Global Challenge Insight Report*. World Economic Forum. <https://www.weforum.org/publications/the-future-of-jobs-2016/>
77. World Economic Forum. (2023). *The Future of Jobs Report 2023: Insight Report*. World Economic Forum. <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>
78. Yi, Y. (2021). Uspostavljanje koncepta UI pismenosti: Fokusiranje na kompetenciju i svrhu. *Jahr: Europski časopis za bioetiku*, 12(2), 353-368.
<https://doi.org/10.21860/j.12.2.8>