

UNIVERZA V LJUBLJANI  
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

**ETIČNI VIDIKI ODLOČANJA S POMOČJO ALGORITMOV**

Ljubljana, junij 2022

TINA KOŽELJ

## IZJAVA O AVTORSTVU

Podpisana Tina Koželj, študentka Ekonomske fakultete Univerze v Ljubljani, avtorica predloženega dela z naslovom Etični vidiki odločanja s pomočjo algoritmov, pripravljena v sodelovanju s svetovalcem prof. dr. Alešem Popovičem

### IZJAVLJAM

1. da sem predloženo delo pripravila samostojno;
2. da je tiskana oblika predloženega dela istovetna njegovi elektronski obliki;
3. da je besedilo predloženega dela jezikovno korektno in tehnično pripravljeno v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani, kar pomeni, da sem poskrbela, da so dela in mnenja drugih avtorjev oziroma avtoric, ki jih uporabljam oziroma navajam v besedilu, citirana oziroma povzeta v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani;
4. da se zavedam, da je plagiatorstvo – predstavljanje tujih del (v pisni ali grafični obliki) kot mojih lastnih – kaznivo po Kazenskem zakoniku Republike Slovenije;
5. da se zavedam posledic, ki bi jih na osnovi predloženega dela dokazano plagiatorstvo lahko predstavljalo za moj status na Ekonomski fakulteti Univerze v Ljubljani v skladu z relevantnim pravilnikom;
6. da sem pridobila vsa potrebna dovoljenja za uporabo podatkov in avtorskih del v predloženem delu in jih v njem jasno označila;
7. da sem pri pripravi predloženega dela ravnal/-a v skladu z etičnimi načeli in, kjer je to potrebno, za raziskavo pridobila soglasje etične komisije;
8. da soglašam, da se elektronska oblika predloženega dela uporabi za preverjanje podobnosti vsebine z drugimi deli s programsko opremo za preverjanje podobnosti vsebine, ki je povezana s študijskim informacijskim sistemom članice;
9. da na Univerzo v Ljubljani neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravico shranitve predloženega dela v elektronski obliki, pravico reproduciranja ter pravico dajanja predloženega dela na voljo javnosti na svetovnem spletu preko Repozitorija Univerze v Ljubljani;
10. da hkrati z objavo predloženega dela dovoljujem objavo svojih osebnih podatkov, ki so navedeni v njem in v tej izjavi.

V Ljubljani, dne \_\_\_\_\_

Podpis študentke: \_\_\_\_\_

# KAZALO

<b>UVOD .....</b>	<b>1</b>
<b>1 UMETNA INTELIGENCA.....</b>	<b>3</b>
<b>1.1 Učenje umetne inteligence.....</b>	<b>4</b>
1.1.1 Strojno učenje.....	5
1.1.2 Globoko učenje .....	6
<b>1.2 Glavni dejavniki delovanja umetne inteligence .....</b>	<b>7</b>
1.2.1 Kakovost podatkov.....	7
1.2.2 Odločitve ustvarjalcev umetne inteligence .....	8
1.2.3 Kompleksnost interakcije z zunanjim okoljem .....	9
<b>2 ALGORITMI UMETNE INTELIGENCE.....</b>	<b>9</b>
<b>2.1 Prednosti algoritmov .....</b>	<b>12</b>
2.1.1 Hitrejši postopek izkazovanja identitete .....	13
2.1.2 Avtonomna vozila .....	15
2.1.3 Varnost .....	17
2.1.4 Zdravstvo.....	18
<b>2.2 Negotovosti in izzivi povezani z algoritmi.....</b>	<b>20</b>
2.2.1 Nasprotujoča si stališča .....	22
2.2.2 Izguba človeške presoje .....	23
2.2.3 Pristranskost in diskriminacija .....	25
2.2.4 Zlorabe .....	26
<b>2.3 Družbeni izzivi.....</b>	<b>27</b>
2.3.1 Potreba po večji transparentnosti .....	28
2.3.2 Potreba po rasti algoritmične izobraženosti .....	29
2.3.3 Potreba po večjem nadzoru .....	30
<b>3 VLOGA ETIKE IN ETIČNI VIDIKI ODLOČANJA Z ALGORITMI.....</b>	<b>33</b>
<b>4 TVEGANJA IN ETIČNE SMERNICE UPORABE UMETNE INTELIGENCE.....</b>	<b>37</b>
<b>4.1 Tveganja za temeljne pravice in varstvo potrošnikov .....</b>	<b>38</b>
<b>4.2 Tveganja za varnost proizvodov in odgovornost .....</b>	<b>40</b>
<b>4.3 Etične smernice Evropske komisije o uporabi umetne inteligence .....</b>	<b>42</b>
<b>4.4 Univerzalne smernice o umetni inteligenci .....</b>	<b>45</b>

<b>5 RAZISKAVA ETIČNIH VIDIKOV ODLOČANJA S POMOČJO ALGORITMOV .....</b>	<b>47</b>
5.1 Zasnova raziskave.....	47
5.2 Analiza .....	47
5.3 Ključne ugotovitve.....	53
5.4 Priporočila za nadaljnje raziskovanje .....	58
<b>SKLEP.....</b>	<b>59</b>
<b>LITERATURA IN VIRI.....</b>	<b>60</b>
<b>PRILOGE .....</b>	<b>71</b>

## KAZALO SLIK

Slika 1: Razlika v delovanju strojnega in globokega učenja.....	5
Slika 2: Tipi napak pri odločanju z algoritmi.....	22
Slika 3: Etična in moralna vprašanja razvoja in izvajanja umetne inteligence .....	37
Slika 4: Smernice kot okvir za zaupanja vredno umetno inteligenco .....	43
Slika 5: Zavedanje o uporabi algoritmov pri zaposlitvenem postopku .....	48
Slika 6: Zavedanje o uporabi algoritmov pri izračunu kazenskega tveganja .....	49
Slika 7: Zavedanje o uporabi algoritmov pri izračunu ocene osebnih financ .....	50
Slika 8: Zavedanje o uporabi algoritmov pri hitrejšem odkrivanju bolezenskih stanj.....	51
Slika 9: Zavedanje o uporabi algoritmov pri avtonomnih vozilih .....	52

## KAZALO PRILOG

Priloga 1: Tabele .....	1
Priloga 2: Anketni vprašalnik.....	11

## SEZNAM KRATIC

**ACM** – (angl. The Association for Computing Machinery)  
angl. – angleško  
**ANNs** – (angl. Artificial neural network); umetne nevrnske mreže  
**CEO** – (angl. Chief Executive Officer); glavni izvršni direktor  
**CRM** – (angl. Customer relationship management); upravljanje odnosov s strankami  
**EU** – (angl. European Union); Evropska unija  
**GDPR** – (angl. General Data Protection Regulation); Splošna uredba o varstvu podatkov  
**HITL** – (angl. Human in the loop)

**HITLFE** – (angl. Human in the loop for exceptions)

**HOTL** – (angl. Human on the loop)

**HOOTL** – (angl. Human out of the loop)

**HRIS** – (angl. Human resource management system); informacijski sistem upravljanja s človeškimi viri

**NIST** – Nacionalni inštitut za standarde in tehnologijo

**SAE** – (angl. Society of Automobile Engineers)

**UGAI** – (angl. Universal Guidelines for Artificial Intelligence); univerzalne smernice o umetni inteligenci

**ZDA** – Združene države Amerike



## UVOD

V magistrskem delu sem obravnavala vse večjo problematiko vključenosti algoritmov v obliki podpore procesov odločanja. Tako se z njimi srečujemo v vsakdanjem življenju. Njihov vpliv na odločitve ljudi je lahko znatno, saj imajo pomembno vlogo, na primer pri zaposlitvenem procesu, kazenskem tveganju in izračunu ocene osebnih financ (Castelluccia & Le Métayer, 2019).

Zagovorniki uporabe algoritmov za odločanje vidijo v njih priložnosti, med drugim z vidika večje varnosti, saj lahko predstavniki oblasti s pomočjo prepoznavanja obrazov odkrijejo v množici ljudi osumljence kaznivih dejanj in s tem preprečujejo kriminal ter terorizem (Kayser-Bril, 2019b). Omogočen je tudi hitrejši postopek dokazovanja identitete, ki ga na primer vpeljujejo banke pri varnostni zaščiti potrjevanja finančnih transakcij (Islam, 2015). Prednost je na področju zdravstva za hitrejše odkrivanje bolezenskih stanj z visoko stopnjo natančnosti (Davenport & Kalakota, 2019) kot tudi pri zmanjševanju števila prometnih nesreč z uporabo avtonomnih vozil (Nyholm & Smids, 2016).

Vendar se je treba zavedati, da četudi so algoritmi prvotno ustvarjeni z dobrim namenom, lahko v nekaterih primerih vodijo do nenamernih negativnih posledic. Težava tako nastane, ko se tehnologija izkaže za neučinkovito na račun odstotka napačnih identifikacij in ustvarjanja posledičnega tehnološkega rasizma kot tudi pristranskosti (Garcia, 2017). Vse pogostejše pa so tudi zlorabe, kot so namerne lažne kampanje ali pa dezinformacije (Šprohar, 2019).

Vse večji so pozivi k etičnemu premisleku in nujnosti hitre reakcije, saj tehnologija odločanja s pomočjo algoritmov še ni vseprisotna. Evropska komisija je s tem namenom določila smernice za etično in zaupanja vredno uporabo umetne inteligence, ki dopolnjujejo prvotno strategijo za umetno inteligenco (European Commission, 2020).

Razrešitev regulatornega vprašanja, ki sem se ga dotaknila v magistrskem delu, opredeljuje smiselnost in varnost splošne uporabe oziroma ohranitve trenutne rabe posameznih primerov znotraj strok. Vse večja je namreč potreba po regulatornem pristopu in hitrih reakcijah, z namenom zavarovanja uporabnikov pred zlorabami. Poleg tega je tema magistrskega dela pomembna z vidika pomembnosti prizadevanja za rast algoritmične izobraženosti in vzpostavitve zaupanja med ljudmi v samo tehnologijo.

Namen magistrskega dela je bralca spodbuditi k razmisleku o relevantnosti zavedanja vpetosti algoritmov v vsakdanjem življenju in njihovem kontinuiranem širjenju uporabe. Razumevanje učenja algoritmov umetne inteligence kot tudi morebitne pristranskosti ustvarjalcev ter navsezadnje kompleksnosti interakcije z zunanjim okoljem je ključno, če želimo pravočasno vzpostaviti etične smernice, ki bi pripomogle k večji etični uporabi umetne inteligence. Skozi magistrsko delo želim bralca seznaniti s prednostmi uporabe algoritmov, torej k čemu vse lahko pripomorejo, na drugi strani pa prikazati negotovosti in

izzive, povezane z algoritmi, ter s tem odpreti razmislek o etičnosti uporabe in tveganjih za posameznika.

Zato je rdeča nit magistrskega dela raziskovalno vprašanje, kaj velja za etično sprejemljivo odločanje s pomočjo algoritmov. Cilj magistrskega dela je raziskati ozaveščenost ljudi o algoritmih, preveriti vpetost le-teh pri njihovem delu in predvsem pridobiti vpogled v njihovo razmišljanje o etičnosti uporabe algoritmov preko podanih resničnih primerov. Prikazala sem pomembnost etičnih vidikov pri uporabi umetne inteligence, katera s pomočjo algoritmov omogoča predvidevanja in sprejemanja odločitev. S tem so namreč povezane tudi negotovosti in izzivi. Poleg etičnih vidikov sem opredelila tudi tveganja kršitev pravic in pravil v povezavi z umetno inteligenco ter katere etične smernice narekuje Evropska komisija o uporabi umetne inteligence kot tudi navedla univerzalne smernice o umetni inteligenci in ne nazadnje podala priporočila za nadaljnje raziskovanje.

Magistrsko delo je razdeljeno na teoretični in empirični del. Teoretični del se začne s poglavjem o umetni inteligenci, ki je pomembna z vidika razvoja računalniških sistemov, da bi lahko le-ti opravljali naloge človeka inteligentno. Podrobneje sta predstavljena načina učenja umetne inteligence, torej strojno in globoko učenje. Slednja sta podvrsta nje in uporabljata algoritme za svoje delovanje. Opredeljeni so tudi glavni dejavniki delovanja umetne inteligence, med drugim kakovost podatkov, odločitve ustvarjalcev in kompleksnost interakcije z zunanjim okoljem. Sledi poglavje, ki je namenjeno algoritmom umetne inteligence. Nadaljuje se s predstavitev dveh pogledov odločanja s pomočjo algoritmov, in sicer prvi pogled pojasnjuje prednosti algoritmov skozi različne primere uporabe, med drugim pri hitrejšem postopku izkazovanja identitete, avtonomnih vozilih, varnosti in v zdravstvu. Sledi razlaga drugega pogleda odločanja, in sicer o negotovostih in izzivih, povezanih z njimi. Opisana so nasprotujoča si stališča, izguba človeške presoje, pristranskost in diskriminacija ter zlorabe. V zadnjem delu poglavja o algoritmih so predstavljeni družbeni izzivi, kot so potreba po transparentnosti, rast algoritmične izobraženosti in nadzor. Sledi poglavje o vlogi etike oziroma etičnih vidikih odločanja z algoritmi; to je poglavje o tveganjih in etičnih smernicah uporabe umetne inteligence, opredeljene so tudi univerzalne smernice o umetni inteligenci.

Sledi drugi del magistrskega dela, torej empirični, v katerem so najprej predstavljeni zasnova spletne ankete, torej načrt raziskave, izbrana populacija in tudi postopek zbiranja podatkov. Sledi del, v katerem so prikazani rezultati analize podatkov, pridobljeni s pomočjo anketnega vprašalnika. Nato so interpretirane ključne ugotovitve iz empiričnega dela, ki jih primerjam s teoretičnimi izhodišči, povzetimi s strani znanstvenih in strokovnih avtorjev. Dodana pa so tudi priporočila za nadaljnje raziskovanje. V sklepnem delu so predstavljena vsa spoznanja magistrskega dela.



# 1 UMETNA INTELIGENCA

Z izvorom umetne inteligence povezujemo Alana Mathisona Turinga. Začetki segajo v leto 1935, ko je zasnoval sodoben računalnik, danes poznan kot univerzalni Turingov stroj. Leta 1950 je predstavil test računalniške inteligence poimenovan Turingov test (Copeland, 2000). Njegova zamisel je bila zasnovati računalnike tako, da le-ti delajo preproste napake in s tem ustvarjajo vtis človečnosti. Dejal je: »Ideja o inteligenci je sama po sebi čustvena in ne matematična.« (Schwartz, 2019). Opredelitev izračunljivosti po Turingu je tisto, kar računalnik zmore narediti, in to, kar danes poznamo pod izrazom algoritem. Tako je bil prvi, ki mu je uspelo postaviti zasnovo ogrodja stroja, katerega namen je izvajanje vrste diskretnih algoritmov za dosego želene naloge, to pa izvede s pomočjo programiranja (Roos, 2019).

Vrste umetne inteligence se delijo na dva tipa glede na funkcionalnost. Prvi tip so:

- ozka,
- splošna,
- super umetna inteligenca.

Ozka je pravzaprav edina, ki obstaja danes in katere namen je reševanje oziroma odlična izvedba posamezne naloge. Primer so priporočila izdelkov v e-trgovini. V določenih okoliščinah se je zmožna približati človeškemu delovanju ali pa ga celo preseči. Pri splošni gre za teoretični koncept, ki ima kognitivno funkcijo na ravni človeka pri obdelavi jezika, slik, računalniškega delovanja, sklepanja. Napredek je super umetna inteligenca, ki je zmožna preseči človeške zmogljivosti, tako odločanje, sprejemanje odločitev kot tudi gradnjo osebnih odnosov. Razkorak med ozko in splošno je zahteven in dolg postopek, medtem ko med splošno in super umetno inteligenco razmeroma majhen in bi z njim dosegli trenutno nepredstavljava področja (Great Learning Team, 2022).

Drugi tip so:

- reaktivni stroji,
- teorija uma,
- samozavedanje.

Reaktivni stroji so ena najbolj osnovnih oblik in nimajo predhodnega pomnilnika ter tako ne delujejo na podlagi preteklih informacij. Primer je IBM-jev stroj Deep Blue, ki je premagal prvaka v šahu in omejen spomin, kjer gre za sisteme umetne inteligence, ki na podlagi preteklih izkušenj vplivajo na prihodnje odločitve. Sem spadajo praktično vse aplikacije umetne inteligence. Teorija uma je koncept, ki bi zahteval izboljšave in bo zanj potreben še določen čas. Namen je boljše razumevanje čustev, potreb, prepričanj ter misli ljudi. Samozavedanje je tisto, ki še ne obstaja, vendar v primeru obstoja bo največji mejnik na področju umetne inteligence. S tem bi bila razvita do te mere, da bi bila podobna človeškim možnostim. Nevarnost takšne ravni se navezuje na oblikovanje zamisli in idej, ki bi

prekašale intelektualnost človeškega bitja ter mu tako lahko tudi škodila (Great Learning Team, 2022).

Umetna inteligenca je podmnožica avtomatizacije, katera je sposobna delovanja brez pomoči človeka, saj njegovo vlogo pokriva programska ali strojna oprema in s tem posnema človeško inteligenco. Obstajajo različne vrste umetne inteligence kot je na primer strojno učenje (Šprohar, 2019).

Čaka nas sprememba življenja z razvojem umetne inteligence tako s pozitivnega vidika kot tudi s pojavim možnih tveganj med katerimi je za izpostaviti nepreglednost odločanja, diskriminacijo, vdor v zasebna življenja ljudi in zlorabo umetne inteligence. Pomembno je torej zagotoviti, da ji bodo ljudje zaupali. Skupek razpoložljivih podatkov, algoritmov in zmogljivosti, ki jih premorejo računalniki današnjega časa je umetna inteligenca. Vpliv, ki ga ima umetna inteligenca, pa ni samo z vidika posameznika, temveč tudi celotne družbe. Vpliv ima tako na posameznike, kot na podjetja in storitve v javnem interesu (European Commission, 2020).

## **1.1 Učenje umetne inteligence**

Odnos med umetno inteligenco, strojnim učenjem in globokim učenjem si lahko predstavljamo kot koncentrične kroge. Največji krog in kot prvotna je umetna inteligenca, sledi strojno učenje, ki je vzcvetelo kasneje, in nato globoko učenje, ki je zaslužno za današnjo eksplozijo umetne inteligence in se prilega znotraj obeh (Copeland, 2016).

Razlika med strojnim in globokim učenjem je v tem, da strojno učenje uporabi nabor algoritmov, iz katerih se uči in na podlagi pridobljenega znanja sprejme najboljšo možno odločitev, medtem ko globoko učenje algoritme strukturira v več plasti in s tem ustvari umetno nevronske mreže, slednja pa se lahko iz podanih podatkov sama uči in sprejema odločitve. Tako je razlika tudi v tem, da porabi globoko učenje večji nabor podatkov za usposabljanje kot strojno učenje in da je izhod lahko podan v katerikoli obliki, medtem ko je pri strojnem učenju izhodni podatek v numerični obliki (Great Learning Team, 2022).

Proces usposabljanja globokega učenja traja mnogo dlje, vendar po drugi strani potrebujejo algoritmi globokega učenja pri izvedbi testiranja manj časa. Prednost strojnega učenja v primerjavi z globokim je v manjših stroških strojev in grafičnih procesorjev. Najpogosteje se globoko učenje uporablja v primerih, ki zahtevajo velike količine podatkov in pomanjkanje razumevanja značilnosti (Burns & Brush, 2021).

Slika 1: Razlika v delovanju strojnega in globokega učenja.



Prirejeno po Burns (2021).

### 1.1.1 Strojno učenje

Podskupina umetne inteligence je strojno učenje z zmožnostjo avtomatskega učenja iz podatkovnih algoritmov na način, ki ne vključuje neposrednega programiranja kot tudi ne usmerjanja s strani ljudi. Stroj se osredotoča na razvoj računalniških programov, ki z dostopom do podatkov postajajo pametnejši kot tudi kompleksnejši v razmišljanju na podlagi preteklih izkušenj (DeMarco, 2019). Velja za konkurenčno prednost podjetij. Uporablja se v večji meri s pomočjo programske opreme za upravljanje odnosov s strankami (angl. Customer relationship management, v nadaljevanju CRM), pri kadrovskem informacijskem sistemu upravljanja s človeškimi viri (angl. Human resource management system, v nadaljevanju HRIS) za odkrivanje najbolj primernih kandidatov za delovno mesto, avtonomnih vozilih za prepoznavo delno vidnih predmetov in podajanja opozoril, virtualnih pomočnikov za interpretacijo naravnega govora kot tudi gonilo poslovnih modelov, katerega primer je Uber, kjer so vozniki povezani med seboj. Prednosti se kažejo v razumevanju strank na globlji ravni glede prilagajanja izdelkov in trženja, slabosti pa so v ceni kot tudi pristranskosti, saj so izurjeni na nepopolnih podatkih, kar ima za posledico neuspeh algoritma ali diskriminacijo. Razlaga modelov strojnega učenja je zapleten proces in v določenih panogah je potrebna uporaba preprostih modelov za pojasnitev sprejetih odločitev (Burns, 2021).

Ločimo glede na to, kaj je želja napovedati med (Burns, 2021):

- nadzorovanim,

- nenadzorovanim,
- polnadzorovanim,
- učenjem z okrepitevijo.

Nadzorovano učenje deluje tako, da z označenimi primeri in s pomočjo sklepne funkcije za predvidevanje ter analizo znanega nabora podatkov na podlagi tega, kar se je naučil v preteklosti, uporabi za napovedovanje prihodnjih dogodkov. Prav tako pa je algoritem sposoben poiskati napake in prilagoditi model, medtem ko se nenadzorovano uporablja, ko informacije niso razvrščene ali pa označene in tako sistem proučuje sklepanje pod temi pogoji. Rezultat, ki ga sistem ugotovi, sicer ni nujno pravilen, vendar pa je sposoben z raziskovanjem narediti sklepe iz podatkov za opisovanje skritih struktur (Selig, 2022).

Skupno obema je definiranje pravil odločanja z novimi vložki s strani algoritma, kar nakazuje na določeno stopnjo avtonomije. Hkrati je pri tem kritično, da ustvarjenih pravil odločanja ni treba razumeti človeškemu operaterju. Tako je težko vnaprej predvideti naloge, izvedene s strani strojnega učenja kot tudi opredelitev načina sprejetja določene odločitve s strani algoritma kasneje. Prav ta negotovost pripomore k prepreki zgodnjega prepoznavanja etičnih izzivov pri sami zasnovi kot tudi pri odpravi teh izzivov kasneje pri samem delovanju (Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016). Pri nadzorovanem strojnem učenju mora biti želeni rezultat, ki je pogosto povezan z oznakami vključen v podatke za učenje, ki so pravzaprav lastnosti. To je osnova, kako se algoritem uči vzorcev, nato pa se algoritmu dodajajo nove lastnosti. Ti se ocenijo glede na parametre modela. Algoritem na podlagi tega ustvari sklepe in rezultat. Problem nastane, če so podatki slabe kakovosti, saj bo algoritem ustvaril slab rezultat (European Union for Fundamental Rights, 2019).

### 1.1.2 Globoko učenje

Podskupina strojnega učenja je globoko učenje. Navdihujejo ga umetne nevronske mreže oziroma način, kako delujejo človeški možgani, pri tem pa lahko modeli dosežejo visoko stopnjo natančnosti, prav tako se ves čas prilagajajo in podajajo stalne povratne informacije. Z njim pravzaprav avtomatiziramo napovedno analitiko in tako procese naredi hitrejše in lažje za podatkovne znanstvenike. Uspešnost izvedbe modela se sorazmerno povečuje z vhodno količino podatkov v primerjavi s strojnem učenjem, kjer se zmanjšuje s povečanjem količine vhodnih podatkov. Globoko učenje je namenjeno raznoraznim potrebam, med drugim za prepoznavanje obrazov iz nadzornih kamer, pri avtonomnih vozilih, za odkrivanje bolezni (Great Learning Team, 2022). Ti modeli dosežajo zelo visoke natančnosti, vendar je po drugi strani njihova struktura nepregledna in ni točno jasno, kateri vhodni podatki so tisti, ki prispevajo do njihovih odločitev. Od tod tudi poimenovanje črne škatle (angl. Black Boxes). Potrebna je previdnost privzete uporabe, zato se stremi k boljšemu razumevanju učenja modela kot tudi tehnik (Samek, Wiegand & Müller, 2017).

Primeri uporabe globokega učenja se najpogosteje odražajo v prepoznavi slik in govora ter obdelavi naravnega jezika. Izboljšal se je tudi računalniški vid in s tem natančnost zaznave predmetov ter klasifikacije slik. Pojavljajo se v avtonomnih vozilih, storitvah prevajanja jezikov, generiranju besedil, klepetalnicah, namenjenih strankam, letalstvu in vojski, za povečanje varnosti delavcev v industriji, pri medicinskih raziskavah. Pomanjkljivosti se kažejo v zahtevani veliki količini podatkov, delovanju le za specifični primer, saj niso sposobni večopravnosti, pristranskosti in visokih denarnih vložkov (Burns & Brush, 2021).

## 1.2 Glavni dejavniki delovanja umetne inteligence

Delovanje umetne inteligence je odvisno od več dejavnikov. Na kakovost podatkov kot tudi na odločitve ustvarjalcev umetne inteligence imamo lahko vpliv, saj je odvisno od ljudi, kako jo bodo zasnovali. Medtem ko na samo interakcijo z zunanjim okoljem ne, saj se s časom umetna inteligenca sama uči, in tako lahko postanejo ljudem predlagane rešitve umetne inteligence nerazumljive. S tem se posledično razvija zaskrbljenost glede njenega delovanja.

### 1.2.1 Kakovost podatkov

Zbirajo se ogromne količine podatkov, ki se analizirajo in uporabljajo z vse večjo hitrostjo tako imenovani masovni podatki (angl. Big Data) (European Union for Fundamental Rights, 2019). V primeru napačnih ali poškodovanih podatkov lahko postane umetna inteligenca nenamerno škodljiva in pristranska kot odraz ljudi, ki so jo razvili (Šprohar, 2019). Algoritem je lahko pri svoji uporabi dober le toliko kot podatki, ki jih uporablja, zato je pomembno razmisliti o posledicah med drugim kršenje človekovih pravic ter slabe napovedi in ocene (Raso, Hilligoss, Krishnamurthy, Bavitz & Levin, 2018). Dva generična koncepta, povezana s kakovostjo, sta napake predstavitve, kar pomeni, da ni v zadostni meri pokrita populacija in napake meritev. Evropski svet je 28. junija 2018 sklenil, da so visokokakovostni podatki nujno potrebni za razvoj umetne inteligence. Vprašanja in skrbi, povezani s podatki, bi morali biti osrednja tema debate razvijalcev umetne inteligence (European Union for Fundamental Rights, 2019). Pred samim razvojem računalnikov s človeško naklonjenostjo, kot navaja Picard (1995), je potreben razmislek, ali so čustva potrebna v takšnih sistemih, katera to so oziroma kako jih lahko izrazimo ter kakšna je strategija odziva na njih. Negativna čustva v sistemih umetne inteligence imajo lahko pri zahtevnejših nalogah uničujoče rezultate.

Gledano iz drugega zornega kota pa lahko prispevajo k bolj osebnemu načinu sistema za namene, kjer je želja po stimulaciji človeškega vedenja (Martínez-Miranda & Aldea, 2005). Vsak model dela napake s tem, ko odraža ljudi, ki ga zasnujejo. Model lahko poda popolnoma napačno rešitev, s katero smo kot uporabnik oškodovani, ali pa nekaj, kar je v resnici dobro oziroma pravilno vidi kot napačno, kjer sicer ne bomo na večji škodi. Še večji problem nastane z onemogočenim vpogledom in popravkom napak posodobitev mnogih

strojno učenih modelov, saj je le-to nerazumljivo ljudem. Takšno strojno učenje se uporablja za večji del današnje umetne inteligence in ga imenujejo šibko strojno učenje. Če bi se uporabljalo tako imenovano močno strojno učenje, bi to bilo mogoče, saj bi komuniciral s človekom o notranjih posodobitvah v simbolnih oblikah (Lasry, Kobayashi & Unesco, 2018, str. 28).

### 1.2.2 Odločitve ustvarjalcev umetne inteligence

Ustvarjalci pri svojem oblikovanju dajejo prednost določenim spremenljivkam rezultat tega pa se odraža v pozitivnih in negativnih posledicah (Raso, Hilligoss, Krishnamurthy, Bavitz & Levin, 2018). Pri odločanju ima trde preference, torej omejitve, ki jih je treba pri končni odločitvi upoštevati, in mehke, kamor sodijo mnenja (Vasegaard, 2021). Kot opisuje Colson (2019) imajo tako ljudje kot umetna inteligenca vlogo procesorja pri sprejemanju odločitev, a hkrati povsem drugačne zmožnosti. Kot lahko umetna inteligenca dela napake, jih lahko tudi človek. Tudi njegova intuicija ni vedno pravilna. Človeški možgani lahko zaradi pristranskosti presodijo napačno, medtem ko je nagnjenost umetne inteligence k pristranskosti bistveno manjša, saj je sposobna odkriti nepoštenosti. Smiselno je, da je primarni obdelovalec rutinskih odločitev, temelječih na strukturiranih podatkih, umetna inteligenca. S povečevanjem vedno večje vloge odločanja s pomočjo umetne inteligence so vse večji tudi etični pomisleki. Z odgovorno umetno inteligenco zagotovimo zmanjšanje tveganj in zagotovimo večjo zasebnost, povezljivost, upravljanje in manj pristranskosti (PwC, 2021).

Vendar da bi zaupali sistemom umetne inteligence, moramo tudi razumeti, kako umetna inteligenca sprejema odločitve (Kundu, 2019). Pri odločanju je človek lahko neracionalen in ravna po intuiciji. Racionalnost in intuicija pa sta vzporedna sistema. Sodelovanje prvega in drugega sistema spodbuja proces odločanja. Umetna inteligenca tako nudi podporo človeku pri sprejemanju odločitev. Dopolnjujeta se z vidika oblikovanja novih idej, pri katerih imajo stroji težavo zaradi omejenosti s kodami in algoritmi. Interakcija med njima poteka tako, da človek predstavi problem umetni inteligenci, nato ta analizira podatke in predlaga vzorce človeku. Slednji ima dve možnosti ali da umetna inteligenca izbere vzorec in ga avtomatizira ali pa on izbere vzorec glede na njegove vrednote in cilje. Tako je proces odločanja mogoč v celoti s strani človeka ali umetne inteligence oziroma vzajemnega sodelovanja (Claudé & Combe, 2018).

V primeru vpletenosti človeka v odločitve je treba zagotoviti smiselno interakcijo med človekom in strojem. Prva izmed štirih možnosti upravljanja je imenovana Human in the loop (v nadaljevanju HITL), pri kateri človeku podporo pri odločanju zagotavlja stroj, vendar odločitve sprejema človek. Pri Human in the loop for exceptions (v nadaljevanju HITLFE) so odločitve avtomatizirane in je prisotnost človeka samo za določene izjeme, ko sistem zahteva presojo človeka za odločitev ter nadzor nad tem katere izmed izjem so mišljene za njegov pregled. Tretji pristop je Human on the loop (v nadaljevanju HOTL), ko

so vse odločitve v rokah stroja, nato jih človek pregleda in ima možnost prilagoditve pravil ter parametrov v prihodnje ali pa le odobri predlagano s strani stroja. Zadnji pristop je Human out of the loop (v nadaljevanju HOOTL), kjer je vsaka odločitev naloga stroja, človek le postavi nove omejitve in cilje in so tako vse prilagoditve temelječe na povratnih informacijah ljudi avtomatizirane (Ross & Taylor, 2021).

### 1.2.3 Kompleksnost interakcije z zunanjim okoljem

S postavitvijo sistema umetne inteligence sledi njegova interakcija z zunanjim okoljem, kjer lahko pride do nepredvidenih situacij. Nekatere izmed teh je možno zaznati s pomočjo analitičnih tehnik, vendar imajo zaradi težkega razumevanja delovanja samega sistema umetne inteligence lahko neznani vpliv na človekove pravice (Raso, Hilligoss, Krishnamurthy, Bavitz & Levin, 2018). Tako stroji kot algoritmi, ki jih poganjajo, so vse bolj kompleksni, zato se vse pogosteje dogaja pomanjkanje razumevanja njihovega delovanja kot tudi napovedovanja težav. Še tako dovršeni algoritmi kot sistemi umetne inteligence so omejeni z zaznanimi vhodnimi podatki razvijalcev in izhodnimi podatki, ki jih lahko nadzorujejo (Leetaru, 2016). Dejavniki so notranji in zunanji. Prvi so v povezavi s samo strukturo, metodami, znanjem, cilji, vrednotami, dejavnostmi in jezikom, zunanji pa z okoljem, tako naravnim kot tudi družbenim in umetnim. S kompleksnostjo se povečuje tudi ranljivost in slepo zaupanje programski opremi, ker pa so sistemi med seboj večplastno povezani, lahko iz eni ravni preide na vse (Gonzales, 2017).

Če želimo ustvariti nadzor nad algoritmi, je treba razumeti kontekst njihovega razvoja, kar pa je težavno, saj morajo strokovnjaki na tem področju z nepristranskostjo analizirati celoten proces razvoja te tehnologije, da bi prišli do izvornih modelov in formul, ki so ustvarili algoritme. Sledi ocena koristi in pomanjkljivosti oziroma morebitna tveganja, ki so posledica algoritmov. In tukaj se poraja potreba po ljudeh, ki bi imeli čas in sredstva za raziskovanje in podajanje priporočil, da bi nudili koristne ukrepe. Če želimo povečati transparentnost, bi bile potrebne algoritemske predpostavke, ki bi nudile informacije o tem, kako je spodbujena inteligenca določenega izdelka na račun algoritmov in posledice v povezavi s temi predpostavkami (Rainie & Anderson, 2017). Kompleksnost delovanja algoritmov se pokaže z neenakimi pogoji na različnih trgih in območjih, zato je potrebno prilagajanje za vsako podskupino. S tem se povečujejo tudi plasti algoritmov in njegovi stroški razvoja, nadzora kot tudi proizvodni stroški. Zaradi vse višjih stroškov in organizacijske kompleksnosti nekatera podjetja opustijo določene trge (Candelon, Charne di Carlo, De Bondt & Evgenious, 2021).

## 2 ALGORITMI UMETNE INTELIGENCE

Pričakujemo lahko vse večjo prisotnost pristopov, namenjenih reševanju problemov, ki temeljijo na podatkih, izpopolnjevanju kodiranih procesov in reševanju etičnih težav. Cilj je, da bi bili algoritmi boljši kot ljudje, kar pa ne pomeni, da morajo biti popolni, medtem ko

so negotovosti navezujoče na človeško presojo, ki se izgublja z večanjem napovednega modeliranja in podatkov. S tem je mišljena grožnja programiranja z namenom dobička, možnost manipulacije ljudi in izida s strani algoritmov kot tudi sposobnost branja naših misli, kar neizogibno pelje v družbo, temelječo na logiki, in v izgubo sposobnosti odločanja. Vse rešitve bi morale vključevati spoštovanje do posameznika. Problematična je tudi pristranskost, ki se odraža s strani ustvarjalcev algoritmov in sistemi, ki so odvisni od podatkov, ki so lahko omejeni, pomanjkljivi ali napačni. Z algoritmično kategorizacijo se poglobljajo meje in tisti, ki so že zdaj prikrajšani, bodo v prihodnosti še bolj. Mehurčke ustvarjajo tisti, ki zbirajo podatke, in zmanjšuje se izpostavljenost ljudi pred zanesljivimi informacijami in širšo paleto idej in naključij. Družbeni izzivi, ki nam jih prinašajo, so med drugim povečana težnja po algoritmični pismenosti, preglednosti in odgovornosti. Izzivi so tudi glede pravil in nadzora (Rainie & Anderson, 2017).

Z uporabo računalniških algoritmov je mišljena avtomatizacija, ki vnaprej na vhodnih podatkih opravi sprogramirane operacije, nujne za obdelavo podatkov, in privede do rešitve problema oziroma izhoda. V primeru premostitve potrebe človeške interakcije tako z napravami kot algoritmi lahko govorimo o umetni inteligenci (Podpečan, 2019). Po svoji sestavi je algoritem lahko zelo preprost ali pa je sestavljen iz kompleksnih matematičnih enačb. Kompleksnost se povečuje s številom izvedenih korakov kot tudi težavnostjo posameznega koraka. Algoritmi so gradniki strojnega učenja in umetne inteligence. Oba sta niza algoritmov, razlika je le v prejetih podatkih, ki so lahko strukturirani ali nestrukturirani (Quinyx, brez datuma). Uporabljajo se za klasifikacijo in opredelitev kategorije podatkov, za napoved prihodnjih rezultatov glede na vhodne podatke in za združevanja v skupine, kjer glede na nabor podatkov poiščejo podobnosti oziroma razlike (Rock Content Writer, 2021). Značilnosti algoritma so jasnost in nedvoumnost z dobro opredeljenimi vhodi kot tudi izhodi, izvedljivost in končnost ter neodvisnost od jezika (Prabhu, 2022).

Namen algoritmov je optimizacija procesov. Na podlagi njih deluje celoten internet in spletno iskanje. Večina finančnih transakcij se izvede z njihovo pomočjo. Pomagajo pripomočkom pri odzivu na glasovne ukaze, prepoznavi obrazov, razvrščanju fotografij in vožnje avtomobilov. Tudi vdori, kibernetiski napadi in razbijanje kriptografske kode izkoriščajo algoritme. Vse več pa se pojavljajo algoritmi za samoučenje in samoprogramiranje, kar vodi do tega, da bi lahko v prihodnosti algoritmi napisali večino, če ne vseh algoritmov. Gre za neverjetno uporabna orodja, ki pa lahko, čeprav so ustvarjeni z dobrim namenom, povzročijo nenamerne posledice. Ostaja skrb glede prevelikega nadzora v rokah korporacij in vlad, pristranskosti, ustvarjanju filtrskih mehurčkov, manjši izbiri in naključij ter večji brezposelnosti (Rainie & Anderson, 2017). S svojim delovanjem imajo moč nad ogromnimi količinami podatkov prav tako pa imajo izjemen vpliv na zdravstvo, finance, kriminal, prevoz, podnebje kot tudi odločanje, učinkovitost in produktivnost (DeMarco, 2019). Razvoj kompleksnih algoritmov umetne inteligence je dosegel, da stroji lahko razmišljajo in se odločajo na enak način kot ljudje, kar je mogoče videti v kmetijstvu, zdravstvu, trženju, poslovni analitiki, robotiki (Scott, 2021).



K neizogibnosti algoritmov je privedla nujnost sistema zaradi vse večjih količin podatkov kot tudi pojav zmogljivih grafičnih enot in pojav globokega učenja. Vse, kar je bilo nekoč mišljeno kot nemogoče, postaja nujno in tudi pričakovano, kar predstavlja nastanek novih vprašanj o etiki, pristranskosti in pravičnosti (IBM, brez datuma c). Nadzor nad algoritmi je bil do zdaj majhen. S tem ko postajajo temelj delovanja sodobne družbe, je čas za uvedbo reguliranja uporabe, določanja pravil delovanja in nadzora. Algoritmi so posledica digitalne dobe, v kateri je poplava podatkov, s tem pa se je razvila tudi potreba po avtomatizaciji, saj praktično druge izbire ni bilo. S širitvijo neizogibnosti se uporabljajo na vse več področjih in s tem postajajo poleg filtriranja podatkov tudi zunanji odločevalci. Njihovo posredovanje med ljudmi in stvarmi je nejasno, čeprav vplivajo na življenja in iz tukaj izhaja problem nepoznavanja njihovega delovanja (Alang, 2019). Z vse večjim vgrajevanjem umetne inteligence v izdelke, storitve, procese in odločanje je vse več pozornosti, na kakšen način programska oprema uporablja podatke, še posebej z razvojem kompleksnih algoritmov (Candelon, Charme di Carlo, De Bondt & Evgenious, 2021).

Do leta 2024 se napoveduje, da bo umetna inteligenca boljša kot ljudje pri prevajanju jezikov, do 2031 pri prodaji blaga in do leta 2053 pri izvedbi operacij. Že zdaj pa se pojavlja na številnih področjih, kot so klepetalni roboti, pametna vozila, vrhunske izkušnje zdravstvenega, bančnega, logističnega in potovalnega sektorja. Prav tako je postala prisotna v večini domov z glasovnimi pomočniki. Roboti pa bi v prihodnosti lahko nudili pomoč ljudem pri opravljanju nalog, pri katerih je večja stopnja nevarnosti kot tudi v medicini, varovanju, gradnji in industriji (Batok, 2020).

Smo v dobi algoritmov, kar odraža širjenje njihove uporabe kot posledica ustvarjanja, zajemanja in analiziranja ogromne količine podatkov. Prisotni so v vsakdanjih procesih kot tudi tistih, skritih pred očmi javnosti (Rainie & Anderson, 2017). Sprejemanje odločitev se je v večji meri preusmerilo iz ljudi na algoritme. Raven natančnosti in računske moči je preseгла človeka. Etično vprašanje, ki se tukaj poraja, je, ali lahko dopustimo sprejemanje odločitev s strani algoritmov, ki jih ni mogoče razložiti (Nielly, 2020).

Med eno izmed najbolj konkurenčnih področij spadajo platforme strojnega učenja. Velika imena, kot so Amazon, Google in Microsoft, med seboj konstantno tekmujejo. Raziskovalci pa želijo trenutno optimizirane algoritme za opravljanje izključno ene naloge preoblikovati v bolj prilagodljive modele in tako omogočiti stroju, da kontekst ene naloge uporabi za prihodnje. Napoved tehnološko raziskovalnega podjetja IDC je kljub padcu svetovnega gospodarstva kot posledice pandemije koronavirusne bolezni 2019 ocenila 110 milijard dolarjev letno vredno poslovno uporabo umetne inteligence na svetovni ravni do leta 2024. Sistemi umetne inteligence se povečujejo v podjetjih in so tudi sestavni del njihovih strategij. Sprva se je predvidevalo, da se bo v prihodnosti umetna inteligenca pojavljala le pri avtomatizaciji ponavljajočih se nalog, ki so preproste in je stopnja odločanja na nizki ravni, vendar je z razvojem vse močnejših računalnikov in s povečevanjem podatkovnih nizov umetna inteligenca zrastle v sofisticiranosti (Pazzanese, 2020).

Raziskava Stanforda »Artificial Intelligence and Life in 2030« opredeljuje uporabo umetne inteligence v prihodnosti. Do spremembe bo prišlo pri načinu vožnje, specializaciji robotov, ki bodo zagotavljali večjo varnost, na področju izobraževanja je možnost platform za obdelavo jezika in razvoj pouka ljudi poleg obstoječih interaktivnih sistemov poučevanja, združevanje orodij bi pripeljalo do bolj interaktivnih in prilagojenih medijev. Če bi se povečevalo zaupanje medicine v umetno inteligenco, bi se razvijale naprave za spremljavo osebnega zdravja kot tudi pomoč pri operacijah z roboti. Pri analizi vzorcev kriminala bi prispevala k manjši človeški pristranskosti in večji varnosti, ne da bi bila pri tem odvzeta svoboda in dostojanstvo. Kot posledica hitro spreminjajočega gospodarstva pa se bodo delovna mesta izgubljala, a hkrati ustvarjala nova (Abate, 2016).

Svetovni gospodarski forum je v letu 2020 napovedal, da bo umetna inteligenca do leta 2030 gospodarstvu prispevala 15,7 bilijona dolarjev. S tem bi prišlo do rasti bruto domačega proizvoda za 26 %. Do leta 2025 naj bi kar 95 % nalog z naslova storitev za stranke opravila umetna inteligenca. Algoritmi se z leti vse bolj razvijajo in v bližnji prihodnosti je pričakovati samo posodabljanje in popravke vedenja, s tem pa tudi večje zaupanje (Mazurek, brez datuma).

## **2.1 Prednosti algoritmov**

Količine podatkov se iz dneva v dan povečujejo in s pomočjo algoritmov bo upravljanje letih mnogo lažje. S tem bo napredovala tako znanost kot človeške zmogljivosti in vse več bo povezave ljudi z informacijami. Algoritmi pripomorejo k večji hitrosti v primerjavi s tradicionalno izvedbo, krčijo baze podatkov, zmanjšujejo onesnaženost in ekonomske odpadke. Pripomorejo k zmanjšanju trenj glede vedenj, kot sta odločanje in sprejemanje racionalnejših odločitev. Nudijo povezljivost in s tem priporočila glede na posameznikove preference (Rainie & Anderson, 2017). Prednosti se kažejo tudi v avtomatizaciji procesov in večji natančnosti (Nexus Integra, brez datuma).

Primeri dobre uporabe so banke, kjer z bolj interaktivnim odnosom pripomorejo k zmanjšanju tveganja in povečanju možnosti dajanja posojil. Zdravstvene storitve s pomočjo nove tehnologije določen del obremenitev preusmerijo na ljudi, kar je pomembno z vidika naraščajočih režijskih stroškov in upravljanja lastnega zdravja. Največji vpliv bi imelo tudi za analizo podatkov, slikanja in diagnoze, saj bi s tem pri vsaki odločitvi zdravljenja bilo na voljo vse medicinsko znanje o tej bolezni. S pomočjo avtonomnih vozil se zmanjšujejo prometne nesreče. Policijsko delo je bolj proaktivno in usmerjeno k preprečitvi kriminala. Oglaševanje je bolj ciljno usmerjeno (Rainie & Anderson, 2017). Zmanjšala pa se bo tudi birokracija s tem, ko bo vlada izvajala manj nadzora in regulacije (Tranberg, 2017). S kadrovskega procesa je uporabno za analizo in spodbujanje hibridnih delovnih mest. Tudi pri samem opravljanju dela ni njen namen nadomeščanje zaposlenih, temveč prenos tehničnih nalog, ki so pomembne in se tako delavci lahko osredotočijo na druge odgovornosti, s tem pa se poveča produktivnost (Pazzanese, 2020).

Na podlagi neznanstvene ankete je 38 % anketiranih mnenja, da bodo pozitivni učinki pretehtali negativne, 25 % pa, da bo izid pozitivno-negativen. Po mnenju voditeljev bodo sistemi uspešni, če se bo pozornost namenilo zasnovanosti teh orodij in posodabljanju. Napovedi strokovnjakov kažejo, da se bo življenje ljudi z vzponom umetne inteligence izboljšalo. Načrtuje se, da bodo pametni sistemi prispevali k prilagojeni prihodnosti v večji meri in nudili prihranek časa in denarja (Rainie & Anderson, 2017).

### 2.1.1 Hitrejši postopek izkazovanja identitete

Tehnologije, namenjene prepoznavi obrazov, so vse bolj prisotne v splošni rabi zaradi enostavnosti uporabe kot tudi hitrosti. Njihovi pozitivni učinki se odražajo v varnosti in uporabnosti. Nudijo nove načine oglaševanja, nakupovanja, plačevanja, odkrivanja bolezenskih stanj, varnostne zaščite finančnih transakcij in varovanja navezujoč na pregon, iskanje pogrešanih oseb, hitrejši postopek legitimacije (Savič, 2019).

Vse več pa je tudi opozoril glede zasebnosti in problematike podatkovnih zbirk, ki so lahko deležne zlorab in napak. Posvetiti bi se morali etičnim izzivom, saj napake nakazujejo na veliko potrebo po regulatornem pristopu, ki bi zagotovil varovanje uporabnikov (Savič, 2019). Digitalizirana družba je omogočila, da se podatki lahko zbirajo povsod in ves čas. Splošna uredba o varstvu podatkov (angl. General Data Protection Regulation, v nadaljevanju GDPR) ima veliko vlogo pri sami zaščiti podatkov, vendar po drugi strani evropske države, ki so del Schengena, morajo zbirati podatke vseh, ki vstopijo v državo, in jih deliti z drugimi državami preko obveščevalnih služb. Če se biometrični podatki ne uporabljajo preudarno, se lahko zgodi omejevanje svobode gibanja in dostopnost. Potreben je etični premislek o tem, kdo naj jo uporablja in za kakšen namen, saj se z razvojem tehnološkega napredka povečuje tveganje za razvoj negativnih dogodkov. Osnovna pravica ljudi je informirano soglasje, ko se od njih zahtevajo biometrični podatki kot tudi transparentnost glede uporabe, deljenja in hranjena podatkov. S tem bi se močno povečala ozaveščenost. Vprašanje, ki se tukaj poraja, je, ali je naloga države zaščita pred zlorabami, vendar ta zaenkrat lahko poskrbi za osnovno zaščito, nadaljnje razpravljanje o omejevanju pa bi se lahko začelo po seznanjenosti ljudi o razpoložljivi tehnologiji in njeni uporabi (Henigman, 2019).

S pomočjo biometrije se lahko preverja identiteta posameznika na podlagi telesnih oziroma vedenjskih značilnosti. Te so prstni odtis, dlan, obraz, šarenica, očesna mrežnica, DNK, vonj, uho, preplet ven na roki, lastnoročni podpis, govor, tipkanje in gibanje. Razširjenost uporabe biometrije je posledica unikatnosti, neprenosljivosti, težke ponaredbe ali prekritja. Vse večja pa je tudi potreba po avtomatizaciji, natančnosti in hitrosti ugotavljanja identitete. Sporna je iz vidika razkrivanja zdravstvenega stanja ljudi na primer iz očesne šarenice ali pa ovira v primeru poškodbe (Informacijski pooblaščenec, brez datuma). Odstotek napačnih identifikacij je visok, in sicer 0,3 odstotka. V Evropski uniji je s pomočjo GDPR urejeno prepovedano zbiranje vseh biometričnih podatkov. Vendar se kršitve na tem področju

pojavnjajo s strani oblasti z uporabo pametnih kamer in s strani ponudnikov tehnologije s tem, ko ni zagotovil o izbrisu podatkov (Rajšek, 2019).

Z vse večjo količino vizualnih podatkov je postala dostopna računalniška moč, ki je nujna za analizo podatkov. S tem pa so sistemi dosegli od začetne 50 % natančnosti 99 %, kar je več kot pri ljudeh. Računalniški vid je področje strojnega učenja, ki je integriran v že nam poznane izdelke. Tako omogoča avtonomnim vozilom razumevanje okolice z zaznavanjem pešcev, predmetov, prometnih znakov. Nudi pomoč pri avtomatizaciji nalog, kot je pomoč pri odkrivanju bolezenskih stanj na slikah. Uporablja pa se tudi pri prepoznavi obrazov (Mihajlovič, 2019). Prepoznavanje obraza je prednostno biometrično merilo, ki deluje v prvem koraku z zaznavo obraza, nato zajem obraza pretvori iz analognih v digitalne informacije z uporabo algoritma, sledi preverba ujemanja obraza. Dokazovanje identitete je možno tudi s prstnimi odtisi, skeniranjem šarenice, prepoznavo glasu, vedenjskimi meritvami in digitalizacijo žil na dlaneh (Ford, 2021).

Facebook ponuja DeepFace s 97,25-% stopnjo natančnosti presojanja dveh fotografij obraza iste osebe, Google nudi FaceNet z natančnostjo 99,63 %, ki je vključen v Google Foto in je namenjen razvrstitvi slik ter samodejni označbi glede na prepoznane osebe. Slabše rezultate zaradi visoke stopnje napak identifikacije temnopoltih žensk imajo orodja, temelječa na Microsoftu in IBM-u. Amazon je leta 2018 promoviral Rekognition organom pregona za prepoznavanje obrazov v oblaku. Namen je bil prepoznavanje do 100 ljudi na eni sliki glede na baze podatkov desetine milijonov obrazov. Vendar se je kasneje izkazalo, da je napačno prepoznalo ljudi ameriškega kongresa, saj jih je označilo kot aretirane kaznivih dejanj (Ford, 2021).

Direktorat za znanost in tehnologijo Združenih držav Amerike (v nadaljevanju ZDA) je s pomočjo sponzoriranih testov v Marylandskem testnem centru preverjalo učinkovitost 12 sistemov za prepoznavo obraza na hodniku 2 x 2,5 m. Rezultati so bili 99,44-% stopnja pridobivanja obrazov v manj kot 5 sekundah s stopnjo resnične identifikacije 98 %, stopnja napake je bila 1 %. Nacionalni inštitut za standarde in tehnologijo (v nadaljevanju NIST) je izdal leta 2019 poročilo o tem, kako natančna je tehnologija, namenjena prepoznavanju obrazov. Ugotovilo se je, da je razpon lažno pozitivnih rezultatov 0,003 % oziroma 3 napake od 100.000 poizvedb ter 0,3 % oziroma 3 napake na 1.000 poizvedb. Napake so odvisne od kakovosti slike poizvedbe, osnovnega nabora podatkov in rase obraza. Prav tako so slike, na katerih so bile ženske, dosegle višjo stopnjo lažno pozitivnih napak v primerjavi z moškimi kot tudi slike temnopoltih ljudi v primerjavi s svetlopoltimi (Ford, 2021). Januarja 2020 so dokazali, da so se kritiki motili, saj najboljši algoritmi nimajo rasne ali spolne pristranskosti pri prepoznavanju obrazov. Iz poročila avgusta 2020 in marca 2021 je mogoče videti povečanje učinkovitosti algoritmov v kratkem času glede natančnosti prepoznavanja obrazov z obraznimi maskami zaradi koronavirusne bolezni 2019 (Thales, 2021).

Uporaba je prisotna za nadzor množice na zasebnih in javnih prireditvah, pri novejših fotoaparatih za samodejno ostrenje kot ekstrakcijo obraznih potez, razvrščanje po spolu kot

pomoč pri identifikaciji osebe iz digitalne slike ali video okvirja, tudi vse bolj je pomembno za trženjske namene kot pomoč pri analizi vedenja strank in ciljnega oglaševanja, je pa tudi uporabno za zaznavo prisotnosti ljudi v kombinaciji z biometrijo za upravljanje dostopa (Boesch, brez datuma).

### 2.1.2 Avtonomna vozila

Avtonomno vozilo je sposobno zaznavanja okolja in delovanja brez človeka. Zanašajo se ne senzorje, kompleksne algoritme, sisteme strojnega učenja, procesorje in aktuatorje (Synopsys, brez datuma). Vse bolj postajajo realnost z napredkom tehnologij, umetne inteligence in robotike. Prednosti se kažejo v zaščiti okolja, gospodarski rasti, večji dostopnosti, varnejših cestah, novih vozilih, elektrifikaciji vozil in revoluciji urbanističnega načrtovanja. Učinek bo viden po celotni vrednostni verigi, vse več pa bo tudi novih poslovnih modelov oziroma razvoj obstoječih, kot sta elektronsko poslovanje in mobilnost kot storitev (Evropska komisija, 2018).

Cilj je s povečanjem varnosti zmanjšati smrtne žrtve na cestah in posledično vplivati na manjše zdravstvene stroške. Problem, ki se pojavlja, je v etično-moralnih vprašanjih kot tudi pravnih. Primer je vožnja z avtonomnim vozilom, ki mu zaupamo vožnjo po mestu, medtem ko mi opravljamo drugo delo. Na cesto v določenem trenutku priteče otrok. Hitrost ter razdalja onemogočata varno zaustavitev avtomobila. Vprašanje tukaj je, kako se bo odzval: ali bo zaviral in ohranil isto smer ter posledično trčil v otroka ali se mu bo izognil s tem, ko bi zavil na pločnik, kjer pa so ostali pešci, ali na nasprotni vozni pas, kjer vozijo pravilno vozeča vozila. Znašlo bi se torej na razpotju vprašanj, kako odreagirati. Ravno od tu izhaja problem odgovornosti (Triglav, 2016).

Izzivi so navezujoč na varnost zaradi delitve ceste z drugimi udeleženci v prometu in klasičnimi avtomobili. Nujna je tudi harmonizacija cestnoprometnih predpisov v Evropski uniji (v nadaljevanju EU). Področje civilne odgovornosti mora pojasniti ali odgovornost nosi oseba, ki je v avtomobilu, pa čeprav ga ne upravlja, ali proizvajalec. Naslednji izziv je povezan z obdelavo podatkov. Sicer evropska zakonodaja varstva podatkov pokriva tudi avtonomno vožnjo, vendar zaenkrat ukrepov, ki bi pripomogli h kibernetiki varnosti vozil, še ni bilo. Izziv so tudi etična vprašanja, za katera so bile pripravljene evropske smernice na področju umetne inteligence, saj je pomembna svobodna izbira in spoštovanje človeškega dostojanstva. Potrebni pa so tudi primerni standardi. Ne nazadnje je izziv v vložkih v razvoj infrastrukture, ki bi pripomogla k prenosu podatkov ter komunikaciji v prometu (Evropski parlament, 2019). Tehnološki izzivi, povezani z avtonomnimi vozili, naj bi bili odpravljani v 75–80 %, vendar ostajajo kulturni izzivi povezani s sprejemanjem, saj so nekateri na primer predani vožnji. Predstave ljudi o avtonomnih vozilih so namreč precej različne (Drevenšek, 2019). Študija ameriškega avtomobilskega združenja razkriva, da je vožnje z avtonomnim vozilom strah tri od štirih Američanov (Edmonds, 2019).

Potreben bo kompromis med varnostjo in hitrostjo, saj se kratki reakcijski časi uporabijo za povečanje varnosti in hitrosti. Pride lahko tudi do odpora zaradi prevladujoče se družbe z avtonomno tehnologijo. Na drugi strani pa ravno zaradi pritiska k varnejši vožnji lahko pride do družbenih in političnih sporov. Etično vprašanje je tudi glede nadzora. Policija bo lahko za povečanje varnosti elektronsko prevzela nadzor, obstaja pa tudi nevarnost v primeru, da bi to uspelo kriminalcem. Zloraba je tudi možna pri zbiranju informacij, ki so potrebne za zagotovitev optimiziranih poti. Z namenom izboljšanja varnosti ostalih udeležencev v prometu bi to bilo možno s transponderji, ki bi zagotovili obveščanje avtonomnih vozil o gibanju ostalih udeležencev v prometu, vendar se spet poraja vprašanje zasebnosti. Zaradi tipov in modelov vozil bi se spodbujalo družbeno-ekonomsko ločitev prometa in s tem vplivalo na socialno kohezijo. Vpliv bi imelo tudi na upad delovnih mest poklicnih voznikov. Etične posledice, ki jih prinašajo avtonomna vozila, bodo v veliki meri odvisne od družbenih odločitev. Vprašanja so glede varnosti, porazdelitve odgovornosti, organizacije prometa, usklajevanja med zasebnostjo in drugimi interesi ter prilagoditve glede na podnebno in okolijsko politiko (Hansson, Belin & Lundgren, 2021).

Naklonjenost avtonomnim vozilom je 58 % glede na študijo med evropskimi državljani. Zaradi povzročenih nesreč bo družba sprejela avtomatizirano mobilnost, ko bodo zagotovljeni najvišji varnostni standardi. Med drugim je treba najti ustrezno ravnovesje, ki bi zagotovilo učinkovito konkurenco v smeri inovativnih rešitev ter nudilo varstvo pri izmenjavi javnih in zasebnih podatkov. Evropska komisija stremi k celovitemu pristopu EU povezane in avtomatizirane mobilnosti in v prihodnost usmerjeno evropsko agendo, ki ima skupno vizijo ter podporne ukrepe za sam razvoj infrastrukture, tehnologij in storitev. Z njo bo zagotovljena pripravljenost pravnega in političnega okvira EU za podporo pri uvedbi avtomatizirane mobilnosti, ki bo varna in bo reševala družbene in okolijske pomisleke, ki pa igrajo veliko vlogo za sprejemanje pri javnosti. Ko bo le-ta del celotnega prometnega sistema ter podprta s pravnimi ukrepi in sinergijami, bi to privedlo do »vizije nič«, kar bi pomenilo, da na cestah do leta 2050 ne bi bilo več smrtnih žrtev. To bo možno le, če bodo med seboj sodelovali EU, države članice, zasebni akterji in regionalni ter lokalni organi (Evropska komisija, 2018).

Na lestvici šestih stopenj avtonomne vožnje, ki jo je oblikovalo Society of Automobile Engineers (v nadaljevanju SAE) dosegamo trenutno 3. stopnjo (Schrauth & Funk, 2021). Prva stopnja dopušča, da v določenih razmerah in pogojih vozilo samo nadzira hitrost in zavijanje, druga stopnja pomeni delno avtomatizacijo, torej vozilo lahko samo pospešuje, zavira in zavija. Na tretji stopnji pogojene avtomatizacije je vozilo sposobno opraviti večino funkcij kot tudi spremljati okolico in obvestiti voznika o nevarnostih v določenih primerih, ko vozilo ne zna pravilno reagirati. Visoka avtomatizacija je značilna za četrto stopnjo, ko vozilo samostojno vozi pod določenimi pogoji. Največji približek k temu je bil Googlov prototip vozila, ki je bil brez pedalov in volanskega obroča ter je dosegal hitrost 40 km/h, vendar je bil kasneje umaknjen. Zadnja stopnja je popolna avtomatizacija, torej vozilo vozi

enako dobro kot človek v vseh pogojih, vendar takšnega vozila ni niti še v prototipni obliki (Poženel, 2018).

### 2.1.3 Varnost

Ena izmed najhitreje rastočih tehnologij umetne inteligence na področju varnosti in kazenskega pregona je biometrija. Nudi iskanje milijarde slik iz družbenih omrežij, potnih listov in voznških dovoljenj za identifikacijo osumljencev. S tem se izvaja premik družbe k avtomatizaciji procesov odločanja (Smith & Miller, 2021).

Za finančne institucije so takšne tehnologije zelo učinkovite, saj drastično zmanjšujejo stroške in uspešno preprečujejo pranje denarja. Omogoča lažjo kvalifikacijo ljudi pri večjih hipotekah. Tudi bančništvo s pametnim telefonom je postalo varno in preprosto. Ob pravilni uvedbi je sprejetje med strankami kar 93 %. Poleg bančnega sektorja pa imajo koristi tudi vse storitve, ki obdelujejo občutljive podatke. Pripomore pri manj zahtevnih procesih kot je hitrejši proces za izpolnjevanje obrazcev, saj nismo vezani na lokacijo in zmanjša se čas vnosa podatkov. Z zdravstvenega vidika omogoča hitro preverjanje zdravstvenega zavarovanja, kar pa je pomembneje, je to, da je mogoče bolnike prepoznati v nezavesti, kar prepreči nepravilno zdravljenje. Napreduje tudi varnost s tem, ko pomaga identificirati posameznike, zaznava goljufije in nadzoruje nezakonite premike. Prav tako uporablja napredne nadzorne sisteme policija za prepoznavanje osumljencev (Lunter, 2021). Že več kot desetletje se povečuje pomembnost te tehnologije za organe pregona in je uveljavljena na letališčih. Prav tako je predmet sodnega nadzora na sodiščih (Smith & Miller, 2021).

Obveščevalne službe in varnostni organi uporabljajo napovedne policijske računalnike za napoved prihodnje kriminalitete. Poleg tega je v uporabi program, ki pripomore pri oceni tveganja posameznikov, ki bi lahko bili potencialni teroristi. Podatke pridobiva s pomočjo videonadzora javnih prostorov, ki imajo hkrati vgrajene programe za prepoznavo obrazov ter nato te podatke uporabi za primerjavo z obstoječimi policijskimi podatkovnimi zbirkami (Završnik, 2017).

Pojavljajo pa se etična, pravna in politična vprašanja med zakonitim zbiranjem za povečanje varnosti in pravicami zasebnosti in avtonomije. Z vidika etičnih načel so pomisleki na eni strani glede varnosti in na drugi glede zasebnosti in avtonomije ter tudi demokratičnosti. Opirajo se dvomi zaradi integracije biometričnih podatkov na primer finančnih, zdravstvenih, davčnih in podobno. Kar se čuti na Kitajskem, kjer se uporablja na javnih mestih video nadzor preko CCTV in na podlagi tega sistem socialnih kreditov nagrajuje oziroma kaznuje državljane (Smith & Miller, 2021).

Facewatch za preprečevanje kriminalnih in terorističnih napadov uporablja platformo MXSERVER, s pomočjo katerega sistem izda opozorilo pri vstopu registriranega kriminalca v prostor s primerjavo obraznih biometričnih podatkov. Onifido je spletna platforma za digitalno preverjanje. EyeLock z izdelkom Nano NXT preverja pristnosti šarenice in se

uporablja na bankomatih. PalmSecure Sensor skenira žile na dlani brez vzpostavljanja stika z napravo. S tem je izboljššan brezkontaktni in brezgotovinski plačilni sistem in onemogočene zlorabe. BehaviorSec preverja pristnost glede na vedenjske vzorce (Madhavan, 2020). ShotSpotter opozori oblasti o streljanju in lokaciji. Podjetje Hikvision bo s pomočjo posebnih čipov iskalo potencialne kriminalce in pogrešane ljudi, zaznavalo sumljive nepravilnosti, izvajalo prepoznavanje obrazov in podobno ter s tem doseglo kar 99-% natančnost.

Medtem pa na trgu obstajajo tudi podjetja, kot je Predpol, katerega cilj je napovedovanje zločinov glede na analizo preteklih podatkov. Kitajsko podjetje Cloud Walk Technology želi celo napovedati, ali bo določen posameznik storil kaznivo dejanje, še preden ga dejansko naredi, s pomočjo prepoznave obraza in analize hoje, saj zazna sumljive spremembe. Program Hart glede na zbirko kaznivih dejanj odloča o izpustitvi osumljenca glede na stopnjo tveganja. Na podlagi dosedanjih rezultatov je bil sistem v 98 % uspešen glede nizkega tveganja in 88 % glede visokega tveganja. Podobno se uporabljajo algoritmi za ocenjevanje tveganja pri sprejemanju odločitev o izpustitvi pred sojenjem in pogojni izpustitvi s pomočjo programa Compass, ki je dosegel zadovoljivo napovedno natančnost in učinkovitost. Je bil pa kljub temu na udaru glede pristranskosti, saj so imeli temnopolti ljudje dvakrat večjo možnost napačne razvrstitve v primerjavi z belci (Faggella, 2019).

#### 2.1.4 Zdravstvo

Umetna inteligenca v zdravstvu je kot del informacijskega sistema, programske opreme na napravi ali načrtovanja. Njen namen je izboljšati diagnoze in terapije. Stremi k personalizaciji napovedi in robotsko podprtim kirurškim instrumentom. Vizija je celotna avtomatizacija prepoznavanja in označevanja medicinskih slik ter načrta obsevanj. Obeta se odločanje zdravnika s pomočjo predlogov, ki mu jih bo podala umetna inteligenca (Kolednik, 2019). Bližnja prihodnost bo omogočila, da se bodo nekateri zdravstveni parametri beležili avtomatsko in s pomočjo mobilnih naprav posredovali zdravnikom (Kušar, 2018). Umetna inteligenca in strojno učenje sta že v koraku k preoblikovanju medicine. Vse več orodij za medicino na podlagi umetne inteligence ter porast medicinskih podatkov gre z roko v roki. Za ljudi so zmožnosti za obdelavo kot tudi uporabo že presežene, zato lahko ravno strojno učenje in napovedni algoritmi prispevajo k oblikovanju pomena poplave podatkov in prepoznavi vzorcev, ki se lahko uporabijo za boljše odločitve pri postavitvi diagnoz (Battelle Insider, 2019).

Umetna inteligenca prispeva k večji hitrosti, natančnosti diagnoze in omogoči, da zdravniki svoje znanje nudijo nujnim primerom, medtem ko umetna inteligenca lahko prepozna znake raka, hitreje obdelava diagnostične slike ter jih označi za nadaljnji strokovni pregled ljudi in nudi priporočila. Orodje umetne inteligence bi lahko odkrilo vzorce z branjem in razlago zdravstvenih kartotek, saj bi jih primerjali z drugimi bolniki in s tem postavili diagnozo. Še posebej uporabno bi bilo to pri redkih boleznih. Tudi napovedovanje koristi zdravljenja in



neželenih učinkov za določenega bolnika je možno z umetno inteligenco. Številne naprave za spremljanje, kot sta na primer merilnik glukoze in inzulinska črpalka, ki omogoča samoregulacijo, pa tudi pametni elektrokardiogrami, ki pošiljajo opozorila, če se srčni ritem poslabša, so programirani k pošiljanju opozoril, če so izven določenih parametrov. Umetna inteligenca je sposobna poiskati vzorce, ki nakazujejo na negativne dogodke, iz različnih naprav. S tem bi bile s klinično napovedno analitiko napovedane težave, preden bi do njih prišlo, saj bi ugotovili, kateri bolniki so bolj izpostavljeni določenemu tveganju (Battelle Insider, 2019).

Orodja umetne inteligence se uporabljajo tudi pri analizi mikrobnih genomov za odkrivanje resnosti patogena, nalezljivosti in pa odpornosti na antibiotike. V povezavi z zdravili bi bilo mogoče oblikovati zdravila v virtualnem okolju in izbrati primerne kandidate za sintezo in testiranje. S tem bi izločili formulacije z negativnimi učinki, zmanjšali število kandidatov z možnostjo negativnih učinkov in omogočili hitrejši časovni okvir razvoja. Manjši pa bi bili tudi stroški in tveganja razvoja zdravil (Battelle Insider, 2019).

Vznemirljivost umetne inteligence se kaže tudi na področju nevronske premostitvene tehnologije, ki omogoča paraliziranemu človeku zavestno nadzorovanje zapestja, roke in prstov (Battelle Insider, 2019).

S pomočjo umetne inteligence se je zgodil preboj na področju prepoznavne mutacij v rakavih tkivih, ki je tako dosti bolj natančno in vključuje strojno učenje, ki omogoči avtomatizacijo diagnostičnega procesa tumorske DNK ter tako poda rezultat, na podlagi katerega zdravnik izbere ciljno zdravljenje. Ravno to je bil prej največji problem, ker se rak neprestano spreminja, ravno tako tudi njegov DNK. Programska oprema ima enako natančnost kot človeški radiologi pri odkrivanju raka dojke. V pomoč bi bila lahko v primerih, ko gostota ne bi omogočila jasne diagnoze kot tudi patologom, saj je pri avtomatski detekciji metastatskega raka dojke globoki učni sistem zmanjšal stopnjo človeške napake za 85 %. Več raziskovalnih skupin je razvilo pametne algoritme, namenjene diagnostiki kožnega raka, kot je aplikacija SkinVision (The Medical Futurist, 2019).

Razvrščanja ehokardiogramov s pomočjo globokega učenja, s katerimi preverjamo bolnikove srčne bolezni, je doseglo 92 % natančnost algoritma, medtem ko je bila pri ljudeh 79 %. Razvit je bil algoritem, ki napove srčni infarkt ali možgansko kap v 10 letih s skeniranjem rutinskih zdravstvenih podatkov pacientov. Glede na standardne metode napovedovanja je bilo pravilno napovedanih še 355 primerov usod pacientov (The Medical Futurist, 2019).

Razvit je bil algoritem, ki odkriva zgodnje zanke depresivnega vedenja in tako pomaga pri zmanjšanju pojavnosti hudih duševnih bolezni. Po zbranih podatkih je bila stopnja natančnosti 84 % glede samomora v naslednjem tednu in v 80 % v dveh letih (The Medical Futurist, 2019).

Razvija se strokovni nadzor bolnika, ki bi bil nenehno ob postelji in povezoval različne dejavnike med seboj in tako pravilno usmerjal zdravljenje. To bi pomenilo velik napredek glede na današnje stanje, ko morajo to početi zdravniki in so instrumenti nepovezani med seboj. Algoritmi so sposobni napovedati, kdaj se nekdo od bolnikov bliža koncu življenja in bi tako prejel oskrbo točno takrat, ko je to potrebno. Usposobljen je bil za analizo različnih dejavnikov elektronskih zdravstvenih zapisov v 3- do 12-mesečnem obdobju pred smrtjo bolnika in s tem označbe ustreznih kandidatov paliativne oskrbe. V vseh primerih, ki jih je izbral algoritem za paliativno oskrbo, je to potrdila tudi skupina zdravnikov. Poleg tega pa je algoritem celotno odločitev prepustil zdravnikom in so tako lahko delovali skupaj kot ekipa (The Medical Futurist, 2019).

Spekter uporabe medicinskih algoritmov je širok od diagnostike do napovedi kot tudi spremljanja zdravja in pomoči pri administraciji. Pa vendar so v osnovi zgrajeni na podlagi homogenih nizov podatkov, kar pomeni, da ne zajemajo celotne populacije, poleg tega pa so oblikovani s pomočjo pravil, ki so lahko pristranski (The Medical Futurist, 2019). Ta se lahko pojavi v procesu kadar koli med ustvarjanjem algoritma, lahko pa že pri sami zasnovi študije in izbiri podatkov med vnosom in izbiro podatkov ali pa med izbiro algoritma in modela (Igoe, 2021). Bolniki tudi ne vedo, da se pri njihovem zdravljenju kot pomoč pri svetovanju zdravniku uporablja algoritem (Ruiz, 2020).

## **2.2 Negotovosti in izzivi povezani z algoritmi**

Glavne grožnje umetne inteligence in prihodnosti ljudi so izguba nadzora, zlorabe podatkov, izgube delovnih mest, zmanjšanje kognitivnih, socialnih in preživetvenih veščin, večje število zločinov z avtonomnim orožjem in kibernetškega kriminala (Rainie & Anderson, 2017). Kljub obljubljenim gospodarskim koristim je vse bolj prisotna zaskrbljenost glede družbene škode zaradi uporabe programske opreme za potrebe odločanja v medicini, kadrovanju, kreditni sposobnosti in kazenskem pravosodju s strani zasebnih podjetij, ki to izvajajo brez nadzora in obveze po razlagi (Pazzanese, 2020).

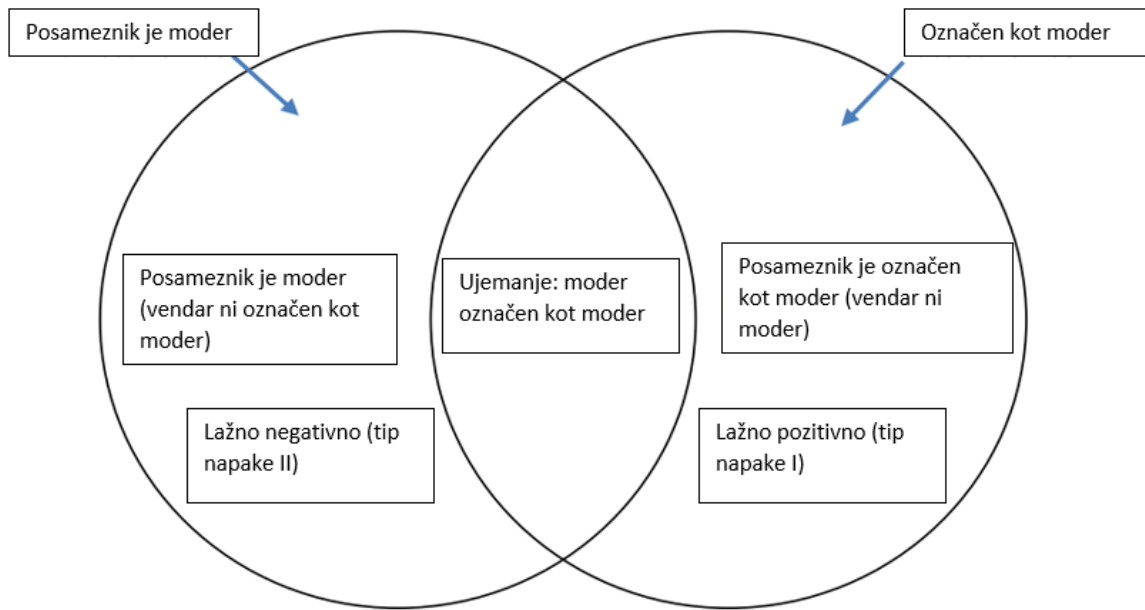
Pride lahko do (ne)namernih posledic. Z vidika posameznika glede fizične varnosti. Primer so avtonomna vozila, ki zaradi nepravilnosti lahko povzročijo poškodbe in smrt, zanašanje na napovedno-vzdrževanje opreme vodi do poškodb delavcev ali pa modeli strojnega učenja nepravilno napovejo zdravstvene pogoje. Glede zasebnosti in ugleda se lahko uporabljajo podatki brez posameznikovega dovoljenja in niso varno shranjeni. Izkrivljanje posameznikovih podatkov lahko vodi do digitalnega obrekovanja. Zaradi slabih finančnih predlogov lahko pride do slabega upravljanja sredstev. Možno je lažno predstavljanje s pomočjo strojne opreme in s tem kraje finančnih podatkov. Glede enakosti in poštenega obravnavanja so primeri v zavarovalništvu, kjer na podlagi rase imajo stranke nenamerno zavrnjeno hipoteko. Tudi pri algoritmih za kredit je opaziti, da upoštevajo povezave iz družbenih omrežij in na podlagi tega opredeljujejo pogoje (Cheatham, Javanmardian & Samandari, 2019).

Z vidika organizacij se algoritmi trgovanja ne morejo pravilno prilagoditi novim okoliščinam, kar vodi do finančnih izgub. Organizacije tudi postavljajo neugodne cene, ki zmotno vplivajo na potrošnikovo elastičnost cene, kar vodi do slabih proizvodnih odločitev. Glede nefinančne uspešnosti se uporabljajo algoritmi pri zaposlovanju, ki spet nezavedno vodijo v neraznoliko delovno silo in nenamerno vedenje. Neoptimalne ocene sredstev in virov, zahtevanih za naravne nesreče in izredne razmere, imajo za rezultat neustrezno pripravo. Glede zakonitosti in predpisov prihaja do sodnih sporov zaradi vgrajene nenamerne diskriminacije pri odločitvah o posojilih in pa v primeru razkritja zdravstvenih podatkov potrošnikov. Ugled se zmanjšuje s pomanjkanjem jasnosti v povezavi z zasebnostjo podatkov potrošnikov in v primeru oglaševalskih algoritmov, ki uporabljajo osebne identifikacijske podatke potrošnikov, ki bi lahko identificirali določenega posameznika (Cheatham, Javanmardian & Samandari, 2019).

Z vidika družbe so posledice glede državne varnosti s tem, ko akterji z zlonamernimi dejanji prevzamejo izdelke, podprte s sistemi umetne inteligence, in jih uporabljajo za nezakonite dejavnosti ali pa v primeru kršitve občutljivih podatkov. Na gospodarsko stabilnost imajo vpliv avtomatizirani algoritmi trgovanja, ki povečujejo nestanovitnost finančnih trgov, algoritmi, ki ustvarjajo nestabilnost valutnih trgov in s tem manjše trgovanje in finančni instrumenti črne škatle, ki vodijo v nenamerna sistemska tveganja. Na politično stabilnost ima vpliv manipulacija nacionalnih procesov. Infrastruktura celovitost pa se odraža glede na tveganja v povezavi s procesi in odločitvami, ki se povezujejo med seboj, in prekomerne ali zlorabljene uporabe infrastrukture zaradi inteligentnih sistemov (Cheatham, Javanmardian & Samandari, 2019).

Algoritemske napake se delijo v dve kategoriji: napake z vidika kategorij in procesov. Pri prvi skupini algoritmi obdelujejo velike količine podatkov in s tem ustvarjajo dve podskupini: lažno pozitivne (na primer, ko je nekdo kategoriziran, da ima raka, pa ga v resnici nima) in lažno negativne (na primer kategoriziranje nekoga, da ni kriminalca, a v resnici je). Pri drugi skupini algoritmi naredijo napake pri načinu odločanja. Razvijalci algoritmov določijo vlogo algoritma v odločitvenem procesu in koliko vpliva ima posameznik na to (Martin, 2019).

Slika 2: Tipi napak pri odločanju z algoritmi



Prerejeno po Martin (2019).

### 2.2.1 Nasprotujoča si stališča

Algoritmi so ustvarjeni s strani znanstvenikov, katerih interesi niso vedno pozitivni. Z njimi se bo povečal digitalni razkorak in razlike v premoženju, saj bo vpliv na revne in neizobražene negativen. Zaradi nepreglednosti bo oteženo vzpostaviti nadzor in preprečiti diskriminacijo. Vse več bo tudi rudarjenja podatkov (Rainie & Anderson, 2017). Ena izmed napovedi je v smeri, da bo umetna inteligenca v celoti prekosila človeka do leta 2050 in s tem bo prišlo tudi do večje možnosti za zlorabe. Na družbi je torej, da jo razvije v najboljši scenarij, in ne v nekaj, česar ne bo mogoče nadzorovati (Kojić, 2018).

Pozitivni vpliv se bo kazal za organizacije, ki se bodo lahko izognile tveganjem in stroškom. S pomočjo umetne inteligence se zmanjša število človeških napak, odlično se izkaže pri tveganih nalogah, saj lahko opravi delo namesto človeka, je na voljo ves čas ter ga razbremeni pri vsakdanjih nalogah in lahko tako človek porabi čas za ustvarjalne naloge. Z njo so v stalnem pogonu novi izumi, ki bodo pripomogli k rešitvam težavnih problemov, hitreje so sprejete odločitve in opravljene naloge. V današnjem času je tudi že težko razlikovati, ali govorimo z robotom ali človekom (Kumar, 2019).

Podjetja se bodo morala slej ko prej sprijazniti s tem, da je umetna inteligenca dvorezen meč. Primer je večje globalno segrevanje zaradi porabljene količine energije kot posledice uporabe računalnikov in mobilnih naprav pa vendar omogoča tudi s pomočjo projektov pomoč pri odkrivanju vodnih virov za območja, ki so v osnovi bolj nagnjena k suši, hitreje se identificira pridelke, ki jim grozi izumrtje, in razvijajo se trajnostna mesta. Tudi v avtomobilski industriji so ljudje sprva mislili, da ni pametno naprej razvijati motornih vozil,

ko se je zgodila prva prometna nesreča s smrtnim izidom leta 1896, a je človeku skupaj s tehnologijo uspelo najti rešitve za napredek v mobilnosti (Ravindran, 2020).

Raziskava Pew Internet Trust je pokazala, da je 38 % anketiranih mnenja o prevladi pozitivnih učinkov nad negativnimi, 37 % je nasprotnega mnenja in 25 % se jih je opredelilo kot izenačen vpliv (Warnock, 2019). Po podatkih raziskave Pew Research Centra je 53 % anketiranih po vsem svetu imelo pozitiven pogled glede preteklega razvoja umetne inteligence kot tudi uporabe računalniških sistemov namenjenih posnemanju človeškega vedenja, medtem ko jih je 33 % mnenja, da je bilo to slabo za družbo. Tudi glede robotov, namenjenih avtomatizaciji delovnih mest, jih 48 % meni, da je bila to dobra uvedba, medtem ko se 42 % s tem ne strinja. Če pogledamo bolj razdrobljeno, lahko vidimo, da je azijska javnost še posebej optimistična glede razvoja umetne inteligence in so tako tudi marsikateri kraji v Aziji svetovni voditelji v umetni inteligenci, medtem ko so v Evropi pogledi pesimistični z izjemo Švedske in Španije. Podobno je tudi glede avtomatizacije delovnih mest, kjer je spet v azijski javnosti optimističen pogled, kar se odraža tudi v napredku razvoja robotike in umetne inteligence. Glede na spol so moški veliko pogosteje pozitivnega mnenja glede razvoja umetne inteligence. Z vidika izobrazbe se pozitiven pogled večja z višjo izobrazbo. Tudi starost je eden izmed pomembnih dejavnikov, saj imajo mlajši bolj pozitiven pogled kot starejši posamezniki (Johnson & Tyson, 2020).

Opravljen raziskava podjetja The Economist Intelligence Unit je pokazala, da če vlada poskrbi za vlaganje v izobraževanja o delovanju algoritmov in regulacijo podatkovnih baz, to privede do sodelovanja med ljudmi in umetno inteligenco na delovnih mestih. V nasprotnem primeru, če vlada dopusti prosto pot, se poveča možnost, da umetna inteligenca prevzema delovna mesta (Šprohar, 2019). Anketna raziskava Svetovnega gospodarskega foruma, opravljena med 20.000 ljudmi v kar 27 državah, je pokazala, da je večina menila, da bi bilo treba zagotavljati večji nadzor pri uporabi umetne inteligence. Ugotovili so, da se zaskrbljenost glede uporabe sistemov umetne inteligence razlikuje med spoloma in starostjo, vpliv imajo tudi dohodki in izobrazba. Sodeč po anketi je v večini kar 40 % zaskrbljenih, podoben delež 32 % je neodločenih, le 27 % pa jih razvoj umetne inteligence ne skrbi. Glede regulacije podjetij, ki uporabljajo umetno inteligenco, je skoraj polovica anketiranih menila, da bi jo bilo treba uvesti, 20 % ljudi pa, da strožja pravila niso potrebna. Glede vladnega sprejetja je 60 % ljudi menilo, da zaostritve niso potrebne (Charlton, 2019).

### 2.2.2 Izguba človeške presoje

Z algoritmi je tehnološkim korporacijam in vladam omogočeno zbiranje, shranjevanje, razvrščanje in analiziranje ogromnih naborov podatkov. Prvotni namen teh algoritmov je v tem, da bi optimizirali učinkovitost kot tudi dobičkonosnost in pri tem izvzeli družbene vplive. Ljudje so v tem procesu kot nekakšen vložek, s čimer se ustvarja družbo, ki temelji na logiki, ki spodbuja pisanje algoritmov, s tem pa ljudje izgubljajo vlogo odločanja (Rainie & Anderson, 2017). Tako je vse večji poseg algoritmov v naloge, ki so bile zasnovane za

človeško presojo, ravno tako se s težavnostjo nalog povečuje zanašanje na njih. V metaanalizi, ki je vključevala 136 študij, so bili algoritmi za 10 % natančnejši od človeške presoje (Bogert, Schechter & Watson, 2021).

Algoritmi vplivajo na tretjino sprejetih odločitev na Amazonu in več kot 80 % na Netflixu, kar 70 % časa na YouTubu prav tako določajo algoritemska priporočila. Algoritemske odločitve imajo vpliv na več področjih. Na primer pri posojilih sprejemajo odločitve o odobritvi hipoteke. Pri procesu zaposlovanju določijo, koga povabiti na razgovor. V sodstvu določajo verjetnost ponovitve kaznivega dejanja. Tudi na področju medicine je vse več personalizacije zdravljenja. Algoritmi pa niso samo v pomoč, temveč tudi delujejo samostojno, kar je mogoče videti pri avtonomnih vozilih. Pri tem imamo ljudje sicer svobodno voljo na neki način, vendar nas hkrati algoritemska priporočila usmerjajo, čeprav mislimo, da ne (Knowledge at Wharton Staff, 2019).

Sistemi odločanja so v pomoč pravosodnim organom tako v predkazenskem postopku kot pri izbiri in odmeri sankcije ter pri odločitvi pogojnega odpusta (Završnik, 2017). Storilcem kaznivih dejanj je lahko dodeljena drugačna kazen tako v višini kot vrsti, kljub enakemu dejanju kot posledica sodnikove osebnosti, pomanjkanja smernic in abstraktnosti zakonodaje na področju kazni. Ravno zato nekatere države vključujejo v pravne sisteme točkovno ovrednotenje okoliščin in seštevek uvrstijo v preglednico, ki določa vrsto in višino kazni. Ta način je v današnjih časih avtomatiziran s pomočjo računalniškega programa. Vendar se pojavljajo dvomi glede pravičnosti, transparentnosti in sistemske moralnosti glede potrebe po sodnikih. Izrekanje kazni so vnaprej določeni koraki sodnika (Križnar, 2019).

Četudi je tradicionalno veljalo, da je vse odločitve sprejemal sodnik, je dandanes njegovo delo v nekaterih segmentih v ZDA nadomestilo 60 različnih vrst računalniških algoritmov. Eksperimentalna računalniška oprema je z 79 % natančnostjo odločala tudi v Evropi ter v 70 % na najvišjih sodiščih. Odraz uporabe programov je, da so kaznjencem, podvrženim ponovitvi kaznivega dejanja, že prej dodeljevali daljše zaporne kazni. Računalniški program tako glede na dostopne informacije s pomočjo različnih komponent ocenjuje stopnjo ponovitve nevarnosti. Te komponente so lahko dinamične in se lahko spremenijo, na primer starost in zaposlitev, ali statične in so neločljivo pripisane določenemu posamezniku na primer spol. S tem nudijo sodniku oceno oziroma stopnjo verjetnosti, ki mu je v pomoč pri izbiri in višini kazni. Kot najzanesljivejše se je izkazalo strojno učenje, saj nudi oceno dejstev v prihodnosti in je mnogo učinkovitejše, kot kar bi sodnik drugače uporabil. Omogočena je boljša optimizacija glede odločanja kot prilagodljivost na nova stanja (Križnar, 2019).

Čeprav je eden izmed razlogov za uporabo algoritmov v sodstvu ravno zmanjšanje pristranskosti sodnikov, pa se ta problem še vedno pojavlja, saj algoritme programirajo ljudje ter se učijo iz realnih vzorcev iz preteklosti. Študija iz leta 2016, ki jo je objavil ProPublic, je pokazala primerjavo algoritmskega programa za izrekanje kazni Compas in dejanskega povratništva: temnopoltim obtožencem je bila napovedana večja verjetnost kot

belcem in da pristranskost algoritma ni bila ustavljena (Huš, 2019). Prisoten je tudi strah, povezan z izgubo delovnih mest sodnikov, vendar je glede na trenutno situacijo razvoja umetne inteligence ta nepotreben. Vseh znanj in izkušenj trenutna umetna inteligenca ne more doseči, kvečjemu je večja možnost vzporedne pomoči sodnikom pri odločanju (Podpečan, 2019). Glede na to, da algoritmom manjka človeški dejavnik, bi bilo treba za namen pomoči sodnikov razumeti zasnovo, če si želimo objektivno odločitev. Vse večja je uporabnost računalniških algoritmov za oceno dejanskega stanja, analizo sodne prakse kot tudi koraka k večji enotnosti. Velika prednost razvoja se kaže pri uporabi za tipizirane pravne in upravne postopke (Drevenšek, 2018).

### 2.2.3 Pristranskost in diskriminacija

Digitalne tehnologije naj bi rešile družbene probleme s pomočjo velikega podatkovja (angl. big data) kot tudi s samoučečimi se algoritmi, ki so pravzaprav srčika umetne inteligence (Mayer-Schönberger & Cukier, 2013). Tako v današnjem času ni več statičnosti algoritmov, saj njihov razvoj poteka v smeri samodejnega spreminjanja vedenja, ki bo morda presegel ljudi ter s tem prav tako ustvarjal morebitno pristranskost (Michael, 2019).

Umetna inteligenca je naredila korak s časom tudi na področju zaposlovanja. Zaposlitveni algoritmi pripravijo celovito podobo prosilca zaposlitve tako z vidika osebnostnih lastnosti kot strokovne primernosti in pri tem lahko prihaja do krivic (Završnik, 2017). Zmožna je spremljati čustva kandidata in oceniti, ali je njegova osebnost primerna za tovrstno delo. Prednost takšnega načina zaposlovanja je zagotovo v udobju, saj imata obe strani prilagodljiv čas tako poteka kot pregleda intervjuja. Pri množičnem zaposlovanju se spet takšen pristop pokaže kot koristen. Vendar po drugi strani je razpoložanje spremenljiva stvar, kot tudi obstajajo fizične težave, ki imajo lahko vpliv na mimiko obraza (Zetlin, brez datuma). Pri avstrijski agenciji za zaposlovanje so leta 2016 začeli program, s katerim bi ugotovili dejavnike, ki v največji meri kažejo na možnost zaposlitve posameznika. Dokument je razkril, da je prišlo do diskriminacije, zato je bil razglašen kot nezakonit sistem, čeprav povečuje učinkovitost z vidika manjše porabe sredstev za ljudi, ki jim na koncu ne bi koristili. Algoritem je ženskam dodelil negativno težo, ravno tako tudi invalidom in osebam, starejšim od 30 let. Za primerjavo velja omeniti, da so bile tudi ženske z otroki negativno prepoznane, medtem ko moški ne (Kayser-Bril, 2019a).

Algoritmi se zanašajo na podatke in na podlagi tega je razširjeno prepričanje o njihovi objektivnosti. Vendar algoritmi so prvotno oblikovani s strani ljudi, kateri so izbrali podatke in jih izučili. Tako se neizogibno širi pristranskost ljudi v modele umetne inteligence in se odraža kot krepitev človeških predsodkov. Primer je, če v iskalniku Google Images iščemo izraz glavni izvršni direktor (angl. Chief Executive Officer, v nadaljevanju CEO) nam prikaže le 11 % žensk, čeprav jih je na tem položaju 27 % v Združenih državah. Izziv človeštva je v pomanjkanju skupnega razumevanja morale, saj se preference držav zelo

razlikujejo. Vprašanje, ki se tu poraja, je, kako naj na primer avtonomno vozilo sprejme pravo odločitev, ki bo uskladila vse medkulturne etične variacije (Michael, 2019).

Povzročijo lahko katastrofalne posledice med drugim smrt, razizem in denarne izgube. Po mnenju strokovnjakov algoritmi sami po sebi niso zlobni, temveč je to rezultat nepredvidnosti pri oblikovanju. Pristranskost je samodejno vnešena zaradi podatkov iz preteklosti, ki so jih izbrali ravno ljudje in jih analizirali. Tako so primeri škode na primer visoke denarne izgube podjetij zaradi napačnega testa programske opreme in algoritma, strmoglavljenje letala zaradi napačnih signalov in podobno. Klepetalni robot Tay podjetja Microsoft je v 24 urah postal rasističen s svojimi izjavami, algoritmi namenjeni prepoznavi obrazov, so nedolžnim ljudem na letališčih prisodili, da so kriminalci, pristranskost se je pojavila tudi na lepotnem tekmovanju Beauty.AI, kjer so želeli s pomočjo mnenj ljudi z vsega sveta videti, kdo bi bil najboljši. Kmalu so ugotovili, da je algoritem povezoval lepoto z barvo kože (Hebbar, 2017).

#### 2.2.4 Zlorabe

Odrptih etičnih dilem je veliko kot tudi možnosti za zlorabo. Postavitev meja je zato ključna ne samo z vidika osebnih podatkov, temveč tudi odgovornosti za nastale situacije. Zaradi hitrejše in učinkovitejše obdelave podatkov s pomočjo algoritmov strojnega učenja v primerjavi s človekom se zanašanje na umetno inteligenco povečuje. S tem pa se povečujejo tveganja, kar se še posebej opazi pri odločitvah, ki jih človek ne more nadzorovati. Tako bo v primeru, da podatki odražajo pristranskost, tudi odločitev takšna. Z avtomatiziranim sprejemanjem se krepijo socialne razlike (Bird in drugi, 2020).

Tehnologija lahko povzroči digitalne, fizične in politične grožnje. Da bi se zaščitili pred zlonamernimi akterji, je treba stremeti k razumevanju tveganj in prepoznavi le-teh. Najpogostejša zloraba je uporaba tehnike umetne inteligence, ki izdelava oziroma manipulira tako zvok kot vizualno vsebino pri tem pa je pristen rezultat, ki ga je težko ločiti od legitimnih vsebin. Ta vsebina lahko v današnjih časih nato doseže milijone posameznikov v izredno hitrem času. Ljudje se morajo zavedati, kako realistično izgleda in kateri so možni načini zlorabe, saj privedejo do uničenja imena in statusa posameznika, privabljajo k nakazovanju denarja in deljenju pomembnih informacij, ponarejanju spletnih potrditvenih sistemov, ki so namenjeni potrjevanju identitete, prav tako manipulaciji elektronskih dokazov pri kazensko-pravnih raziskavah, možna je tudi manipulacija trga delnic, podpira pa se lahko tudi dezinformacijske kampanje, družbene nemire in politične polarizacije (Trend Micro, 2020).

Druga kritična zloraba je izboljšanje algoritmov, ki so namenjeni ugotavljanje gesel in s tem omogočajo vdore kibernetским kriminalcem. Že zdaj so v uporabi pristopi, ki primerjajo različne različice gesla, bo pa z uporabo nevronske in globalne omrežije omogočena kibernetским kriminalcem analiza obsežnih naborov podatkov in ustvarjanje različic gesel. Zlorabe so tudi na področju posebitve ljudi in ustvarjanju lažnih računov (Trend Micro, 2020). Tako z moralnega kot varnostnega vidika predstavljajo skrb ubijalski droni. Še ena



izmed skrbi je tudi nadzor, ki ga lahko poleg potencialno pozitivnih učinkov zlorablja avtoritativne vlade. Tisti, ki imajo moč že zdaj, jo bodo lahko samo še stopnjevali. Večje je tudi tveganje glede diskriminacije kot tudi pristranskosti zaradi podatkov, ki jih uporablja tehnologija, in le-ti temeljijo na ljudeh (Castelvecchi, 2019). Biometrične tehnologije, ki med drugim služijo za prepoznavanje obrazov, se vse več uporabljajo tudi za neomejeno sledenje ljudi, kar je sporno z vidika zasebnosti kot tudi varstva podatkov (United Nations, 2021).

Primer zlorabe algoritma je objavljen posnetek leta 2017 raziskovalcev univerze v Washingtonu, na katerem je govoril Barack Obama, vendar obraz ni bil njegov. To jim je uspelo narediti na podlagi dostopnih videoposnetkov in sestave prepričljivega vizualnega modela za namen reprodukcije njegovega javnega nastopanja. Glede na neverjetno dobre rezultate bo z razvojem strojne opreme to še veliko lažje in hitrejše. Problem je v dostopnosti tehnologije, ki se tako lahko uporablja v maščevalne namene. Tudi v politiki lahko pride do napačnih informacij in preplaha. Glede na to, da smo v začetnem obdobju nove dobe, je še čas za okvire etičnega delovanja in predvsem, da inženirji in snovalci niso tisti, ki določajo uporabo, temveč celotna družba (Savič, 2018).

Podatki glede na poročilo IDG State of the CIO 2021 razkrivajo, da je 39 % vodilnih IT-jevcov mnenja, da bo prav analitika podatkov privedla do največ naložb v njihovih organizacijah. To dokazuje, da omogoča konkurenčno prednost, ampak se je treba zavedati posledic napak, ki lahko vplivajo na ugled podjetja, njegove prihodke in celo življenje ter poznati orodja katera se uporabljajo kot tudi spoštovati vrednote podjetja (Olavsrud, 2021).

### **2.3 Družbeni izzivi**

Uporaba algoritmov je prisotna že vrsto let, vendar pa je vse bolj prisotna zaskrbljenost ljudi navezujoč na pomanjkanje razumevanja algoritmov in njihovega delovanja, vse obširnejše pristranskosti ter nezadostne preglednosti in odgovornosti. S tem se večajo družbeni izzivi glede transparentnosti, zaupanja v odločanje algoritmov in zagotavljanje celovitosti kar pa je vse kompleksnejše za ustvarjalce algoritmov. Vprašanje, ki se zato poraja, je, ali resnično obstaja potreba po izdelavi algoritmov za odločanje. Če je odgovor na to pritrdilen, se je treba vprašati, na katerem področju in zakaj prav na tem. Pred samo izdelavo oziroma uvedbo je smiselno pomisliti, kaj bomo s tem dosegli oziroma katero pomanjkljivost rešili in tako skrbno načrtovati njihov razvoj (Pandya, 2019). Na to bi se lahko odzvali s povečanim javnim izobraževanjem o pismenosti delovanja algoritmov. Treba bi bilo tudi opredeliti odgovornosti (Rainie & Anderson, 2017). Družbeni izziv je tudi v prepričanju glede vpliva umetne inteligence na zmanjšanje zaposlovanja, vendar le na delovne kategorije, ki so bile spremenjene že pred tem z vsakim tehnološkim valom; tako se ustvarjajo nove. Raziskave kažejo, da umetna inteligenca ne bo mogla v celoti nadomestiti človeškega dela, temveč bodo s pomočjo nje podjetja lahko človeško silo usmerila v delo z višjo vrednostjo (Walch, 2019).

### 2.3.1 Potreba po večji transparentnosti

Pravica uporabnika je razumeti vpliv algoritmov, vendar težava nastane pri samem razumevanju, ki je precej težavno, saj je odvisno od zahtevnosti algoritma, toka obdelanih podatkov ter števila uporabnikov, ki si želijo razlago. Ta je v primeru strukturiranih podatkov mnogo preprostejše kot pa nestrukturiranih, kar velja za globoko učenje. Potreba po tem, da bi razumeli sprejemanje odločitev umetne inteligence, je odvisna od kritičnosti odločitve (Nielly, 2020).

Preglednost algoritmov igra ključno vlogo, če želimo jasen potek odločanja kot tudi, na podlagi katerih informacij je bilo nekaj oblikovano. Tehnična preglednost omogoča razkritje izvorne kode kot tudi vhodov in izhodov samega algoritma. S tem se povečuje zaupanje. Vendar izvorna koda ni vedno razkrita, saj se lahko šteje za intelektualno lastnino. K preglednosti bi pripomoglo dokumentiranje razvoja in delovanje algoritma ter določitev ljudi, ki bi do evidenc imeli dostop. Organizacije bi morale omogočiti preglednost okvirov upravljanja informacij za tiste, ki si to želijo (Sherman, brez datuma). Preglednost bi omogočil odprtokodni algoritem, vendar če relevantno občinstvo kljub deljenju vseh postopkov, kod in dokumentacije, ne razume podanih informacij, ne moremo govoriti o preglednosti (Watson, 2019).

Na transparentnost vpliva več dejavnikov. Med drugim so to organizacijski ukrepi, pri katerih je kot posledica zbiranja, shranjevanja in analize velikega števila podatkov podjetjem omogočen dostop do širše uporabe algoritmov in avtomatizacije odločitev. Vpliv imata tudi dobiček in stroški z vidika izboljšanih priložnosti na področju storitev za stranke, večje hitrosti ter zmanjšanja stroškov z uporabo algoritmov. Podjetja, ki poslujejo z etičnim pristopom, imajo zagotovljen uspeh na dolgi rok. Hkrati pa je treba poudariti stroške v povezavi z razvojem in upravljanjem algoritmov. Če se podjetje ne drži zakonov in regulacij, lahko sledi kazen in škoda podjetju. Pomemben dejavnik so nove tehnologije, ki so povečale možnosti zajema, analize in uporabe podatkov. Ozaveščanje javnosti in odziv igra tudi ključno vlogo, saj javnost ne razume v celoti delovanja in imajo pomisleke glede vrste osebnih podatkov, ki se zbirajo, analizirajo in uporabljajo. Prav tako le 51 % ljudi po raziskavi Pew Research Centra meni, da bi morala biti tehnološka podjetja bolj regulirana, kot so sedaj. Šele ko algoritmi negativno vplivajo na ljudi, se njihovi pomisleki krepijo. Vpliv imajo tudi etična, pravna in regulativna vprašanja, pri katerih tehnološka podjetja bolje sodelujejo glede razvoja zakonov in regulacij (Watson, 2019).

The Association for Computing Machinery (v nadaljevanju ACM) je določilo 7 načel algoritemske odgovornosti v kontekstu transparentnosti. Vključujoč zavedanje ljudje o deljenju podatkov, ki jih zbirajo, analizirajo in uporabljajo podjetja. V primeru napake zaradi slabega razvojnega procesa je pomembna dostopnost in odškodnina. Prenos krivde izključno na algoritem ni pravi način. Za povečanje zanesljivosti je odgovornost za ustvarjene sisteme na strani podatkovnih znanstvenikov, managerjev in podjetij. Potrebna je razložljivost posamezniku, ki mu je bila povzročena škoda. Organizacije morajo vedeti, kako pojasniti

proces, ki jih ustvarjajo, testirajo, posodablajo in uporabljajo algoritmi za povečanje kredibilnosti. Izvoru podatkov z vidika nenamernih pristranskosti, vnesenih v modele s strani razvijalcev, je treba posvetiti čas že pri samem ustvarjanju in testiranju modelov. Pregled podatkov o tem, kako organizacije razvijajo, uporabljajo in vzdržujejo algoritme, bi moral biti omogočen notranjim in zunanjim deležnikom. V povezavi s potrjevanjem in testiranjem bi morale organizacije sprejemati odločitve po najvišjih standardih glede na različne dejavnike in spremljati delovanje algoritma ter vzpostaviti modele, ki bodo beležili neugodne reakcije (Watson, 2019).

31. mednarodna konferenca za varstvo podatkov in zasebnosti je bila tista, na kateri se je priznala pomembnost transparentnosti. Leta 2012 se je tudi vključila še v takratni predlog GDPR in je dandanes zapisana v členu. Vključena je perspektivna preglednost, ki zagotavlja obveščanje o tekoči obdelavi podatkov kot tudi zahteva od organizacij, ki obdelujejo podatke obveščanje o tem, na katere osebne podatke se nanaša z razumljivim jezikom. Prav tako tudi retrospektivna, ki se nanaša na post hoc pojasnila in omogoča, da ima posameznik pravico do pridobitve razlage odločitve in izpodbijanja (Felzmann, Fosch Villaronga, Lutz & Tamò-Larrieux, 2019).

Vlada Združenega kraljestva je naredila korak naprej, saj je v poskusno izvajanje uveljavila nacionalni standard, namenjen algoritemski preglednosti. S preučitvijo vloge algoritmov pri odločanju želi vlada graditi na zaupanju javnosti. Standard vključuje opis orodja in namen ter način uporabe kot tudi podrobne podatke o samem delovanju, uporabljenih podatkih in vlogi nadzora človeka (Open Access Government, 2021).

Pozivi k preglednosti naraščajo, vendar popolna preglednost bi pomenila, da bi ljudje svoje podatke usmerili tako, da bi algoritem nanje boljše gledal in bi s tem pravzaprav bili del sistema. Preglednost bi bila smiselna le za določene namene oziroma odločitve. Torej pomembno je omogočiti obseg preglednosti (Martin, 2018).

### 2.3.2 Potreba po rasti algoritmične izobraženosti

Za zagotovitev uspešne umetne inteligence, ki bi ji ljudje zaupali in bi bila hkrati etična, je zelo pomembno razumevanje sprejemanja odločitev računalnika kar lahko dosežemo z izobraževanjem o algoritmih. Smiselno je ljudem omogočiti svetovanje o tehnični razvitosti kot tudi o zaupanju in etiki. Podjetja morajo poznati proces celotne razvojne verige od delovanja modela umetne inteligence in tudi, katere podatke potrebuje za svoje delovanje in o kasnejšem posodabljanju (De Jonghe & Latinne, 2019).

Ljudje v osnovi precej bolj zaupajo ljudem, četudi se motijo kot pa umetni inteligenci. Njihova stopnja zaupanja se lahko izboljša z izkušnjami in uporabo kot tudi z večjo transparentnostjo delovanja. Tako bi ljudje lažje razumeli algoritemske odločitve. Nekatera podjetja so že ubrala pot k temu z objavljenimi poročili o vladnih zahtevah kot tudi nadzoru. Povečanje zaupanja je možno z večjo vključenostjo v proces odločanja in hkrati učenjem

umetne inteligence iz človeških izkušenj. Če so ljudem omogočene manjše spremembe algoritma, je po študiji tudi delež zadovoljnih s kasnejšimi sprejetimi odločitvami večji. Poveča se zaupanje v izboljšave z uporabo algoritmov v primerjavi s človekom kot tudi večja je možnost uporabe v prihodnosti. Ljudem natančne informacije o samem delovanju algoritmov ne bi kaj dosti pomagale, če se ne spoznajo na procese pa vendar, če bi jim bile omogočene delne informacije o njihovem delovanju in nadziranju teh sistemov pri izvajanju torej, da bi se algoritemsko izobrazili, bi zagotovo lažje sprejeli umetno inteligenco (Polonski, 2018).

Zaupanje v umetno inteligenco se manjša z nepravimi predstavami glede njene sposobnosti, kot tudi ena izmed pomembnih stvari je razrešitev nejasnosti o vplivu umetne inteligence na delovna mesta. V splošnem so vodje umetne inteligence enotnega mnenja glede izobraževanja ljudi o potencialnih motnjah kot tudi ključnih veščinah pri opravljanju novih delovnih mest, ki jih bo ustvarila umetna inteligenca. Že v šolah pa bi bilo treba povečati miselnost o vseživljenjskem učenju (IBM, brez datuma b).

Z vse večjo prisotnostjo umetne inteligence prihaja do vse večjega problema glede razumevanja sistemov umetne inteligence. Ljudje praktično ne vedo, zakaj so bile določene odločitve sprejete. Še večji problem je pri omejenem razumevanju ne samo povprečnih ljudi, temveč tudi raziskovalcev in ustvarjalcev umetne inteligence. Gre pravzaprav za neobičajen pojav, hkrati pa bi lahko rekli, da je to njena moč, saj ljudje ustvarjajo nekaj, kar se je sposobno kasneje samousposobiti in izvajati naloge, ki presegajo človeško zmogljivost. S tem se ustvarja vse večja vrzel. Javnost ima z etičnega vidika pravico vedeti način delovanja umetne inteligence kot tudi možnost zahteve po pojasnilu. Ne zahtevajo pa vsi primeri uporabe nujno tudi razlago, ker enostavno niso dovolj pomembni. Nasprotno pri primerih, kjer ni moč preverbe pravilnega delovanja, je pomembno razumevanje delovanja. Obstaja tudi že področje umetne inteligence, ki se imenuje razložljiva umetna inteligenca. Gre za splošne tehnike, ki so uporabne pri kakršnih koli modelih strojnega učenja in ni nujno, da so informacije na koncu uporabne (Xu, 2021).

### 2.3.3 Potreba po večjem nadzoru

Negativni učinki algoritmov se kažejo pri specifičnih programih za kriminaliteto, kjer je prisotna diskriminacija živečih na določeni geografski legi oziroma pogojno obsojenih, pri katerih je prišlo do problema rasne pristranskosti. Kazalnik negativnih učinkov je tudi v negativnih zadetkih, ki v določenih primerih vplivajo na posameznikove pravice. Ta problem nima vpliva na človekove pravice na primer pri oglaševanju, medtem ko pri seznamu prepovedi letenja letalskih družb ima. Netransparentnost algoritmov in njihovega načina odločanja je prav tako eden izmed negativnih učinkov, saj če želimo priti do podatkov, so le-ti na primer pravno zaščiteni. Posledica algoritmične napovedi je začaran krog, ki tvegane skupine daje v ospredje nadzora, a hkrati ravno na ta račun se kazniva dejanja kažejo predvsem pri njih. Primer so policijski programi, ki oblikujejo napovedi. Eden najbolj

kritičnih pa je zagotovo vpliv na varstvo osebnih podatkov kot tudi zasebnost (Završnik, 2017).

Družbeno sortiranje, ki je vidno predvsem v bančništvu in zavarovalništvu s tem, ko posamezniki ne vedo, da je bila določena odločitev sprejeta na podlagi specifičnih podatkov in kakšni so ti podatki, ki vplivajo na njihove odločitve. Njihova vloga je oblikovanje personaliziranih storitev, s tem so mišljene zavarovalne premije, ki so usklajene z gensko sliko posameznika in njegovega življenjskega sloga, kot tudi obrestne mere, ki so zopet prilagojene finančnemu tveganju posameznika (Završnik, 2017). Poleg tega so pri oceni kredita, ki uporablja umetno inteligenco, opazili pristranskost z vidika spola, kjer so imeli moški več možnosti tako za odobritev kot tudi glede višine. Vendar kljub svoji nepopolnosti nudi pomoč pri kreditnih odločitvah z uporabo netradicionalnih metod. Za kreditne točke so uporabljali podatke navezujoče na identiteto, bančne transakcije, zgodovino kreditov, zavarovanje in prihodke. V netradicionalnem pogledu pa umetna inteligenca skupaj s strojnim učenjem uporablja vire podatkov, ki jih pridobita iz družbenih medijev, posnetkov dronov in satelitov ter vseh podatkov, ki so dostopni kjerkoli. S tem je ocena tveganja mnogo bolj izvedena in sposobnost odplačevanja. Podjetja so s tem tudi učinkovitejša in lažje obvladujejo kreditna tveganja (Pandey, 2022).

Raziskovalni projekt je ugotovil več kot 95 % natančnost sprejetih odločitev s strojnim učenjem glede na zaposlenega v banki, kar pomeni, da niso boljši kot ljudje, vendar so sposobni priti do enakih zaključkov kot ljudje. Razlog se skriva tudi v predvidljivem odločanju ljudi (Anderson, 2022). Ob pravilni izvedbi predstavljajo priložnost za vse deležnike, saj so s tem preučene vse razpoložljive možnosti glede na okvir omejitev in opredelitve, kaj želijo doseči (Floridi & Lee, 2020).

Ključne lastnosti, ki pripomorejo k zmanjšanju tveganj, povezanih z algoritmičnim odločanjem, se delijo na notranje, med katere štejemo pravičnost, nepristranskost, nediskriminacijo in se lahko izražajo kot lastnosti samega algoritma v kontekstu njegove uporabe, in zunanje, ki vključujejo razumljivost med vhodnimi in izhodnimi podatki, ki jih da algoritmično odločanje in dve glavni obliki tega sta transparentnost in razložljivost (Castelluccia & Le Métayer, 2019).

Tehnične težave so lahko z vidika varnosti predvsem pri vpetosti algoritmičnih odločitvenih sistemov v fizične sisteme, saj lahko povzročijo škodo. Številne napake so lahko odpravljene z ad-hoc rešitvami, ampak še vseeno je treba opredeliti enoten pristop k temu problemu. Testiranje in vrednotenje bi morala postati osnovna zahteva in zagotovitev ustrezne odgovornosti. Druga težava je v ogrožanju celovitosti in razpoložljivost, saj ko se uporabljajo v kritičnih kontekstih, je potrebna zaščita pred zlonamernimi dejanji. Upoštevati je treba varnostne lastnosti, ker se večina algoritmičnih odločitvenih sistemov zanaša na strojno učenje. Pri zaupnosti in zasebnosti gre spet za ogrožitev z napadi na algoritmične odločitvene sisteme in rešitve so v smeri anonimizacije usposabljanja nabora podatkov in generiranih modelov ali pa s porazdelitvijo faze učenja na način, da podatki ne zapustijo

naprave, ki jih zbira. Seveda pa te rešitve še niso razvite. Nenazadnje pa tudi poštenost, ki je problematična z vidika potencialnih možnosti virov nepravičnosti med procesom, saj algoritmični odločitveni sistemi temeljijo na strojnem učenju algoritmov in so usposobljeni tako, da zbirajo podatke. Nepoštenost lahko izhaja iz vsebine podatkov, načina označevanja podatkov ali prihodnjega izbora (Castelluccia & Le Métayer, 2019).

Pri razložljivosti lahko sledimo trem pristopom:

- pristop črne škatle,
- pristop bele škatle,
- konstruktivni pristop.

Pristop črne škatle (angl. The black box approach) pomeni, da analizira obnašanje algoritmičnih odločitvenih sistemov brez poznavanja njegove kode. Je edini pristop, kjer operater ali ponudnik algoritmičnih odločitvenih sistemov ne sodeluje ter je razlaga samo iz vidika opazovanja odnosa med vhodi in izhodi sistema. Pristop bele škatle (angl. The white box approach) v nasprotju s prejšnjim predpostavlja, da je analiza kode algoritmičnih odločitvenih sistemov možna. Konstruktivni pristop v primerjavi s prvima dvema predpostavlja, da algoritmični odločitveni sistem že obstaja in je mišljen k oblikovanju na način, da zahteve razložljivosti že upošteva. Tehnične rešitve ne morejo samostojno rešiti težav v povezavi z algoritmičnimi odločitvenimi sistemi, ampak jih je treba povezati z ostalimi ukrepi še posebej z zahtevami po transparentnosti, razložljivosti in odgovornosti. Nekateri obstoječi zakoni že veljajo za algoritmične odločitvene sisteme in v določeni meri obravnavajo zgornje zahteve (Castelluccia & Le Métayer, 2019).

V prihodnosti je pričakovati kriminalce, ki bodo umetno inteligenco izrabljali na vse možne načine. Kibernetski napadalci bi lahko še povečali svoje napade kot tudi obseg. Umetno inteligenco bi lahko uporabili s pomočjo socialnega inženiringa. Tako lahko prve korake napada avtomatizirajo in ogrozijo zaposlene ter poslovne procese. S tem so goljufije podjetja še hitrejša in bolj natančna, prihaja lahko do lažnega predstavljanja in prevar pri e-pošti. Tudi na področju kriptovalut lahko pride do zlorab s pomočjo robotov umetne inteligence, ki se naučijo strategij trgovanja. Ne izključuje se škoda posameznikom med drugim tudi fizična. Droni s prepoznavanjem obraza tako lahko nosijo gram eksploziva in izgledajo kot ptice ali pa žuželke, upravlja se jih preko mobilnega interneta in so popolnoma neopazni (Trend Micro, 2020). Potrebna je pazljivost glede zlonamernih uporabnikov, ki lahko pripeljejo do hude škode. Razmislek v smeri načina gradnje in upravljanja digitalne infrastrukture ter načrtovanja in distribucije sistemov umetne inteligence je ključno sedaj, saj bo odkrivanje zlonamernih napadov težavno. Poleg tega umetna inteligenca s pomočjo avtomatizacije in inteligentnega načina spremlja dogajanje ljudi in daje vpogled podjetjem in vladam (Walch, 2019).

### **3 VLOGA ETIKE IN ETIČNI VIDIKI ODLOČANJA Z ALGORITMI**

Etika umetne inteligence razlaga, kako s primernim vedenjem razvijalcev, proizvajalcev in operaterjev minimizirati etično škodo, ki lahko izhaja iz umetne inteligence kot posledica neetičnega oblikovanja, neprimerne uporabe ali pa zlorabe. Namenov etičnega razmišljanja o umetni inteligenci je več, med drugim spodbuditi razmišljanje, ali je potrebna zaščita posameznikov in skupin že na najosnovnejši ravni. Prav tako pa tudi spodbuditi vrsto novih inovacij, ki bi pripomogle k doseganju ciljev trajnostnega razvoja in hkrati le-te spodbujajo etične vrednote. Prav zaupanja vredna umetna inteligenca lahko pripomore k pravičnejši družbi z ustvarjanjem blaginje in vrednosti kot tudi s povečanjem bogastva na načine, ki pripomorejo k enakosti porazdelitve tako gospodarskih kot družbenih in političnih priložnosti (European Commission, 2019b).

Ključnega pomena za razumevanje in raziskovanje etičnega vidika odločanja s pomočjo algoritmov je v tem, ali imajo posamezniki dostop do družbenih dobrin in pravic ter kako so algoritmi razviti in implementirani s strani razvijalcev kot tudi v kakšni meri. Če ne moremo upravljati vedenja tega, kar ustvarjamo, tudi ne moremo preverjati etičnosti. Pomembno je vgraditi etiko v prvotno zamisel razvijanja specifične tehnologije, opremljene z umetno inteligenco, kot tudi kasneje nadzirati delovanje tehnologije, da bi lahko razumeli njeno vedenje in preverili upoštevanje moralnega kompasa (Hoes, 2019).

Algoritmi so vrednostno obremenjeni kar je neizogibno, namreč operativne parametre, ki sicer niso pogojeni z etično spremenljivim vedenjem, opredeljujejo razvijalci, kasneje pa jih izoblikujejo še uporabniki s svojimi preferencami. Da bi ugotovili morebiten in dejanski etični vpliv, je potreben izredno težaven in dolgoročen postopek raziskave procesov vpletenosti človeške subjektivnosti. Namreč tudi pri zadostnih virih problem kot tudi osnovna vrednost ne bosta očitno vidna, dokler se ne pojavi težaven primer uporabe. S slabim interpretiranjem in predvidevanjem učenja algoritmov je ugotavljanje izvora problematične odločitve skoraj nemogoče oziroma zelo zahtevno. Medsebojni vpliv algoritmov omenjen problem le še pogloblja ter povečuje vrzel med samo zasnovano in kasnejšim delovanjem algoritmov in razumevanjem etičnih posledic, kar ima lahko kasnejši odraz za družbo (Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016).

Poleg obsega analize in zahtevnosti odločanja, k problematiki etičnosti prihaja tudi zaradi negotovosti in nepreglednosti dela algoritmov kot tudi njihovega vpliva. Algoritmi se vse bolj zanašajo na učne zmogljivosti kot pa na ročno programiranje. Etični pomisleki, ki jih sprožajo algoritmi, so v tem, da napačni dokazi vodijo v pristranskost, medtem ko nepošteni v diskriminacijo. Rezultat nerazumljivih dokazov je nepreglednost. Za razliko od sledljivosti, ki vodi do moralne odgovornosti (Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016).

Etika se odraža v kakovosti napovedi kot tudi rezultatih in vplivu na ljudi. Etična kakovost tudi opredeljuje moralne dolžnosti umetne inteligence in njenih razvijalcev, ki so v tem, da ublaži vse škode, do katerih lahko pride z uporabo umetne inteligence. S tem je mišljena napačna predstavitev posameznika v samem sistemu strojnega učenja kot tudi vse nadaljnje odločitve, sprejete v povezavi z njim in netransparentnost sistema pred posameznikom. Na področju, kaj je umetna inteligenca, se kritična vprašanja v povezavi z etiko navezujejo na pristranskost in poštenost, odgovornost in popravljivost ter na preglednost, interpretiranje in razložljivost. Z vse večjo interakcijo med umetno inteligenco in človekom so raziskovalci prepoznali potrebo po oceni glede vpliva na varnost v primeru negotovosti, kibernetško varnost, zlonamernost uporabe kot tudi pogled ljudi na omenjeno interakcijo ter vpliv na zasebnost, kontrolo in nadzor. V povezavi z vplivom, ki jih ima umetna inteligenca, se etična vprašanja navezujejo na avtomatizacijo, izgubo delovnih mest in trende dela. Poleg tega pa tudi na demokracijo in državljanske pravice kot tudi človeške interakcije. Ne nazadnje se etična vprašanja glede tega, kaj lahko umetna inteligenca postane, navezujejo na grožnje, ki jih predstavlja za človeške kognitivne sposobnosti in s tem prevlado nad človekom (Nalini, 2019).

Glavne etične dileme in moralna vprašanja zelo vplivajo na družbo. Strah ljudi pred pomanjkanjem delovne sile je prisoten že stoletja in z uvedbo umetne inteligence se le-ta povečuje. Po študiji, ki jo je opravilo 1896 strokovnjakov, jih 52 % meni, da tehnologija do leta 2025 ne bo izrinila več delovnih mest, kot bi jih ustvarila. Drugo področje naslavlja neenakost z vidika izkoriščanja delavcev. Spremembe, ki se obetajo, bodo tudi glede kakovosti dela. Nova delovna mesta bodo lahko zahtevala visoko usposobljene delavce. Umetna inteligenca bi se morala razumeti kot korist vseh in ne kot zasebna dobrina. Problematična točka znotraj tega je koncentracija moči med elito. Rast umetne inteligence bi vodila tudi do premoženjske neenakosti in političnega preobrata (Bird in drugi, 2020).

Naslednje vprašanje naslavlja zasebnost, človekove pravice in dostojanstvo, ki je zelo pomembno pri oblikovanju robotov, saj se nahajajo v domovih. Digitalne zapise lahko iščemo z uporabo algoritmov za razpoznavanje vzorcev. Tudi prepoznavna obraza vpliva na našo identiteto kot tudi politične in ekonomske predispozicije. Strojno učenje omogoča pridobivanje informacij iz podatkov in odkrivanje novih vzorcev ter tako navidezno neškodljive podatke lahko spremeni v škodljive. Ljudje imajo pravico do tega, kako so njihovi podatki uporabljeni in hranjeni, vendar glede tega, kako so učeni modeli, pa le delno pravico, kar povzroča mnogo etičnih težav. Umetna inteligenca izvaja poleg tega nadzor z opazovanjem in napovedovanjem. S tem pa se povečuje tudi kršenje zasebnosti in človekovih pravic, saj se vsi ti podatki zbirajo in tudi delijo brez privolitve. In ne nazadnje tehnološka podjetja omenjajo umetno inteligenco kot rešitev sovražnega govora, digitalnih dezinformacij in nasilnega ekstremizma, vendar s tem krepijo samocenzuro in svobodo izražanja (Bird in drugi, 2020).

Naslednja etična dilema je pristranskost, saj je umetno inteligenco ustvaril človek. Tako so se že pojavljali primeri, kot je na primer Compas, programska oprema, delujoča na podlagi



strojnega učenja, ki je označila temnopolte za kazenske povratnike. Prav tako naj bi bili bolje plačani poklici bolj v domeni moških kot žensk. Tudi fotografije ljudi so izbrane samo iz določenih držav (USA, UK), kar vodi do različnih nesoglasij, na primer, da kamera javlja fotografu, da ima oseba zaprte oči, v resnici pa je azijske rase (Bird in drugi, 2020).

Opozoriti je treba še na demokratičnost, ki je lahko ogrožena z nepravimi novicami in družbenimi omrežji, primer je vpliv na volivce. Poleg tega je nedavno poročilo ugotovilo, da je najmanj 28 držav zaposlilo tako imenovane kibernetске skupine (angl. cyber troops) za manipuliranje javnega mnenja preko družbenih omrežij. Mediji tudi uporabljajo algoritmične predloge novic, ki so prilagojene ljudem glede na njihove interese. S tem so ljudje izpostavljeni samo njihovem stališču in tako živijo v filtriranem mehurčku, s tem pa deluje proti demokratičnosti, kjer je ključna interakcija različnih mnenj. Demokratičnost je ključna za spodbujanje inovacij, krepitev gospodarstva in ustvarjanja premožnih ljudi (Bird in drugi, 2020).

Naslednji sklop dilem se nanaša na vpliv na človeško psihologijo, saj se v prihodnosti pričakuje, da bodo roboti služili ljudem v vlogi negovanja, gospodinjstva, skrbi za otroke in starejše, poučevanja. Ti naj bi izgledali in govorili tako kot ljudje in z njimi bo mogoče oblikovati čustveno povezanost. Tukaj pa se lahko pojavi manipulacija in prevare, ker bi heker prevzel kontrolo nad robotom in vplival na človeka s tem, ko bi človek bolj zaupal robotu. Ljudje bi lahko postali psihološko odvisni od robotov, saj tehnologija, kot je znano, vpliva na dobro počutje. Ko je enkrat stroj kot subjekt zmožen zaznavanja, čutenja in delovanja, ni daleč od razmišljanja o njihovem pravnem statusu (Bird in drugi, 2020).

Naslednje etično vprašanje je glede vpliva na finančni sistem. Trgi so primerni za avtomatizacijo, saj delujejo skoraj v celoti elektronsko in ustvarjajo velike količine podatkov z veliko hitrostjo, kar zahteva algoritme. Dinamičnost pa zahteva hiter odzivni čas. Algoritemsko trgovanje tako lahko ustvarja dobiček v takšni hitrosti in pogostosti, kot je za človeka nemogoče. Obstajajo pa tudi različni načini, v katerih avtonomni finančni agenti lahko povzročijo finančna kazniva dejanja, kot na primer tržno manipulacijo in dogovarjanja (Bird in drugi, 2020).

Etično vprašanje se pojavlja glede vpliva na pravni sistem, ki je bil narejen s predpostavko, da bodo sprejemali odločitve ljudje in ne roboti. Težke in pomembne odločitve naj bi se tako preusmerjale v roke algoritmov, kar bi pomenilo tveganje, da bo trenutni pravni okvir nezadosten glede odgovornosti za odškodninsko, kaznivo in pogodbeno kršitev, ki vključuje umetno inteligenco in povezane pogoje. Kdo je torej odgovoren? Ali je to programer, proizvajalec ali uporabnik ali pa celo umetna inteligenca sama. Ali je odgovoren isti človek za celotno umetno inteligenco ali se razlikuje. Pravna odgovornost za (ne)dejanja umetne inteligence bi bila tradicionalno pripisana lastniku, razvijalcu ali proizvajalcu. Primer so samovozeča vozila v Nemčiji, kjer je odgovoren lastnik, vendar problem nastane z naprednimi sistemi, kot so samoučeče nevronske mreže. Vprašanje je torej, ali je še vedno on odgovoren, če ni mogel predvideti, da bo umetna inteligenca nenadoma spremenila

dizajn. Sedanja zakonodajna infrastruktura in pomanjkanje učinkovitih regulatornih mehanizmov predstavlja izziv pri urejanju omenjenih težav in bi bila morda potrebna nova pravna kategorija, ki bi opredelila značilnosti in omejitve (Bird in drugi, 2020).

Etična dilema se pojavlja tudi glede vpliva na okolje in planet. Tehnologija zahteva računalniško moč, ki pride s stroški energije. Ali lahko umetna inteligenca raste iz energetskega vidika, ko pa smo soočeni s klimatskimi spremembami? Že električna vozila so škodovala okolju, umetna inteligenca pa bo to še povečala. Zaradi pomanjkanja zalog bodo tudi ljudje morali delati v nevarnih okoljih, kar vodi v nadaljnjo avtomatizacijo rudarjenja in pridobivanja kovin in s tem ogrožitev okolja. Z zasnovano zastarelostjo elektronskih naprav se bo povečala težava kopičenja odpadkov in s tem težkih kovin in strupenih materialov. Težava je tudi s pojavom potrebe po ogromnih količinah energije potrebne zaradi vse večje količine podatkov in njihove obdelave kot tudi za proizvodnjo in učenje, ki ju zahteva umetna inteligenca (Bird in drugi, 2020).

Po drugi strani pa lahko umetna inteligenca pripomore k upravljanju odpadkov in zmanjšanju onesnaževanja. Na primer z avtonomnimi vozili bi zmanjšali toplogredne pline, saj bi lahko programirali ekološko vožnjo in s tem zmanjšali porabo goriva za 20 % ter v podobnem obsegu emisije toplogrednih plinov. S pomočjo priporočil alternativnih poti ter izmenjavo informacij z drugimi vozniki v prometu bi se zmanjšali zastoji v prometu. Tehnologija globokega učenja bi omogočila analizo slik živali, zajetih s kamerami s senzorjem gibanja in zagotovila potrebne informacije za ohranjanje biodiverzitete (Bird in drugi, 2020).

Da bi zaupali umetni inteligenci, mora biti ta poštena in nepristranska. Vendar se problem v povezavi s poštenostjo pojavlja zaradi modernih sistemov umetne inteligence, ki temeljijo na globokem učenju in zaradi česar je skoraj nemogoče ugotoviti, zakaj in kako so bile odločitve sprejete. Označeni so tudi kot črne škatle (angl. black boxes), saj temeljijo na umetnih nevronskih mrežah (angl. Artificial neural network, v nadaljevanju ANNs) oziroma umetnih nevronskih mrežah. Obstaja prepričanje, da sistem nikoli ne spremeni svojega vedenja, vendar sistemi, ki temeljijo na strojnem učenju že po definiciji ga. Mnoga podjetja programske opreme in algoritme smatrajo kot poslovno skrivnost za ohranitev konkurenčnega položaja na trgu. Za večjo transparentnost bi morali opazovati zunanje vedenje algoritmov, saj tudi pri vedenju ljudi ne gledamo v njihove možgane, temveč opazujemo in ocenjujemo glede na standarde obnašanja. Meriti bi bilo treba nepravilnosti v procesu razvoja, saj je najbolj enostaven način razumevanja tehnologije, da razumemo, kako je bila oblikovana, zakaj je bila ravno tako in za kakšen namen. Naslednji način za povečanje zaupanja je z zanesljivostjo in s tem odgovornostjo v primeru napak kot tudi regulacijo in pomemben element je etično upravljanje, ki bi ga morali vsi oblikovalci in organizacije upoštevati (Bird in drugi, 2020).

Zadnji sklop naslavlja še kontrolo. Raziskovalci so predlagali, da bi moral vedno biti prisoten model HITL, ki je nujno potrebna komponenta. V tem primeru gre za človeško interakcijo,

ko na primer posameznik določeno e-sporočilo označi kot vsiljeno pošto ter tako postane samo eden izmed ljudi HITL-kompleksnega strojnega učenja algoritmov, saj pripomore k izboljšanju opredelitve mailov. Raziskovalci so tudi predlagali načine zaustavitve sistemov umetne inteligence z tako imenovanim rdečim gumbom (angl. big red button) in tako imenovanim stikalom za ubijanje (angl. kill switch), ki bi omogočil, da človek prekine sistem, ne da bi sistem to smatral kot grožnjo. In tukaj obstaja strah pred tem, da bi se sistem naučil braniti in onemogočiti to (Bird in drugi, 2020).

Slika 3: Etična in moralna vprašanja razvoja in izvajanja umetne inteligence



Prirejeno po Bird in drugi (2020).

Vse predstavljene izzive, tako etične kot tudi pravne, se je v povezavi z umetno inteligenco do sedaj poskušalo sproti razrešiti. V zadnjem času pa je moč opaziti, da se gledajo kot celota in stremijo k razumevanju, kaj pomeni odgovorna umetna inteligenca. Ta naj bi po sprejetem dogovoru vsebovala načela, ki omogočajo, da uporaba sistemov umetne inteligence temelji na odgovornosti, transparentnosti kot tudi etičnosti. Z namenom uresničevanja vseh pričakovanj v povezavi z umetno inteligenco pa je ključno spremljati celoten življenjski cikel sistema (Mikalef, Conboy, Eriksson Lundström & Popovič, 2022).

#### 4 TVEGANJA IN ETIČNE SMERNICE UPORABE UMETNE INTELIIGENCE

Umetna inteligenca poleg dobrega lahko ustvari tudi škodo. Škoda s stvarnega vidika zadeva posameznikovo varnost in zdravje ali pa premoženjsko škodo in tudi smrt, medtem ko je iz

nestvarnega vidika mišljena predvsem v pomanjkanju zasebnosti, svobode podajanja mnenja, diskriminacije ter omejenem spoštovanju dostojanstva posameznika. S pomočjo regulativnega okvirja bi se morala tveganja zmanjšati. Uporaba pravil varstva temeljnih pravic, kamor sodita varstvo osebnih podatkov in zasebnosti ter nediskriminacija, je eno izmed glavnih tveganj umetne inteligence poleg tveganj za varstvo proizvodov in odgovornost (Evropska komisija, 2020).

Evropska unija je z novimi pravili in omejitvami pri uporabi visoko tvegane tehnologije dosegla, da ob ne upoštevanju grozijo podjetjem kazni. Poleg tega je Evropska komisija predlagala določene omejitve pri uporabi aplikacij umetne inteligence. Tako bi bili v splošnem biometrični sistemi, namenjeni identifikaciji, prepovedani, razen v primerih terorističnih napadov, pogrešanih oseb oziroma za ohranjanje javne varnosti. S tem bi se povečevalo zaupanje v umetno inteligenco. Sistemi, ki imajo nesprejemljivo tveganje, bodo prepovedani. Takšna tehnologija se uporablja na različnih področjih, med drugim v kritični infrastrukturi, na primer v prometu, kjer je ogroženo življenje, izobraževanju in usposabljanju, kjer lahko odloča o dostopu, v varnostnih komponentah proizvodov na primer v robotsko podprti kirurgiji, pri zaposlovanju glede razvrščanja življenjepisov, zasebnih in javnih storitev pri najemih posojil, ki je odvisno od kreditnega točkovanja, kaznivih dejanjih, kjer je možen poseg v človekove pravice z ocenjevanjem zanesljivosti dokazov, pri nadzorih meja na primer s pristnostjo potnih listin ali pa v pravosodju in demokratičnih procesih (Stromboli, 2021).

Takšni sistemi bodo primorani izpolnjevati stroge obveznosti pred dajanjem na trg. Kamor sodijo sistemi ocenjevanja, kakovosti podatkov, sledljivosti podatkov z dnevniki, temeljita dokumentacija, dostopne jasne informacije, ukrepi za človeški nadzor kot tudi v čim večji meri varnost, točnost in robustnost. Nova uredba o strojih nadomešča tudi direktivo o strojih, v kateri so opredeljene zahteve, ki jih morajo izpolnjevati stroji z vidika zdravstva in varnosti. S tem se bo zagotovila varnost in vse več bo inovacij. Obravnavana so tudi varnostna tveganja, zagotovljena bo varna vključitev v stroje. Upoštevale se bodo tudi razmere na trgu. Tako bo večja mera pravne varnosti, manjši stroški za podjetja in pa zagotovljena bo skladnost z zakonodajnim okvirjem EU za proizvode (Stromboli, 2021).

#### **4.1 Tveganja za temeljne pravice in varstvo potrošnikov**

Z uporabo umetne inteligence in njenih vplivov na vrednote je možnost kršenja pravic, ki veljajo za temeljne v EU. To so svoboda izražanja, zbiranja, dostojanstva in nediskriminacije. Prav tako pa je treba zagotoviti varovanje osebnih podatkov in zasebnosti. Tveganja, ki se pojavljajo, so plod same zasnove in rabe podatkov, pri katerih ni bila odpravljena pristranskost. Vse večja pa je tudi možnost spremljave in analize ljudi in s tem nadzor nad njimi tako s strani državnih organov kot delodajalcev, kar ni del pravil varstva podatkov EU. Analiziranje velike količine podatkov in njihovih povezav lahko s pomočjo umetne inteligence spet privede do novih tveganj iz naslova varnosti osebnih podatkov.

Medtem pa velja omeniti tudi tveganje za pristranskost in diskriminacijo. Čeprav je vsak človek zmotljiv, bi bil ta učinek mnogo večji pri umetni inteligenci, ker človeško vedenje vključuje mehanizme družbenega nadzora, pri sistemih umetne inteligence pa tega ni. Zaradi vseh posebnosti, ki jih prinaša umetna inteligenca, pa bo ovirano tudi primerjanje skladnosti v primerjavi s pravili, ki veljajo po zakonodaji EU za varstvo temeljnih pravic. S tem so mišljeni netransparentnost, kompleksnost, nepredvidljivost in delno avtonomno vedenje (Evropska komisija, 2020).

Temeljne pravice so osnova za zaupanja vredno umetno inteligenco in v nadaljevanju so izpostavljene tiste, ki še posebej pripomorejo za zajem sistemov umetne inteligence izhajajoče iz celovitega sklopa nedeljivih pravic mednarodnega prava o človekovih pravicah, Listin EU in Pogodb EU. Ideja spoštovanja človekovega dostojanstva je v notranji vrednosti vsakega človeka, ki se ne sme pod nobenim pogojem zmanjšati, zatirati ali kakorkoli ogroziti na račun drugih oseb ali tehnologij, med katere sodijo tudi sistemi umetne inteligence (European Commission, 2019b). Zato je potreba po razvijanju umetne inteligence, ki služi ljudem v skladu s spoštovanjem in varovanjem telesne in duševne celovitosti ljudi, njihove identitete tako osebno kot kulturno in njihovo zadovoljitev osnovnih potreb (European Commission, 2019b).

Druga je spoštovanje demokracije, pravičnosti in pravne države. Uporaba sistemov umetne inteligence bi morala biti namenjena za ohranitev in spodbuditev demokratičnih procesov kot tudi za spoštovanje vrednot in odločitev ljudi. Hkrati pa ne smejo ovirati demokratičnih volilnih sistemov in procesov ter človeškega razmišljanja. Prav tako njihov način delovanja ne sme ogroziti temeljnih zavez pravne države, obveznih zakonov in predpisov (European Commission, 2019b).

Tretji so enakost, nediskriminacija in solidarnost. Delovanje sistemov umetne inteligence tako ne more ponuditi nepoštenih pristranskih rezultatov in mora ustrezno spoštovati potencialno ranljive osebe in skupine, ki jim grozi izključenost, in ne nazadnje pravic državljanov. Velik potencial sistemov umetne inteligence je v boljši zagotovitvi javnih dobrin in storitev za družbo. Hkrati pa ima umetna inteligenca lahko negativen vpliv na pravice državljanov, ki bi jih zato bilo treba zaščititi (European Commission, 2019b).

Veliko pozornosti je bilo namenjeno pozitivnemu vplivu umetne inteligence na gospodarsko rast, manj pa v smeri vpliva na temeljne pravice. Zaenkrat ni dovolj empiričnih dokazov v povezavi s pravicami in ukrepi, ki bi omogočili skladnost s temeljnimi pravicami. Listina EU o temeljnih pravicah v kombinaciji z Evropsko konvencijo o človekovih pravicah tvorita splošni okvir za temeljne pravice. Pomembno vlogo imajo tudi Splošna deklaracija o človekovih pravicah in pa glavne konvencije Združenih narodov o človekovih pravicah. K varovanju temeljnih pravicah prispeva tudi sekundarno pravo EU še zlasti pravni red varstva podatkov EU in ne diskriminacijska zakonodaja EU. Na nacionalni ravni pa tudi zakonodaje držav članic EU. Z ocenami učinka, ki jih zahteva Zakonodaja EU v skladu s Splošno uredbo

varstva podatkov, ublažimo negativne učinke dejavnosti podjetij in javne uprave na temeljne pravice (European Union for Fundamental Rights, 2021).

Raziskava je pokazala, da se pri izvedbi osredotoča predvsem na tehnične vidike in ne toliko na morebitne učinke. Tudi samo znanje intervjuvancev o samih temeljnih pravicah je slabo izključujoče zaradi varstva podatkov in nediskriminacijo. Velika večina pravi, da je prisoten vpliv na temeljne pravice z uporabo umetne inteligence. Razlike so se pokazale tudi med zasebnim in javnim sektorjem, saj se zasebni manj zaveda širine temeljnih pravic, ki bi morebitno lahko bile kršene. Poznajo pa vprašanja, ki se navezujejo na varstvo podatkov. Predstavniki podjetij, ki si pomagajo z umetno inteligenco, imajo slabo znanje glede drugih pravic. Potencialnih težav so se nekateri v celoti zavedali, medtem ko drugi so odgovornost za preverjanje vprašanj v povezavi s tem preložili na stranke (European Union for Fundamental Rights, 2021).

V EU so razviti sklopi neodvisnih organov, ki so zadolženi za varstvo temeljnih pravic. Med drugim tudi organi za varstvo podatkov, na katere so se pogosto obrnili uporabniki umetne inteligence. Raziskava Eurobarometra je ugotovila, da se 40 % Evropejcev zaveda vpliva na avtomatizirane odločitve v povezavi z varstvom podatkov, ki pa ima vpliv tudi na druge temeljne pravice. Večje zavedanje je pri uporabnikih umetne inteligence, so si pa enotni glede potrebe po večji jasnosti uporabe in pomena avtomatiziranega sprejemanja odločitev s pravnega vidika. Temeljni pogoji so tako zavedanje s strani vsakega posameznika o tem, kdaj vse se uporablja umetna inteligenca, in seznanjenost s tem, kako in kje se je možno pritožiti. Vse organizacije, ki uporabljajo umetno inteligenco, morajo javnost obvestiti o tem kot tudi o sprejetih odločitvah, temelječih na sistemu (European Union for Fundamental Rights, 2021).

## **4.2 Tveganja za varnost proizvodov in odgovornost**

Z uporabo tehnologije umetne inteligence so večja tudi tveganja za uporabnike v primeru vgradnje v proizvode ali storitve. Tako lahko na primer avtonomno vozilo zaradi določene napake napačno prepozna predmete na cesti, kar lahko povzroči nesrečo. Razlog za to je v sami zasnovi tehnologije kot tudi zaradi težav, povezanih s samo razpoložljivostjo in kakovostjo podatkov. Možne pa so tudi druge težave izhajajoče iz strojnega učenja. Z uporabo se tako povečajo tveganja ne glede na to, ali so omejena točno na proizvode in storitve temelječe na umetni inteligenci. Obstajati morajo varnostne določbe, pri katerih, če ni opredeljene jasnosti, se lahko ustvarijo tudi pravne negotovosti za podjetja, ki ponujajo proizvode z umetno inteligenco. Pride lahko do nejasnega položaja glede posredovanja tako nadzornih kot izvršilnih organov, s tem pa je manjša tudi varnost in ogrožena konkurenčnost podjetij. S primanjkljajem jasnosti zahtev in značilnosti tehnologije je izslediti odločitve, ki so lahko potencialno problematične, oteženo. Oškodovanim osebam pa je s tem težje priti do odškodnine. Tudi zaradi same dostopnosti do dokazov in manjše učinkovitosti v

primerjavi s tradicionalnimi tehnologijami do pravnih sredstev (European Commission, 2019b).

Direktiva o odgovornosti za proizvode narekuje, da je za škodo, povzročeno s strani proizvoda z napako, odgovoren proizvajalec, le da je s sistemi, temelječimi na umetni inteligenci, napake oteženo dokazati kot tudi škodo in povezavo med njima. Negotovost je tudi z vidika obširnosti direktive glede na vrste napak za primer, če so te recimo vzrok v primanjkljaju kibernetike varnosti. Zakonodaja EU o varnosti proizvodov obravnava predvsem vidik dajanja proizvodov na trg. Tako programska oprema, ki je del končnega proizvoda, mora po zakonodaji izpolnjevati pravila, pa vendar to ni jasno opredeljeno glede samostojne programske opreme. Poleg tega se navezuje le na proizvode, ne pa tudi na storitve, na primer zdravstvene, finančne. Funkcionalnost proizvodov in sistemov se lahko preko življenjskega cikla spreminja z na primer posodobitvami, odvisnimi od strojnega učenja. S tem pa se odprejo nova tveganja, ki sprva niso bila prisotna. Zakonodaja takšnih tveganj ne obravnava ustrezno, ker je osredotočena izključno na tiste, ki so bila prisotna ob dajanju na trg (European Commission, 2019b).

Nejasnosti so tudi glede prelaganja odgovornosti po dobavni verigi. Sodeč po zakonodaji EU je odgovornost na proizvajalcu proizvoda, ki je bil dan na trg s sestavnimi deli in tako vključuje tudi sisteme umetne inteligence. Do problema pride, če umetno inteligenco doda nekdo, ki ni proizvajalec zatem, ko pride proizvod na trg. Odgovornosti drugih v dobavni verigi pa urejajo nacionalna pravila (European Commission, 2019b).

Možnost je tudi za tveganja, ki jih zakonodaja EU ne pokriva. Med drugim kibernetike grožnje, tveganja za osebno varnost v povezavi z gospodinjskimi aparati in izguba povezljivosti. Pojavijo se lahko pri vstopu proizvodov na trg ali kasneje kot odraz posodobitev in samoučenja. Za oceno groženj bi bilo treba uporabiti vsa razpoložljiva sredstva. V poročilu glede varnosti in odgovornosti v povezavi z umetno inteligenco, internetom stvari in robotiko so ugotovili, da trenutna zakonodaja zajema koncept varnosti pred vsemi tveganji z uporabo proizvoda, bi pa bile potrebne določbe, ki bi zajemale vsa nova tveganja za zagotovitev večje pravne varnosti. Avtonomnost lahko vodi do sprememb proizvoda, zato je, da bi zagotovili varnost, treba zagotavljati nove ocene tveganja in človeški nadzor vse življenje. Smiselno je razmisliti obveznosti proizvajalcev glede duševnega zdravja uporabe humanoidnih robotov. Določiti bi morali zahteve glede napačnih podatkov, do katerih lahko pride pri oblikovanju proizvoda, in mehanizme k ohranjanju kakovosti podatkov. Prilagoditi bo treba tudi pravila za samostojno programsko opremo, zahtevati preglednost in povečevati pravno varnost s sodelovanjem subjektov v dobavni verigi in uporabniki (European Commission, 2019b).

Da bi zagotovili zaupanja vredno umetno inteligenco, moramo ustrezno razporediti odgovornost med tistimi, ki izdelujejo orodja in storitve umetne inteligence, ter tistimi, ki jo uporabljajo (Weitzner, 2019). Podjetja, ki razvijajo algoritme, bi morala prevzeti odgovornost tako za obremenjenost algoritma z vrednostjo kot tudi za oblikovanje velikosti

vloge posameznika pri nadaljnemu prevzemu algoritemskih odločitev. Če je zasnovan na način, da posameznikom preprečuje prevzem odgovornosti posamezne odločitve je ustvarjalec tisti, ki je odgovoren za etične posledice. V nasprotnem primeru bi lahko podjetja poskušala prevaliti krivdo na sisteme umetne inteligence (Martin, 2018).

Možnost je tudi vrzeli v določitvi odgovornosti, kar pomeni, da ni jasne opredelitve, kdo je odgovoren. Razvijalci se morajo zavedati vplivov sistemov, saj so oni tisti, ki pišejo algoritme in s tem določajo uspeh oziroma neuspeh kot tudi uporabo sistemov. Prav tako se mora opravljati revizija, spremljanje in ocenjevanje učinka. Če se kot posledica uporabe tehnologije pojavi težava, je breme na organizacijskem uporabniku umetne inteligence. Pomembno je, da se kot prvo organizacije zavedajo morebitnih težav in znotraj njih določijo odgovornosti in predvidijo vnaprej tako scenarije kot tudi pristope k obravnavi. Škoda mora vedno prevzeti pravna oseba in je ne gre pripisovati strojem. Razčleniti je treba, kdo so oblikovalci in kdo organizacijski uporabniki, in dokončno določiti odgovornost, v pomoč pa so lahko evidence in sistemi registracije in dokumentacije (Ryan & Stahl, 2021). Po mnenju skoraj 50 % anketirancev Developer Survey Results 2018 je odgovornost za posledice tehnologije na ramenih ustvarjalcev umetne inteligence in ne na vodilnih ljudeh v podjetjih (IBM, brez datuma a).

### **4.3 Etične smernice Evropske komisije o uporabi umetne inteligence**

Na področju umetne inteligence se pričakuje samoregulacija podjetij, kar pomeni samoustvarjanje in izvajanje smernic, ki stremijo k odgovornim sistemom umetne inteligence. Smiselno pa bi bilo, da bi vsa podjetja sprejela enake smernice in izpolnjevala enaka merila. S tem namenom je tudi Evropska komisija objavila Etične smernice za zaupanja vredno umetno inteligenco. Vendar je vprašanje zaupanja glede tega, ali se bodo podjetja samoregulirala. Podatki raziskav kažejo na to, da v večji meri ne, saj je poročilo The 2020 State of AI and Machine Learning, ki je vključevalo 374 organizacij, ki delajo s podatki umetne inteligence, ugotovilo, da jih 75 % meni, da je kritičen del njihovega poslovanja umetna inteligenca, le 25 % anketiranim pa je pomembna poštena umetna inteligenca. Smernice bi torej morale postati zakoni oziroma predpisi, ki bi se ob nespoštovanju morali kaznovati (O'Sullivan, 2021).

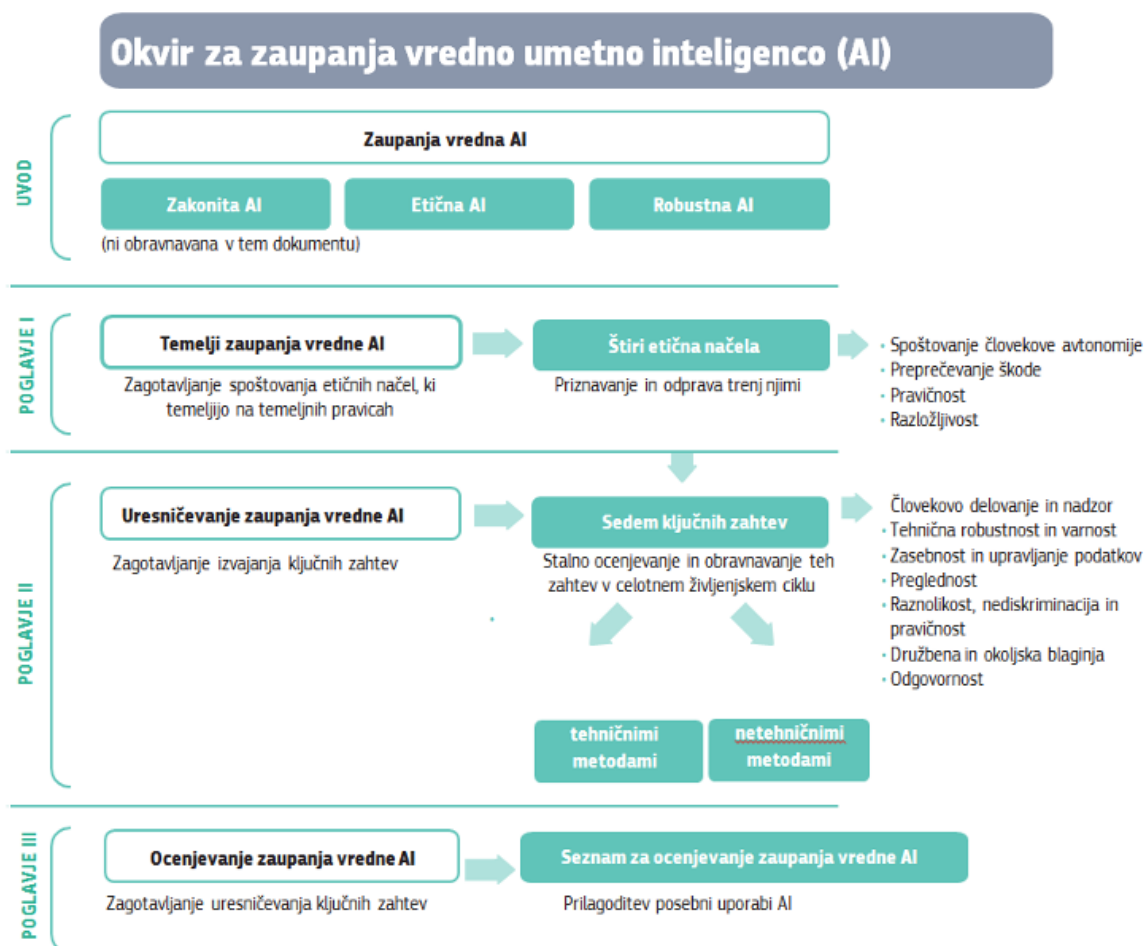
Z namenom povečanja zaupanja v umetno inteligenco je Evropska komisija ustanovila strokovno skupino katere del je 52 neodvisnih strokovnjakov. Povod za to so bile mednarodne razprave glede družbenih in etičnih vprašanjih navezujočih na vse večjo uporabo umetne inteligence kot del proizvodov in postopkov (Delakorda, 2019).

Evropska komisija je v dokumentu Etične smernice za zaupanja vredno umetno inteligenco opredelila, da so potrebni trije elementi, ki naj bi bili doseženi v življenjskem ciklu sistema. Tako bi morala biti umetna inteligenca zakonita ter s tem zagotavljati, da se spoštujejo zakoni in drugi predpisi. Prav tako bi morala biti etična in s tem zagotavljati, da se spoštujejo etična načela in vrednote. Etičnost pa je tesno prepletena s tretjim elementom, in sicer



moralo, ki mora biti robustna tako s tehničnega kot tudi družbenega vidika, namreč kljub dobrim namenom obstaja možnost, da sistemi umetne inteligence naredijo škodo pa čeprav nenamerno (European Commission, 2019b).

Slika 4: Smernice kot okvir za zaupanja vredno umetno inteligenco



Vir: European Commission (2019b).

Za zaupanja vredne sisteme umetne inteligence je treba spoštovati štiri etična načela, ki izhajajo iz temeljnih pravic. Načelo spoštovanja človekove avtonomije, torej možnost samostojnega odločanja in sodelovanja v demokratičnem procesu, mora kljub komuniciranju s sistemi umetne inteligence ostati prisotno. Njihova zasnova mora temeljiti na krepitvi kognitivnih, socialnih in kulturnih znanj ter spretnosti ljudi. Prav tako pa možnost vzpostavitve človeškega nadzora nad delovnimi procesi kot tudi njihovo upravljanje (European Commission, 2019b).

Načelo preprečevanja škode pomeni, da je treba stremeti k varstvu dostojanstva kot tudi duševne in fizične celovitosti ljudi, da sistemi umetne inteligence ne bi povzročali ali povečevali škode. Posebno pozornost pa bi bilo treba nameniti ranljivim osebam in situacijam, pri katerih lahko na račun asimetrij moči in informacij pride do povzročitve ali

pa poslabšanja škodljivih učinkov. Poleg ljudi je treba upoštevati okolje in vsa ostala živa bitja (European Commission, 2019b).

Načelo pravičnosti glede vsebinske razsežnosti pravičnosti pomeni pravično porazdelitev koristi in stroškov kot tudi, da ne prihaja do pristranskosti, diskriminacije in stigmatizacije. Poleg tega postopkovna razsežnost pravičnosti pomeni, da je mogoče izpodbijati odločitve sistemov umetne inteligence in ljudi, ki jih upravljajo, ter sprožiti pravna sredstva proti njim. Pomembno je, da subjekt, ki je sprejel odločitev, je mogoče identificirati in da so postopki odločitve jasni (European Commission, 2019b).

Načelo razložljivosti pomeni, da je za zaupanje vrednih sistemov potrebna preglednost postopkov in odločitve, razložljive ljudem. Dodatna pozornost pa je potrebna pri algoritmih »črne škatle«, kar pomeni, da je brez zgoraj navedenih informacij težko izpodbijati odločitve, ki jih je model sprejel, saj jih ni vedno moč pojasniti ter vhodne dejavnike, ki so prispevali k njej (European Commission, 2019b).

Na podlagi omenjenih štirih etičnih načel je podanih sedem zahtev, ki pripomorejo k zaupanju vredni umetni inteligenci. Za samo izvedbo teh zahtev so na voljo (ne)tehnične metode. Zahteve so iz naslova človeškega delovanja in nadzora. Kar pomeni, da bi bila potrebna ocena negativnega vpliva in vzpostavitev mehanizmov pridobivanja zunanjih informacij. Kot tudi zagotovitev orodij in znanja za uporabnike in s tem sprejemanja samostojnih odločitev. Pomembno vlogo ima tudi človeški nadzor za zagotovitev neogroženosti človeške avtonomije in drugih škodljivih učinkov (European Commission, 2019a).

Naslednja zahteva je tehnična robustnost in varnost, ki vključuje zaščito pred napadi, ter vzpostavitev ukrepov za izvedbo rezervnega načrta pred škodo za živa bitja in okolje. Potreba je tudi po uvedbi postopkov pojasnitve in ocene, prav tako tudi točnost glede ustrezne presoje sistema in navedbe verjetnosti napak. S tega naslova je ključna tudi ponovljivost enakega ravnanja ob enakih pogojih in zanesljivost (European Commission, 2019a).

Tretja zahteva je glede zasebnosti in upravljanja podatkov tekom celotnega življenjskega cikla sistema. Pred učenjem sistema je potrebno zagotoviti celovitost podatkov. Tiste organizacije, ki obravnavajo podatke, bi morale omogočiti dostop in pooblastila izključno usposobljenemu osebju (European Commission, 2019a).

Naslednja zahteva je preglednost vključujoč sledljivost, ki izboljšuje preglednost in bodoče napake kot tudi razložljivost torej, da so sprejete odločitve razumljive ljudem in obveščanje glede stika ali z ljudmi ali z umetno inteligenco (European Commission, 2019a).

Zahteve so tudi z naslova raznolikosti, nediskriminacije in pravičnosti, saj če je vzpostavljen nadzor, ne prihaja do nepoštene nepristranskosti. Zasnova bi morala stremeti k dostopnosti

za vse skupine ljudi in univerzalnosti. Tako je zagotovljena vključenost vseh ljudi (European Commission, 2019a).

Zahteva, ki igra pomembno vlogo, je tudi glede družbene in okolijske blaginje, kjer bi bilo potrebno pogledati ne samo z vidika vpliva na posameznika, temveč na celotno družbo, predvsem v povezavi z demokratičnim procesom (European Commission, 2019a).

In ne nazadnje odgovornost, ki se krepi z možnostjo revidiranja in s tem povečuje zaupanje. Za zmanjšanje negativnega vpliva je zaželena uporaba ocene učinka skupaj s tveganji sistema. Morebitna trenja se rešujejo s kompromisi oziroma na voljo morajo biti pravna sredstva. Za zagotavljanje uresničevanja ključnih zahtev pa je potrebno ocenjevanje umetne inteligence, k čemur lahko pripomore seznam za ocenjevanje zaupanja vredne umetne inteligence. Gre za nenehen proces, ki zadeva tako opredeljevanje, ocenjevanje in zagotavljanje boljšega razpleta (European Commission, 2019a).

#### **4.4 Univerzalne smernice o umetni inteligenci**

The Public voice (2018) je sprejel univerzalne smernice o umetni inteligenci (angl. Universal Guidelines for Artificial Intelligence, v nadaljevanju UGAI) oktobra 2018, katerih namen je ozaveščanje o vse bolj prisotnih izzivih računalniških sistemov. Z njimi želijo stremeti k večji transparentnosti kot tudi odgovornosti ter ohranitvi nadzora nad sistemi s podanimi priporočili za izboljšanje. Smernice UGAI temeljijo že na predhodnih delih Think tanks, mednarodnih in nevladnih organizacij kot znanstvenih društev. Vsebujejo tako človekove pravice kot etične smernice in zakone o varstvu podatkov. Tako lahko na podlagi njih pridobimo že uveljavljena načela kot tudi predloge novih. Kot osnovni cilj smernic je zaščita posameznika, zato se nanašajo predvsem na pravice posameznikov kot tudi obveznosti do institucij, ki razvijajo in uvajajo sisteme umetne inteligence. Smernice bi bilo treba vključiti tako v etične standarde kot tudi zagotoviti, da bi bile sprejete v nacionalnih zakonodajah in sporazumih na mednarodni ravni, predvsem pa, da so že v osnovi del načrta sistemov. Smernice tako pojasnijo načela za sisteme, predvsem tiste, ki imajo vpliv na pravice ljudi.

Prvo načelo je transparentnost, ki ga je moč zaslediti v različnih zakonih. Čeprav morda razlaga sprejete odločitve ne bi bila mogoča, je posamezniku treba zagotoviti možnost razlage, kar je cilj načela poleg neodvisnosti in odgovornosti za sprejete odločitve.

Drugo načelo o odločanju pojasnjuje, da so za avtomatizirane odločitve odgovorni ljudje in ne stroji. Cilj je torej opredeliti odgovornost odločitev umetne inteligence. V posameznih primerih človeška interakcija ni mogoča, vendar v primeru, da sistem odpove mora odločitev sprejeti človek.

Obveznost po identifikaciji je temelj za odgovornost, saj je s tem opredeljena tako identiteta sistema umetne inteligence kot tudi institucija, ki je odgovorna za to, s čimer je obravnavana

asimetrija identifikacije med interakcijo sistema in posameznika. V mnogih primerih sistem o posamezniku ve marsikaj, posameznik pa ne ve, kdo je upravljavec sistema.

Načelo pravičnosti opredeljuje, da kljub sprejemanju odločitev umetne inteligence glede na predsodke in diskriminacijo le-te ne bi smele veljati za nepravilne, saj gre za prenos iz človeka na stroj. Opredelitev nepoštenosti je v odvisnosti od konteksta. Pa vendar izključno ocena objektivnih rezultatov še ni dovolj, da bi lahko ocenili sistem umetne inteligence, ampak je treba vključiti tudi že obstoječe normativne posledice kot tudi tiste, ki jih sistem dopolnjuje.

Načelo ocenjevanja in odgovornosti pojasnjuje pomen ocenjevanja sistema pred in med uvedbo, da bi ugotovili pomembnost vzpostavitve sistema, v primeru tveganj pa projekt prekiniti.

Natančnost, zanesljivost in veljavnost skupaj določajo odgovornosti v povezavi z avtomatiziranimi odločitvami in jih je treba interpretirati tako skupaj kot narazen. Izhajajoč iz tega velja tudi načelo kakovosti.

Načelo javne varnosti priznava nadzor sistemov umetne inteligence nad napravami v fizičnem svetu. Ravno zato je naloga institucij sprejetje preventivnih ukrepov glede na oceno tveganja.

Načelo kibernetske varnosti je v povezavi z javno varnostjo in poudarja, da so kljub dobri zasnovi sistemov lahko le-ti tarča sovražnikov, zato morajo ljudje, ki razvijajo in uvajajo sisteme, v zakup vzeti tudi tveganja.

Iz načela obveznosti identifikacije izhaja prepoved tajnega profiliranja, katerega cilj je, izogibati se informacijski asimetriji kot tudi omogočiti neodvisno odgovornost. Prepoved glede enotnega točkovanja pojasnjuje tveganje, ki nastane z večnamenskim številom, dodeljenim posamezniku. Že v sami zakonodaji o varstvu podatkov je moč videti nenaklonjenost glede profiliranja posameznikov. Identifikatorji so v večini primerov regulirani kot tudi v določenih primerih prepovedani. Univerzalno enotno točkovanje je toliko bolj kritično, ker odraža vnaprej določen rezultat in ne samo enotni profil.

In ne nazadnje: pri obveznosti ustavitve gre za odgovornost nad sistemi umetne inteligence. Človek mora imeti vedno nadzor nad sistemi v nasprotnem primeru jih je treba ustaviti.

## **5 RAZISKAVA ETIČNIH VIDIKOV ODLOČANJA S POMOČJO ALGORITMOV**

### **5.1 Zasnova raziskave**

V empiričnem delu sem analizirala etične vidike odločanja s pomočjo algoritmov med ljudmi skozi različne resnične primere, poleg tega pa tudi uporabo algoritmov za odločanje pri delu. Raziskovala sem poznavanje umetne inteligence in algoritmov med ljudmi ter njihov pogled vpliva algoritmov za odločanje. Temelji na kvantitativni raziskavi, ki sem jo opravila s pomočjo spletne ankete, oblikovane v spletni aplikaciji Ika, ki je namenjena spletnemu anketiranju. Anketna vprašanja sem oblikovala glede na zastavljeno kazalo magistrskega dela oziroma sosledje obravnavanih tem. Večinoma se anketna vprašanja navezujejo na primere uporabe algoritmov umetne inteligence, ki so bila natančneje predstavljena v teoretičnem delu magistrskega dela. Podatke sem statistično obdelala v programu Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA, ZDA) in v programu IBM SPSS 25 (SPSS Inc., Armonk, NY, ZDA). Opisnim spremenljivkam sem izračunala frekvenčno porazdelitev (frekvence in frekvenčne deleže) in jo predstavila v tabelah. Velikost vzorca naključnih anketirancev, katerih odgovori so ustrezni, je 225, medtem ko je anketo v celoti končalo 200 anketirancev. Kot je razvidno iz tabele 1 v Prilogi 1, je v vzorcu zajeto večje število žensk. Najvišji delež oseb je v starostni skupini od 25 do 34 let, najmanj je tistih, ki so stari nad 55 let. Največje število anketirancev glede na najvišjo doseženo izobrazbo je v 6. razredu izobrazbe. Največ anketiranih opravlja poklic v smeri družbenih, poslovnih, upravnih in pravnih ved ter je del zasebnega sektorja.

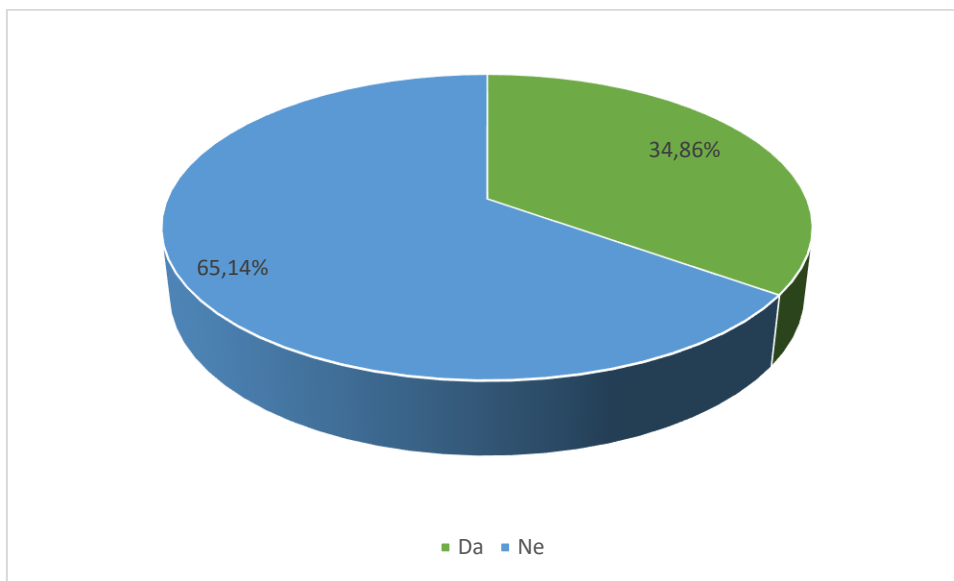
### **5.2 Analiza**

Opravljen analiza anketnega vprašalnika je podala naslednje ugotovitve. Kot je razvidno iz tabele 2 v Prilogi 1, skoraj vsi anketiranci poznajo izraz umetna inteligenca. Prav tako se jih velika večina zaveda, da nezavedno uporabljamo umetno inteligenco, kar lahko vidimo v tabeli 3 v Prilogi 1. Kot je razvidno v tabeli 4 v Prilogi 1, je delež anketiranih ljudi, ki razmišljajo o tem, da se algoritmi umetne inteligence odločajo namesto ljudi, večji v primerjavi s tistimi, ki menijo, da se ne. Malo manj kot polovica anketiranih je bilo mnenja, da lahko računalniški programi sprejemajo odločitve brez vpletenosti človeške pristranskosti, medtem ko malo več kot polovica meni, da bodo le-ti vedno odražali pristranskost ljudi, ki so jih zasnovali, kar lahko vidimo v tabeli 5 v Prilogi 1.

Na podlagi podanega primera uporabe algoritmov umetne inteligence pri zaposlitvenem postopku lahko vidimo v spodnjem grafu, da se večina anketiranih ne zaveda takšne uporabe. Prav tako se malo manj kot polovici anketiranih zdi tovrsten program pretežno nepravičen, najmanj pa jih meni, da je zelo pravičen, kar prikazuje tabela 7 v Prilogi 1. Hkrati so si bila mnenja glede pretežne (ne)učinkovitosti tovrstnega programa skorajda enaka glede na podatke iz tabele 8 v Prilogi 1. Najmanj jih je bilo mnenja, da so zelo učinkoviti. Tabela 9

v Prilogi 1 dokazuje, da se večini anketiranim tovrstni program ni zdel etično sprejemljiv. Na podlagi tabele 10 v Prilogi 1 ugotovimo, da je večina anektiranih kot razlog za etično (ne)sprejemljivost podala mnenje, da so algoritmi v redu, dokler ne posegajo v celoten proces izbire, najmanj se jih je odločilo za to, da gre za bolj objektivni način, pod možnostjo drugo pa so podali svoje mnenje. Odgovori so bili, da se zmanjšuje različnost zaposlenih, da je težko iz življenjepisa prepoznati značaj človeka, da so lahko algoritmi pristranski in pomagajo pri učinkovitosti, vendar ne samostojno, da ne podpirajo lenobe v podjetjih in da so lahko profili na Facebooku, preko katerih algoritmi izbirajo kandidate, olepšani.

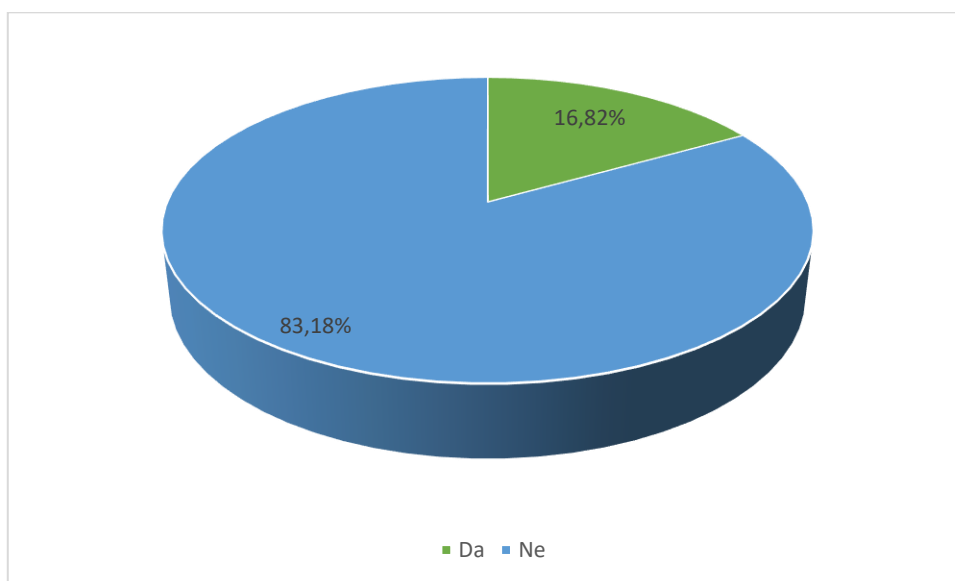
*Slika 5: Zavedanje o uporabi algoritmov pri zaposlitvenem postopku*



*Vir: lastno delo.*

Na podlagi podanega primera uporabe algoritmov umetne inteligence pri izračunu kazenskega tveganja lahko v spodnjem grafu vidimo, da se večina anketiranih ne zaveda takšne uporabe. Prav tako se malo manj kot polovici anketiranih zdi tovrsten program pretežno nepravičen, najmanj pa jih meni, da je zelo pravičen, kar je razvidno iz tabele 12 v Prilogi 1. Hkrati jih je največ menilo, da je tovrstni program pretežno neučinkovit, najmanj pa, da je zelo učinkovit, kar lahko razberemo iz tabele 13 v Prilogi 1. Tabela 14 v Prilogi 1 dokazuje, da se večini anketiranim tovrstni program ni zdel etično sprejemljiv. Na podlagi tabele 15 v Prilogi 1 ugotovimo, da je večina anketiranih kot razlog za etično (ne)sprejemljivost podala dejstvo, da je vsak človek in situacija specifična, najmanj se jih je odločilo za pristranskost in drugo, kjer so podali svoje mnenje, kjer je bil odgovor, da je etično sprejemljivo.

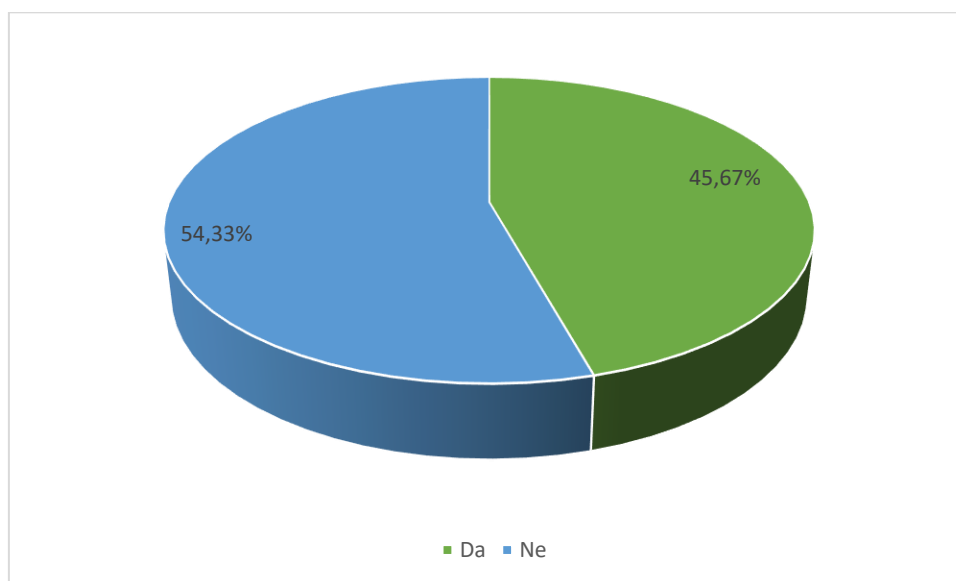
Slika 6: Zavedanje o uporabi algoritmov pri izračunu kazenskega tveganja



Vir: lastno delo.

Na podlagi podanega primera uporabe algoritmov umetne inteligence pri izračunu ocene osebnih financ lahko v spodnjem grafu vidimo, da se večina anketiranih ne zaveda takšne uporabe. Prav tako je približno enak delež anketiranih mnenja, da je tovrstni program pretežno (ne)pravičen. Najmanj jih meni, da je zelo pravičen, kot je razvidno iz tabele 17 v Prilogi 1. Hkrati jih je bil največji delež mnenja, da je tovrstni program pretežno učinkovit, in najmanj, da je zelo neučinkovit, kar prikazuje tabela 18 v Prilogi 1. Glede na tabelo 19 v Prilogi 1 je razvidno, da je delež anketiranih približno enak glede etične sprejemljivosti tovrstnega programa. Tabela 20 v Prilogi 1 prikazuje, da je večina anketiranih kot razlog za etično (ne)sprejemljivost navedla, da se morajo ljudje, ki so dejavni na internetu, sami zavedati, da puščajo za seboj podatke. Najmanj se jih je odločilo za odgovor, da ni možnosti vpliva na podatke oziroma jih popraviti, in za odgovor drugo.

Slika 7: Zavedanje o uporabi algoritmov pri izračunu ocene osebnih financ

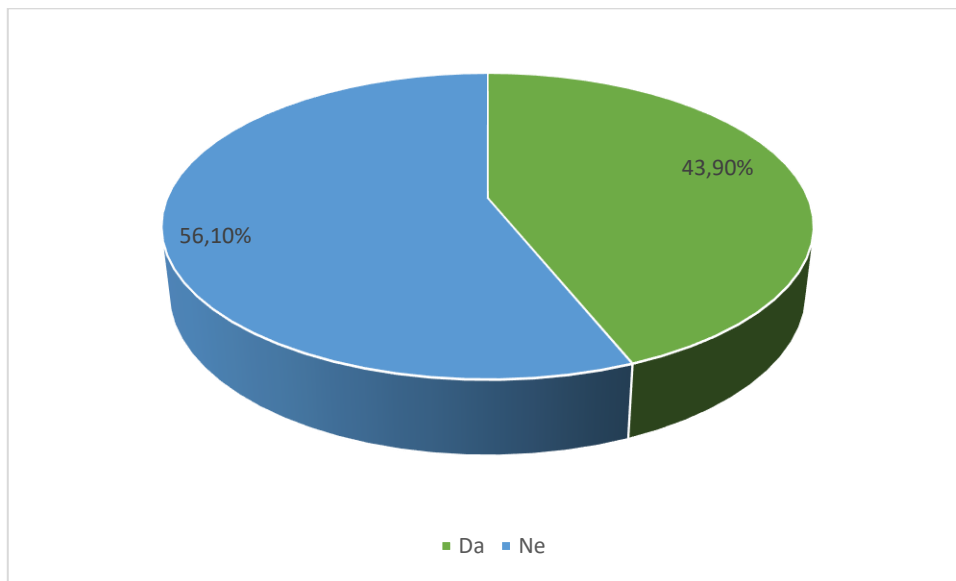


Vir: lastno delo.

Na podlagi podanega primera uporabe algoritmov umetne inteligence pri hitrejšem odkrivanju bolezenskih stanj lahko v tabeli 21 v Prilogi 1 vidimo, da se večina anketiranih ne zaveda takšne uporabe. Hkrati jih je največji delež menil, da je tovrstna tehnologija pretežno učinkovita, in najmanj, da je zelo neučinkovita, kar prikazuje tabela 22 v Prilogi 1. Glede na tabelo 23 v Prilogi 1 je delež anketiranih, ki je podal odgovor, da je tovrstna tehnologija etično sprejemljiva mnogo večji kot tisti, ki govori nasprotno. Na podlagi tabele 24 v Prilogi 1 razberemo, da je večina anketiranih kot razlog za etično (ne)sprejemljivost podala hitrejšo diagnozo, najmanj pa se jih je odločilo za prevzemanje odgovornosti in možnost drugo.



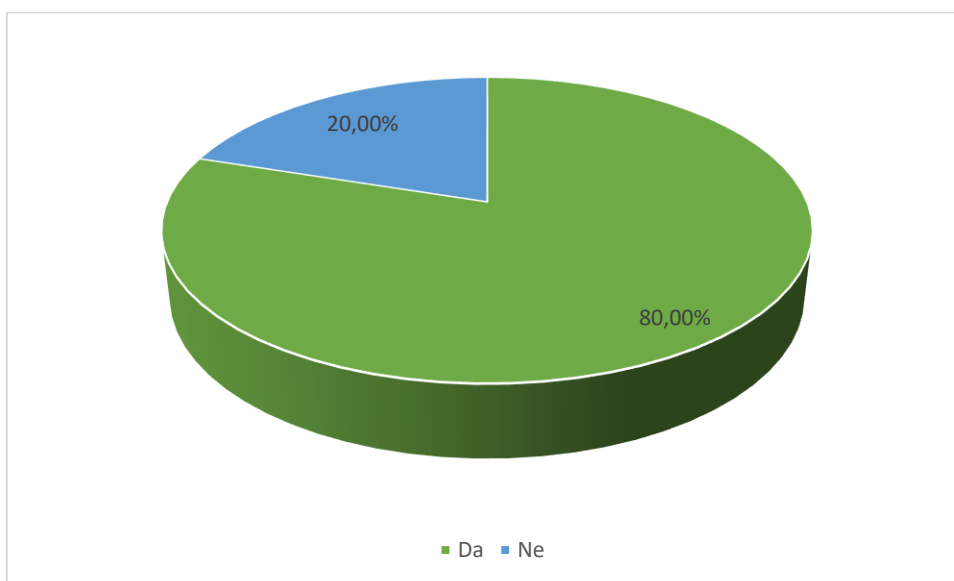
Slika 8: Zavedanje o uporabi algoritmov pri hitrejšem odkrivanju bolezenskih stanj



Vir: lastno delo.

Na podlagi podanega primera uporabe algoritmov umetne inteligence pri avtonomnih vozilih lahko vidimo v spodnjem grafu, da se večina anketiranih zaveda takšne uporabe. Hkrati jih je bilo največji delež mnenja, da je tovrstna tehnologija pretežno učinkovita, in najmanj, da je zelo neučinkovita (tabela 26 v Prilogi 1). Tabela 27 v Prilogi 1 prikazuje, da je delež anketiranih, ki so podali odgovor, da je tovrstna tehnologija etično sprejemljiva, in tistih, ki se s tem ne strinjajo, približno enak. Na podlagi tabele 28 v Prilogi 1 ugotavljamo, da je večina anketiranih kot razlog za etično (ne)sprejemljivost podala odgovor, da je potrebna človeška vpetost, najmanj se jih je odločilo za dodano vrednost in odgovor drugo ter podalo svoje mnenje. Ta so bila da je manjša verjetnost za kršenje cestnih predpisov, da so v resnici takšna vozila programirana, da vzamejo najbolj blago opcijo, ki je varna za potnike in v primeru napake (na primer senzor zazna muho kot otrok) nihče ne bi kupil takšnega vozila, ki te lahko takoj ubije. Poleg tega so napisali tudi, da bolj zaupajo sebi kot tehnologiji. Navedli so tudi, da bi pogrešali svobodo pri vožnji, saj da te hekerji lahko preusmerijo v prepad ter da je velik vzrok za nesreče alkohol, zato je bolje, da vozilo vozi samo.

Slika 9: Zavedanje o uporabi algoritmov pri avtonomnih vozilih



Vir: lastno delo.

Kot prikazuje tabela 29 v Prilogi 1, se večina anketiranih zaveda zbiranja in obdelave biometričnih podatkov z biometričnimi fotografijami v potnem listu, prstnimi odtisi in uporabo kamer za prepoznavanje obrazov. Delež anketiranih, ki je podal odgovor, da je uporaba biometričnih podatkov etično sprejemljiva, je večji od tistega, ki prikazuje nestrinjanje (tabela 30 v Prilogi 1). Večina anketiranih je kot razlog za etično (ne)sprejemljivost (tabela 31 v Prilogi 1) podala vzrok z vidika splošne varnosti, najmanj se jih je odločilo za omejevanje svobode gibanja in možnost drugo ter podalo svoje mnenje, da je sprejemljivo, vendar bi se morali državljani strinjati z zbiranjem podatkov, ne pa da bi jih kdo v to silil, s čimer bi se zmanjšala osebna svoboda. Prav tako je bilo njihovo mnenje, da bodo kmalu na osnovi nekaj napak določili, da nisi ustrezna oseba, podobno kot Hitler, ki je želel imeti superiorno raso.

Tabela 32 v Prilogi 1 prikazuje, da je velika večina mnenja, da bi morala biti pri odločanju algoritmov večja preglednost, prizadevanje za rast algoritmične izobraženosti, večji nadzor nad uporabo, prav tako bi morala država ščititi pred zlorabami, razvijati etiko podatkov in poskrbeti, da bodo pozitivni vplivi odtehtali negativne. Približno enak delež je podalo odgovor glede zaupanja v tehnologijo algoritmov, da podaja kakovostne in nepristranske podatke. Največji delež anketiranih glede na tabelo 33 v Prilogi 1 je mnenja, da je za neetično odločitev odgovorno podjetje, ki uporablja omenjeno tehnologijo, medtem ko najmanjši, da je za to kriv stroj. Anketirancem je v večini pomembno, da je odločitev zakonita, uravnotežena med obema stranema in pomembno kakšno mnenje bodo imeli o sebi po sprejeti odločitvi, kar lahko vidimo v tabeli 34 v Prilogi 1.

Z algoritmi se pri opravljanju svojega poklica anketirani večinoma ne srečujejo. Najmanj odgovorov anketiranih je bilo, da se nameravajo v prihodnosti in izbralo možnost drugo ter

navedlo uporabo Google Businessa pri svojem delu, kar je razvidno iz tabele 35 v Prilogi 1. Odgovore tistih, ki se pri svojem delu srečujejo z uporabo algoritmov prikazuje tabela 36 v Prilogi 1. Največji delež teh jih uporablja za varnostno zaščito potrjevanja transakcij. Medtem ko najmanj jih je izbralo podano možnost za samodejne preglede življenjepisov za delovno mesto ter drugo, kjer so podali svoj odgovor in sicer za analizo spletnega obnašanja in nakupovanja ter energetiko. Tisti anketirani, ki so podali odgovor, da se ne srečujejo z uporabo algoritmov pri svojem delu, so glede na podatke v tabeli 37 v Prilogi 1, podali odgovore, da bi večinoma uporabljali tehnologijo v primeru ustreznega kadra oziroma sredstev. Izzive, s katerimi se srečujejo anketirani, ki nimajo vzpostavljene omenjene tehnologije, prikazuje tabela 38 v Prilogi 1, in sicer v največji meri glede varovanja zasebnosti podatkov. Najmanj pogosto pa so anketirani izbrali izzive v povezavi z zagotovitvijo finančnih sredstev in možnosti drugo, kjer so navedli, da zaradi časovne zahtevnosti implementacije.

### 5.3 Ključne ugotovitve

K razširjenosti umetne inteligence sta prispevala strojno in predvsem globoko učenje. Njena uporaba je prisotna na vse več področjih kot tudi v večini domov. Dandanes prav tako velja za konkurenčno prednost podjetij. Ravno zaradi njene razširjenosti ni presenetljivo, da je med anketiranimi poznavanje področja umetne inteligence v večji meri prisotno kot tudi, da se jih večina zaveda njene nezavedne uporabe. Na podlagi predstavljenih primerov uporabe algoritmov umetne inteligence za namene odločanja namesto ljudi lahko vidimo, da so se razširili na vse več pomembnih področij o čemer razmišlja tudi večina anketiranih. V splošnem so bili tudi mnenja, da računalniški programi odražajo pristranskost ljudi, kar je seveda neizogibno, saj jih oblikujejo ljudje, ki prispevajo svoj vložek k temu, ali jo bodo odražali ali ne. To dokazujejo različni primeri uporabe, na primer iskalnik Google Images, ki za izraz »CEO« prikaže le 11 % žensk, čeprav jih je 27 % na tem položaju v Združenih državah, ravno zaradi označbe podatkov, ki daje prednost moški populaciji za to specifično delovno mesto.

Osnovna pravica vseh ljudi je soglasje za uporabo, deljenje in hranjenje biometričnih podatkov. S tem narašča ozaveščenost. Razširjenost pa se veča tudi zaradi same unikatnosti, neprenosljivosti, težkih ponaredb in prekritja. Med anketiranimi je v veliki večini prisotno zavedanje o zbiranju in obdelavi biometričnih podatkov, kot so biometrične fotografije v potnih listih ali uporaba prstnih odtisov in kamer s tehnologijo prepoznave obraza, kar se jim zdi etično sprejemljivo v največji meri z vidika splošne varnosti in preprečevanja kriminala. Po podatkih iz literature je ob pravilni uvedbi sprejetje med strankami 93 %, večja se tudi povpraševanje in ponudba na trgu. Čeprav se pozitivni učinki odražajo v varnosti in kazenskem pregonu, je delež napak identifikacije glede na opravljene teste sprva znašal 1 %, vendar se je z leti njihova učinkovitost povečevala. V primeru, da se podatki ne uporabljajo preudarno, lahko pride do omejevanja svobode gibanja, kar pa je anketirane najmanj skrbelo poleg kršitev varstva osebnih podatkov in človekovih pravic. Pa vendar se

tehnologije prepoznavne obraza uporabljajo za neomejeno sledenje ljudi in so s tem sporne ravno zaradi kršenja zasebnosti in varstva podatkov. Hkrati so kritične zlorabe, katerih primer je uporaba govora osebe, h kateremu se doda drug obraz. Prav tako hranjenje osebnih podatkov, ki se pojavlja s strani oblasti z uporabo pametnih kamer kot tudi ponudnikov tehnologije, ki ne zagotavljajo izbrisa podatkov. Pomembno je zagotoviti varovanje uporabnikov in ozaveščanje o negativnih posledicah. Evropska komisija je tudi predlagala omejitve; biometrični sistemi bi bili prepovedani, razen v izjemnih primerih, kot je iskanje pogrešanih oseb, v primeru terorističnih napadov oziroma za ohranjanje javne varnosti.

Avtonomna vozila postajajo z napredkom tehnologij, umetne inteligence in robotike vse bolj realnost. Anketirani se tako tudi pri njih zavedajo uporabe algoritmov in so mnenja, da je tovrstna tehnologija v večji meri učinkovita, kar potrjuje literatura, saj se prednosti kažejo predvsem v zaščiti okolja, gospodarski rasti, dostopnosti in varnosti. Hkrati pa je mnenje deljeno glede etične sprejemljivosti. Razlog je v potrebni človeški vpetosti in odgovornosti, a hkrati tudi trdijo, da se bo z njimi povečala varnost na cesti, kar je tudi namen, saj prometne nesreče, ki jih povzroča človek, povzročajo velik problem. Veliko jih je namreč menilo, da je možnost človeške napake večja kot tehnološke. Etično-moralna vprašanja, s katerimi se spopadajo pri vzpostavitvi avtonomnih vozil, zadevajo predvsem odločanje vozila v primeru nesreče, iz česar izhaja problem odgovornosti. Tehnološki izzivi, s katerimi se spopadajo, med drugim varnost delitve ceste z drugimi udeleženci v prometu, harmonizacije cestnoprometnih predpisov, obdelavi podatkov in vložkih v razvoj infrastrukture, naj bi bili odpravljeni v 75 do 80 %, vendar gre za vprašanje sprejemanja ljudi. Glede na študijo med evropskimi državljani ta znaša 58 %. Nekateri so namreč predani vožnji in v njej vidijo užitek, tudi predstave o avtonomnih vozilih so drugačne med ljudmi, prav tako se pojavlja strah pred vožnjo z njimi. Treba je tudi poudariti, da se premalo ljudi zaveda, da tako kot lahko policija prevzame nadzor nad vozilom ga lahko tudi kriminalci.

Malo več kot polovica anketiranih se ne zaveda uporabe algoritmov v zdravstvu, čeprav že preoblikujejo medicino, saj gre tovrstna orodja z roko v roki s porastom medicinskih podatkov. Tovrstna tehnologija se anketiranim zdi učinkovita in etično sprejemljiva zaradi hitrejše diagnoze in odkrivanja bolezenskih stanj kot tudi natančnosti. Najmanj jih skrbi prevzemanje odgovornosti in izvzetje človeške presoje. Kar je tudi namen, da pripelje do hitrejših in natančnejših diagnoz ter omogoči ekipno delovanje. Glede na porast podatkov so za ljudi zmožnosti obdelave presežene, globoki učni sistem pa je prispeval k zmanjšanju človeške napake za 85 % pri odkrivanju raka, višji stopnjo natančnosti pri razvrščanju ehokardiogramov, in sicer 92 % proti 79 %, v številu napovedanih srčnih infarktov ali možganskih kapi, ki je bilo večje kot pri človeku kot tudi odkrivanju depresivnega vedenja, kjer je bila stopnja natančnosti 84 %. To pomeni, da ljudi upravičeno ne skrbi učinkovitost tehnologije in odgovornost, glede na to, da rezultate algoritmov potrjujejo zdravniki in jim algoritem dopušča lastno presojo.

Tudi uporabe algoritmov za izračun kazenskega tveganja se ljudje ne zavedajo, pa vendar je odločitve sodnika v nekaterih segmentih že nadomestilo 60 različnih vrst računalniških

algoritmov. Uporaba se jim zdi nepravična in tudi neučinkovita. Menijo, da gre za etično nesprejemljivost, saj je smiselno tovrstni program uporabiti kot dodatno orodje in ne kot odločilni faktor, kot tudi je vsak človek in situacija specifična in dopuščajo možnost spremembe človeka. Vendar se ne zavedajo, da je lahko za enako dejanje dodeljena drugačna kazen zaradi sodnikove osebnosti in pomanjkanja smernic. Že sedaj so za določitev kazni vnaprej določeni koraki sodnika. Najmanj jih je bilo mnenja, da gre za pristranski način in da bi to pripomoglo k izboljšanju trenutnega sistema. Vendar podatki iz literature navajajo, da je namen uporabe algoritmov v sodstvu zmanjšanje pristranskosti, a se ta problem še vedno pojavlja, ker so programirani s strani ljudi. Odraž je ravno to, da program glede na dostopne informacije ocenjuje stopnjo ponovitve dejanja glede na prejšnje vzorce, ko so jim dodelili daljšo zaporno kazen. Poudariti pa je treba tudi, da omenjena tehnologija izboljšuje trenutni sistem, saj se je kot najbolj zanesljiv in mnogo učinkovitejši način v primerjavi s sodnikom izkazalo strojno učenje, ki nudi oceno dejstev v prihodnosti. Natančnost računalniške opreme je bila dosežena v 79 % v Evropi in 70 % na najvišjih sodiščih. Tako je namen nuditi sodniku stopnjo verjetnosti in pomoč pri višini kazni.

Zavedanje na področju zaposlovanja v večji meri ni prisotno, četudi je umetna inteligenca šla v korak s časom tudi na tem področju. Med anketiranimi velja za nepravičen, neučinkovit in etično nesprejemljiv način odločanja. Vendar če pogledamo odgovore, jih je največ v smeri, da so algoritmi ustrezni, dokler ne posegajo v celoten proces izbire in da lahko podjetja zaposlujejo, kogarkoli želijo ter jim to prihrani čas, kar je tudi namen. Algoritem namreč pripravi celovito podobo kandidata za tovrstno delo tako z vidika ocene skladnosti osebnosti kot tudi strokovnosti. Prav tako pa imata obe strani prilagodljiv čas poteka in pregleda intervjuja. Hkrati so bili odgovori v smeri, da se jim zdi tovrstni način pristranski do nekaterih prosilcev na primer žensk in najmanj se jih odločilo, da gre za bolj objektivni način. Kar drži, saj čeprav se algoritmi zanašajo na podatke odražajo oblikovanje s strani ljudi in glede na prakso je do sedaj prihajalo do diskriminacije. Algoritem je dodeljeval negativno težo ženskam, invalidom in mlajšim od 30 let ter se izkazal v določenih primerih za nezakonitega. Kot tudi je težko objektivni glede na to, da je razpoloženje kandidata spremenljiva stvar.

Družbeno sortiranje, ki je vidno v bančništvu in zavarovalništvu, kjer posamezniki ne vedo, na podlagi katerih kriterijev je bila sprejeta odločitev, odraža tudi mnenje anketirancev, ki se uporabe algoritmov za izračun ocene osebnih financ v večji meri ne zavedajo, vendar se jim zdi učinkovito. So si pa zato mnenja deljena glede pravičnosti kot tudi etične sprejemljivosti. Tako nekateri trdijo, da se morajo ljudje zavedati, kje puščajo podatke, kot tudi, da gre za odločitev podjetij, na kakšen način bodo pridobivali stranke. Za kreditne točke se namreč poleg podatkov navezujočih na identiteto, bančne transakcije, zgodovino kreditov, zavarovanj in prihodkov uporabljajo podatke pridobljene iz družbenih medijev, posnetkov dronov in satelitov ter vseh skratka vseh dostopnih podatkov. Tako lahko posameznik vpliva v določeni meri na to, kakšno sled bo puščal za seboj. Hkrati pa takšna analiza pripomore podjetjem pri boljši oceni tveganja in sposobnosti odplačevanja, s čimer

so učinkovitejša in lažje obvladujejo tveganja, povezana s krediti. Ostali so mnenja, da spletni podatki ne morejo podati prave slike o ljudeh in da gre za kršenje zasebnosti. Vendar s pridobivanjem celostnih podatkov je ocena tveganja in sposobnost odplačevanja mnogo bolje izvedena. Sodeč po raziskavi je natančnost sprejetih odločitev več kot 95 %, v primerjavi z zaposlenimi v banki kar pomeni, da so sposobni priti do enakih zaključkov kot ljudje in preučiti vse razpoložljive možnosti. Najmanj jih skrbi možnost vpliva na podatke kar dokazuje tudi, da so se v večini odločili, da je na vsakem posamezniku kje bo puščal svoje podatke in pa diskriminacija. Pa vendar so v praksi opazili pristranskost z vidika spola, kjer imajo moški več možnosti tako pri odobritvi kot višini kredita.

Na podlagi neznanstvene ankete 38 % anketiranih, predvsem mlajših izobraženih moških, meni, da bodo pozitivni učinki premagali negativne, 37 % meni nasprotno, 25 % je neodločenih. Glede na opravljeno anketo je bilo mnenja, glede prevlade pozitivnih nad negativnimi učinki, kar 61 % anketiranih. Mnenje strokovnjakov je, da bodo sistemi uspešni, če se bo namenilo dovolj pozornosti njihovi zasnovi ter s tem razviti nekaj, kar je moč nadzorovati. V tem primeru se bo tudi življenje ljudi izboljšalo, večji bo prihranek časa in denarja in zmanjšalo bi se število napak. V nasprotnem primeru se bo povečal digitalni razkorak in razlike v premoženju. Poleg tega bo težko ohranjati nadzor in preprečevati diskriminacijo. Povečevale pa bi se tudi zlorabe predvsem zaradi tega, ker bi lahko umetna inteligenca prekosila človeka do leta 2050, kar bi omogočilo več kibernetskih vdorov, možnosti posebitve ljudi in ustvarjanja lažnih računov. Po podatkih raziskave Pew Research Centra se je za prevlado pozitivnega pogleda glede preteklega razvoja umetne inteligence in uporabe sistemov namenjenih za posnemanje človeškega vedenja odločilo 53 % anketiranih, medtem ko jih 33 % meni nasprotno.

S porastom napak in zlorab je potreben etičen premislek o tem, kdo in za kakšne namene lahko uporablja tehnologijo. Mnenje anketiranih je, da bi jih morala država ščititi pred zlorabami, vendar je zaenkrat to možno le kot osnovna zaščita, šele po seznanjenosti ljudi glede razpoložljive tehnologije in njeni uporabi pa bi se lahko začelo razpravljati o omejevanju.

Odgovornost za avtomatizirane odločitve s strani tehnologije je po mnenju anketiranih na strani podjetij, ki jo uporabljajo in ne stroju ali razvijalcu. V raziskavi opravljeni leta 2018 je 50 % anketiranih menilo, da je odgovornost na ramenih ustvarjalcev umetne inteligence in ne na vodilnih ljudeh v podjetjih. Tudi univerzalne smernice o umetni inteligence določajo, da so odgovorni ljudje in ne stroji. Hkrati Direktiva o odgovornosti za proizvode kot tudi zakonodaja EU narekuje, da je za škodo, ki jo povzroči proizvod z napako, odgovoren proizvajalec. V primeru, da je umetna inteligenca naknadno dodana, pa to urejajo nacionalna pravila. Problem v povezavi z odgovornostjo, ki nastaja, je v tem, da je bil pravni sistem ustvarjen s predpostavko, da odločitev ne bo sprejemala umetna inteligenca, ampak ljudje. V zadnjem času pa vse bolj opažamo porast preusmerjanja zahtevnejših odločitev v roke algoritmov in s tem tveganja glede nezadostnosti obstoječega pravnega sistema. Vprašanja, ki se pojavljajo, so zato, kdo nosi odgovornost, programer, proizvajalec,

uporabnik ali umetna inteligenca. Hkrati se pojavi vprašanje, ali je odgovornost samo na enem človeku ali se porazdeljuje. Vse težje jo je opredeliti zaradi samoučočih se sistemov.

Prav tako so anketirani menili, da bi bil potreben večji nadzor glede uporabe umetne inteligence, kar potrjuje tudi raziskava, opravljena v 27 državah po svetu. Poleg tega je pokazala skrb ljudi glede odločitev, sprejetih na podlagi algoritmov v 40 %, popolnoma brez skrbi je bilo 27 % anketiranih in 32 % neodločenih. Skrb se deli med spoloma, starostjo, dohodki in izobraženostjo. Raziskava je tudi pokazala, da bi bilo treba uvesti regulacijo podjetij, ki uporabljajo umetno inteligenco; 20 % jih meni, da strožja pravila niso potrebna. Glede na v magistrskem delu opravljeno anketo je skoraj polovica anketiranih mnenja, da bi moral biti večji nadzor nad uporabo algoritmov. Tudi univerzalne smernice o umetni inteligenci narekujejo, da mora imeti človek vedno nadzor nad sistemi in da jih je v nasprotnem primeru treba ustaviti. Prav tako Etične smernice o uporabi umetne inteligence pojasnjujejo, da v primeru vzpostavitve nadzora ne prihaja do nepristranskosti. Hkrati pa avtonomnost vodi do sprememb proizvoda in je potrebno za povečanje varnosti imeti nadzor ves čas življenja.

V povezavi z zaupanjem, da algoritmi podajajo kakovostne in nepristranske podatke, je bilo mnenje anketiranih skoraj deljeno z večjo večino, da jih ne, kar je upravičeno, saj v primeru slabih podatkov lahko postane umetna inteligenca nenamerno škodljiva in pristranska kot odraz ljudi, ki so jo razvili. Sistem je namreč dober le toliko, kot so dobri uporabljeni podatki. Tudi Evropski svet je sklenil, da so visokokakovostni podatki nujno potrebni za sam razvoj ter da bi morale skrbi, povezane s podatki, postati osrednja tema debate razvijalcev. Tudi etika se odraža v kakovosti in opredeljuje dolžnost umetne inteligence in njenih razvijalcev v smeri ublažitve škode, do katere lahko pride z uporabo. Vendar glede pristranskosti bi lahko rekli, da je umetna inteligenca boljša kot človek, saj je sposobna odkriti nepoštenosti. Prav tako so kljub primerom pristranskosti algoritmov dokazali, da je najboljši algoritmi ne vsebujejo in so tudi vse boljši pri svojem delu.

Mnenje večine anketiranih je potreba po večji preglednosti, vendar težava nastane pri težavnem razumevanju kot posledice zahtevnosti algoritma, toka obdelanih podatkov, števila uporabnikov, ki si želijo razlago in skriti izvorni kodi zaradi intelektualne lastnine. Omejeno razumevanje ni samo s strani povprečnih ljudi, temveč tudi raziskovalcev in ustvarjalcev. Kljub temu bi morale organizacije omogočiti javnosti z etičnega vidika preglednost in vedeti, kako pojasniti procese algoritmov, saj se s tem poveča kredibilnost. Ljudem mora biti omogočena odškodnina v primeru napak sistema ter ustrezna razlaga. Le na takšen način lahko obvladujemo umetno inteligenco.

V prvotno zamisel je treba vgraditi etiko in kasneje nadzorovati tehnologijo z namenom razumevanja njenega vedenja in ravnanja skladno z moralnim kompasom. Neizogibno namreč prihaja do vrednostne obremenitve algoritmov zaradi razvijalcev in kasneje uporabnikov. Anketirani, so bili prav tako mnenja, da bi bilo treba razviti etiko podatkov. Sicer je v sam razvoj etičnih smernic začelo v zadnjih letih stekati veliko denarja glede na

prejšnja leta kot odraz konkurenčne prednosti za podjetja in zadovoljnejših strank. Premik je bil tudi z objavo Etičnih smernic za zaupanja vredno umetno inteligenco s strani Evropske komisije, vendar ostaja vprašanje zaupanja, ali se bodo podjetja samoregulirala, saj podatki kažejo, da v večji meri ne in se njihovi pomisleki krepijo šele, ko algoritmi negativno vplivajo na ljudi. Zato bi morale smernice postati zakoni oziroma predpisi.

Sodeč po opravljenih raziskavah glede vlaganja vlade v izobraževanja o delovanju algoritmov in regulacijo podatkovnih baz to vodi do sodelovanja med ljudmi in umetno inteligenco ter zaupanja v njo, saj drugače v splošnem ljudje bolj zaupajo ljudem, četudi se ti motijo. Zelo pomembno je ljudem omogočiti razumevanje sprejemanja odločitev s strani algoritmov, saj tako začnejo razumeti, kako poteka proces. Med izkušnjami in uporabo ter večjo transparentnostjo se poveča njihova stopnja zaupanja v tehnologijo. Z večjo vključenostjo v sam proces odločanja in omogočenim doprinosom k manjšim spremembam algoritma se stopnja zadovoljstva povečuje in zaupanje v izboljšave. Tudi praktično vsi anketirani so menili, da bi si morali prizadevati za algoritmično izobraženost. S povečevanjem izobraženosti glede delovanja algoritmov bi vplivali na skrbno načrtovan razvoj, saj bi pred samo izdelavo najprej razmislili, kaj bomo dosegli in katero pomanjkljivost rešili. Ljudi pa bi bilo treba tudi izobraziti glede potencialnih motenj in ključnih veščin opravljanja novih delovnih mest, ki jih bo ustvarila umetna inteligenca in povečevati miselnost o vseživljenjskem učenju že v šolah.

Anketiranci, ki so sodelovali v anketi, se v večji meri pri svojem delu ne srečujejo z uporabo algoritmov. V primeru, da bi se odločili za uporabo, jim predstavlja izziv varovanje zasebnosti podatkov in zagotovitev kakovosti. Tisti, ki uporabljajo algoritme, teh je 32 %, jih predvsem na področju varnostne zaščite potrjevanja transakcij kot tudi pri hitrejšem postopku izkazovanja identitete, najmanj pa pri zaposlovanju.

#### **5.4 Priporočila za nadaljnje raziskovanje**

Smiselno bi bilo analizirati etični vidik odločanja s pomočjo algoritmov na posameznih področjih po različnih starostnih skupinah, spolu, doseženi izobrazbi, saj kot je bilo že rečeno, so uporabi algoritmov bolj naklonjeni moški, mlajši ljudje bolj in v sorazmerju s stopnjo izobrazbe.

Zanimiv vidik prihodnjega raziskovanja je tudi v smeri kulturne raznolikosti in etičnih smernic. Torej to, kar se smatra kot etično sprejemljivo oziroma nesprejemljivo v določeni kulturi in kakšne zahteve imajo oblikovane, ni nujno, da velja tudi za drugo kulturo.

Z vidika zmanjševanja pristranskosti vhodnih podatkov, ki so ključni za delovanje algoritma in kasneje oblikovanja odločitev, je zanimivo pogledati, kako organizacije pristopajo k temu in kakšne prakse uporabljajo, da bi stremeli k cilju kakovostnih podatkov.



Še eno izmed področij potencialnih raziskav je lahko v smeri, kot sem že omenila v magistrskem delu, glede vpliva na krožno gospodarstvo, torej, kakšni pristopi so se do sedaj uporabljali in kako vse še lahko izkoristimo umetno inteligenco v korist, da bi pripomogli k trajnostnemu okolju z vidika energije in virov.

Vse več je tudi zahtev po transparentnosti in razložljivosti z namenom razumevanja sistemov umetne inteligence vseh deležnikov, da lahko umetno inteligenco v prvi vrsti obvladujemo in kot drugo razumemo sprejete odločitve s pomočjo algoritmov, ki imajo vpliv na življenje posameznika. Tako bi bilo dobro raziskati mnenje ljudi glede trenutne stopnje transparentnosti in njihovo razumevanje sprejetih odločitev ter kako bi se morale razviti to področje, da bi dosegalo visoko stopnjo razumevanja, med drugim tudi s pomočjo večje algoritmične izobraženosti.

Na koncu bi predlagala tudi raziskanje področja odgovornosti. S tem mislim natančno opredelitev, kdo v verigi je odgovoren za sisteme umetne inteligence oziroma kako bi se morala odgovornost porazdeliti. Ravno tako bi bilo treba raziskati, kaj lahko posameznik pričakuje v primeru, da je oškodovan v skladu z zakoni in predpisi, in kakšni so primeri resničnih praks, v primeru natančne opredelitve in tako ne samo prelaganja odgovornosti pa tudi, kako bi se povečalo zaupanje in vrednost sistemov umetne inteligence.

## **SKLEP**

Skupek razpoložljivih podatkov, algoritmov in zmogljivosti, ki jih premorejo računalniki današnjega časa, imenujemo umetna inteligenca. Z njenim razvojem postaja problematika odločanja s pomočjo algoritmov kompleksnejša. Vse večji so pozivi k etičnosti splošne uporabe kot tudi znotraj posameznih strok. Prav tako prizadevanja za rast algoritmične izobraženosti za vzpostavitev večjega zaupanja. Razumevanje učenja algoritmov umetne inteligence je pomembno za vzpostavljanje etičnih smernic. Umetna inteligenca pa nima vpliva samo z vidika posameznika, temveč tudi celotne družbe.

Glede na opravljeno raziskavo se anketiranci v večini zavedajo splošne vpetosti algoritmov v naša življenja, pa vendar v večji meri še vedno ne pomembnih specifičnih situacij, ki krojijo naša življenja. Skozi podane primere uporabe sem prišla do različnih ugotovitev, kateri načini uporabe se jim zdijo etično sprejemljivi in kateri ne. Tako se jim uporaba za namene zaposlovanja kot pomoč pravosodnim organom zdi etično nesprejemljiva. Skoraj neopredeljeni so glede uporabe v bančništvu in zavarovalništvu kot tudi pri avtonomnih vozilih. Za razliko glede uporabe v zdravstvu in pri zbiranju ter obdelavi biometričnih podatkov, kjer se jim zdi etično sprejemljiva. Pomembno jim je tudi, da so odločitve transparentne, da se izvajata nadzor in zaščita, spodbuja algoritmično izobraževanje, na kakšen način delujejo algoritmi, in da se uredi etika podatkov. Čeprav je rezultat skoraj neodločen glede zaupanja v kakovostne in nepristranske podatke, pa je splošno mnenje, da bo pretehtal pozitiven vpliv.

Za zaključek bom odgovorila na raziskovalno vprašanje, kaj velja za etično sprejemljivo odločanje s pomočjo algoritmov. Glede na podane tako pozitivne kot tudi negativne vidike odločanja pri posameznih primerih uporabe sem prišla do sklepa, da je uporaba algoritmov umetne inteligence smiselna in koristna za vse deležnike, če jo uporabljamo preudarno in razrešimo odprte vrzeli, med drugim kakovost vnesenih podatkov razvijalcev za zmanjšanje diskriminacije in pristranskosti, vprašanje odgovornosti posledic odločanja z algoritmi, tveganja z naslova kršitev temeljnih pravic in morebitne zlorabe s kontinuiranim večanjem uporabe. Hkrati pa si je treba prizadevati za vzajemno delovanje skupaj s sistemi, da preprečimo izgubo človeške presoje in zamenjavo ljudi z algoritmi, transparentnost odločanja za višjo stopnjo zaupanja v sisteme, nadzor nad celotnim delovanjem in prizadevanja za algoritmično izobraževanje, s katerim se poveča zaupanje ljudi v umetno inteligenco ter omogoči ljudem razumevanje sprejetih odločitev algoritmov.

## LITERATURA IN VIRI

1. Abate, T. (2016, 1. september). *Stanford-hosted study examines how AI might affect urban life in 2030*. Pridobljeno 15. junija 2020 iz <https://news.stanford.edu/2016/09/01/ai-might-affect-urban-life-2030/>
2. Alang, N. (2019, 3. april). Algorithms have gotten out of control. It's time to regulate them. *The week*. Pridobljeno 14. aprila 2021 iz <https://theweek.com/articles/832948/algorithms-have-gotten-control-time-regulate>
3. Anderson, M. (2022, 18. februar). *'Simple' AI Can Anticipate Bank Managers' Loan Decisions to Over 95 % Accuracy*. Pridobljeno 13. aprila 2022 iz <https://www.unite.ai/simple-ai-can-anticipate-bank-managers-loan-decisions-to-over-95-accuracy/>
4. Batok, N. (2020, 30. november). Artificial Intelligence has changed our world. *Meer*. Pridobljeno 22. januarja 2021 iz <https://wsimag.com/science-and-technology/64215-artificial-intelligence-has-changed-our-world>
5. Battelle Insider. (2019, 19. oktober). *The Algorithm Is In: 5 Ways AI is Transforming Medicine* [objava na blogu]. Pridobljeno 14. aprila 2021 iz <https://inside.battelle.org/blog-details/the-algorithm-is-in-5-ways-ai-is-transforming-medicine>
6. Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E. & Winfield, A. (2020, marec). The ethics of artificial intelligence: Issues and initiatives. *Panel for the Future of Science and Technology (STOA)*. Pridobljeno 3. aprila 2020 iz [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_STU\(2020\)634452](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)634452)
7. Boesch, G. (brez datuma). *Face Detection in 2021: Real-time applications with deep learning*. Pridobljeno 10. junija 2021 iz <https://viso.ai/deep-learning/face-detection-overview/>
8. Bogert, E., Schechter, A. & Watson, R. T. (2021, april). Humans Rely More on Algorithms than Social Influence as a Task Becomes More Difficult. *Scientific Reports*

- 11(1). Pridobljeno 10. maja 2021 iz <https://www.nature.com/articles/s41598-021-87480-9>
9. Burns, E. (2021, marec). *Machine learning*. Pridobljeno 15. junija 2021 iz <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML>
  10. Burns, E. & Brush, K. (2021, marec). *Deep learning*. Pridobljeno 10. junija 2021 iz <https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network>
  11. Candelon, F., Charne di Carlo, R., De Bondt, M. & Evgenious T. (2021, september-oktober). AI Regulation Is Coming. *Harvard Business Review*. Pridobljeno 15. decembra 2021 iz <https://hbr.org/2021/09/ai-regulation-is-coming>
  12. Castelluccia C. & Le Métayer D. (2019, 3. marec). Understanding algorithmic decision-making: Opportunities and challenges. *Panel for the Future of Science and Technology (STOA)*. Pridobljeno 15. aprila 2020 iz [https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS\\_STU\(2019\)624261](https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2019)624261)
  13. Castelvechi, D. (2019, 4. april). AI pioneer: 'The dangers of abuse are very real'. *Nature*. Pridobljeno 10. oktobra 2021 iz <https://www.nature.com/articles/d41586-019-00505-2>
  14. Charlton, E. (2019, 1. julij). *Artificial Intelligence: These 3 charts show what people really think*. Pridobljeno 10. oktober 2020 <https://www.weforum.org/agenda/2019/07/artificial-intelligence-these-3-charts-show-what-people-really-think/>
  15. Cheatham, B., Javanmardian, K. & Samandari, H. (2019, 16. april). *Confronting the risks of artificial intelligence*. Pridobljeno 17. junija 2020 iz <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence>
  16. Claudé, M. & Combe, D. (2018, 24. maj). *The roles of artificial intelligence and humans in decision making : towards augmented humans?* Umeå: Umeå School of Business, Economics and Statistics.
  17. Colson, E. (2019, 8. julij). What AI-Driven Decision Making Looks Like. *Harvard Business Review*. Pridobljeno 20. oktobra 2020 iz <https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like>
  18. Copeland, J. (2000, maj). *What is Artificial Intelligence?* Pridobljeno 10. maja 2020 iz [http://www.alanturing.net/turing\\_archive/pages/Reference%20Articles/What\\_is\\_AI/What%20is%20AI03.html](http://www.alanturing.net/turing_archive/pages/Reference%20Articles/What_is_AI/What%20is%20AI03.html)
  19. Copeland, M. (2016, 29. julij). *What's the Difference Between Artificial Intelligence, Machine Learning and Deep Learning?* Pridobljeno 11. februar 2021 iz <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>
  20. Davenport, T. & Kalakota, R. (2019, 2. junij). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94-98.

21. De Jonghe, F. & Latinne, P. (2019, 24. september). *When and how will society start trusting AI?* Pridobljeno 13. januarja 2022 iz [https://www.ey.com/en\\_be/ai/when-and-how-will-society-start-trusting-ai](https://www.ey.com/en_be/ai/when-and-how-will-society-start-trusting-ai)
22. Delakorda, S. (2019, 2. december). *Etične smernice za zaupanja vredno umetno inteligenco*. Pridobljeno 14. junija 2021 iz [https://www.ey.com/en\\_be/ai/when-and-how-will-society-start-trusting-ai](https://www.ey.com/en_be/ai/when-and-how-will-society-start-trusting-ai)
23. DeMarco, J. M. (2019, 13. september). *Artificial Intelligence (AI): What it is, Why it Matters, What You Can Do*. Pridobljeno 10. maja 2020 iz <http://johnmichaeldemarco.com/artificial-intelligence-ai-what-it-is-why-it-matters-what-you-can-do/>
24. Drevenšek, S. (2018, 14. februar). *Bodo algoritmi res boljši sodniki?* Pridobljeno 15. decembra 2021 iz <https://svetkapitala.delo.si/ikonomija/bodo-algoritmi-res-boljsi-sodniki/>
25. Drevenšek, S. (2019, 9. marec). *Človeška stran avtonomnih avtomobilov*. Pridobljeno 10. septembra 2021 iz <https://svetkapitala.delo.si/mobilnost/cloveska-stran-avtonomnih-avtomobilov/>
26. Edmonds, E. (2019, 14. marec). *Three in Four Americans Remain Afraid of Fully Self-Driving Vehicles*. Pridobljeno 15. septembra 2020 iz <https://newsroom.aaa.com/2019/03/americans-fear-self-driving-cars-survey/>
27. European Commission. (2019a, 8. april). *Artificial intelligence: Commission takes forward its work on ethics guidelines*. Pridobljeno 4. aprila 2021 iz [https://ec.europa.eu/commission/presscorner/detail/sl/IP\\_19\\_1893](https://ec.europa.eu/commission/presscorner/detail/sl/IP_19_1893)
28. European Commission. (2019b, 8. april). *Ethics Guidelines for Trustworthy AI*. Pridobljeno 3. aprila 2021 iz <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>
29. European Commission. (2020, 19. februar). *White Paper on Artificial Intelligence: a European approach to excellence and trust*. Pridobljeno 20. aprila 2020 iz [https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust\\_en](https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en)
30. European Union for Fundamental Rights. (2019, 7. junij). *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights*. Pridobljeno 21. januarja 2022 iz <https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect>
31. European Union for Fundamental Rights. (2021, 29. januar). *Getting the future right: Artificial intelligence and fundamental rights*. Pridobljeno 3. maja 2021 iz <https://fra.europa.eu/en/publication/2021/getting-future-right-artificial-intelligence-and-fundamental-rights-summary>
32. Evropska komisija. (2018, 17. maj). *Na poti do avtomatizirane mobilnosti: strategija EU za mobilnost prihodnosti*. Pridobljeno 10. septembra 2021 iz <https://eur-lex.europa.eu/legal-content/SL/TXT/?uri=CELEX%3A52018DC0283>
33. Evropska komisija. (2020, 19. februar). *Bela knjiga o umetni inteligenci – evropski pristop k odličnosti in zaupanju*. Pridobljeno 4. maja 2021 iz

<https://op.europa.eu/sl/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1>

34. Evropski parlament. (2019, 14. januar). *Samovozeči avtomobili v Evropi: od znanstvene fantastike do realnosti*. Pridobljeno 11. septembra 2021 iz <https://www.europarl.europa.eu/news/sl/headlines/economy/20190110STO23102/sa-movozeci-avtomobili-v-evropi-od-znanstvene-fantastike-do-realnosti>
35. Evropski parlament. (2020, 26. februar). *Umetna inteligenca: odpravljanje tveganj za potrošnike*. Pridobljeno 2. februarja 2022 iz <https://www.europarl.europa.eu/news/sl/headlines/society/20200206STO72030/umetna-inteligenca-odpravljanje-tveganj-za-potrosnike>
36. Faggella, D. (2019, 2. februar). *AI for Crime Prevention and Detection – 5 Current Applications*. Pridobljeno 10. junija 2021 iz <https://emerj.com/ai-sector-overviews/ai-crime-prevention-5-current-applications/>
37. Felzmann, H., Fosch Villaronga, E., Lutz, C. & Tamò-Larrieux, A. (2019, 27. junij). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *SAGE Journals*, 6(1). Pridobljeno 20. januarja 2022 iz <https://journals.sagepub.com/doi/full/10.1177/2053951719860542>
38. Floridi, L. & Lee, M. S. A. (2020, 9. Junij). Algorithmic Fairness in Mortgage Lending: from Absolute Conditions to Relational Trade-offs. *Minds & Machines*, 31, 165–191.
39. Ford, T. W. (2021, 10. november). It's time to address facial recognition, the most troubling law enforcement AI tool. *Bulletin of the Atomic Scientists*. Pridobljeno 10. junija 2021 iz <https://thebulletin.org/2021/11/its-time-to-address-facial-recognition-the-most-troubling-law-enforcement-ai-tool/>
40. Garcia, M. (2017). Racist in the Machine: The Disturbing Implications of Algorithmic Bias. *World Policy Journal*, 33(4), 111-117.
41. Gonzales, W. J., (2017, 12. september). Artificial Intelligence in a New Context: “Internal” and “External” Factors. *Minds & Machines*, 27, 393-396.
42. Great Learning Team. (2022, 19. januar). *What is Artificial Intelligence? How does AI work, Types and Future of it?* [objava na blogu]. Pridobljeno 11. februarja 2022 iz <https://www.mygreatlearning.com/blog/what-is-artificial-intelligence/>
43. Hansson, S. O., Belin, M. A. & Lundgren, B. (2021, 12. avgust). Self-Driving Vehicles-an Ethical Overview. *Philosophy & Technology*. 34, 1383–1408.
44. Hebbar, P. (2017, 1. september). 8 Real Life Examples When Algorithms Turned Rogue, Causing Disastrous Results. *Analytics India Magazine*. Pridobljeno 17. decembra 2021 iz <https://analyticsindiamag.com/8-real-life-examples-algorithms-turned-rogue-causing-disastrous-results/>
45. Henigman, U. (2019, 27. december). *Biometrični osebni podatki – dokončna izguba zasebnosti?* Pridobljeno 10. februarja 2021 iz <https://www.rtv slo.si/znanost-in-tehnologija/biometricni-osebni-podatki-dokoncna-izguba-zasebnosti/510211>
46. Hoes, F. (2019, 30. december). The Importance of Ethics in Artificial Intelligence. *Towards Data Science*. Pridobljeno 3. aprila 2021 iz

- <https://towardsdatascience.com/the-importance-of-ethics-in-artificial-intelligence-16af073dedf8>
47. Huš, M. (2019, 30. april). Preračunavam...: šest let zapora. *Monitor*. Pridobljeno 5. maja 2021 iz <https://www.monitor.si/clanek/preracunavam-sest-let-zapora/190818/>
  48. IBM. (brez datuma a). *Accountability*. Pridobljeno 6. aprila 2021 iz <https://www.ibm.com/design/ai/ethics/accountability/>
  49. IBM. (brez datuma b). *Building trust in AI*. Pridobljeno 12. marca 2022 iz <https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/building-trust-in-ai.html>
  50. IBM. (brez datuma c). *The new AI innovation equation*. Pridobljeno 14. novembra 2021 iz <https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-innovation-equation.html>
  51. Igoe, K. J. (2021, 12. marec). *Algorithmic Bias in Health Care Exacerbates Social Inequities - How to Prevent It*. Boston: Harvard T.H. Chan School of Public Health.
  52. Informacijski pooblaščenec. (brez datuma). *Prijava biometrijskih ukrepov*. Pridobljeno 10. junija 2021 iz <https://www.ip-rs.si/varstvo-osebni-podatkov/obveznosti-upravljavcev/prijava-biometrijskih-ukrepov/#c395>
  53. Islam, S. (2015, 10. marec). An Algorithm for Electronic Money Transaction Security (Three Layer Security): A New Approach. *International Journal of Information and Computer Security* 9(2), 203–214.
  54. Johnson, C. & Tyson, A. (2020, 18. december). *Here's how opinions on the impact of artificial intelligence differ around the world*. Pridobljeno 6. februarja 2021 iz <https://www.weforum.org/agenda/2020/12/mixed-views-of-the-impact-of-artificial-intelligence>
  55. Kayser-Bril, N. (2019a, 6. oktober). *Austria's employment agency rolls out discriminatory algorithm, sees no problem*. Pridobljeno 7. februarja 2021 iz <https://algorithmwatch.org/en/austrias-employment-agency-ams-rolls-out-discriminatory-algorithm/?fbclid=IwAR2o8viPcnFAAx8Eclhijv3-ak5gkDZahlk8eApaghNlccCasIY7tmX1msk>
  56. Kayser-Bril, N. (2019b, 11. december). *At least 10 police forces use face recognition in the EU, AlgorithmWatch reveals*. Pridobljeno 15. aprila 2020 iz <https://algorithmwatch.org/en/story/face-recognition-police-europe/>
  57. Knowledge at Wharton Staff. (2019, 25. marec). *Who Made That Decision: You or an Algorithm?* Pennsylvania: Wharton School of the University of Pennsylvania.
  58. Kojić, T. (2018, 21. maj). Umetna Inteligenca: Zaveznica ali groznja? *Časnik Finance*. Pridobljeno 11. oktobra 2021 iz <https://viva.finance.si/8990707/Umetna-Inteligenca-Zaveznica-ali-groznja>
  59. Kolednik, A. (2019, 5. marec). *Umetna inteligenca: Slovenci smo v nekaterih nišah v svetovnem vrhu*. Pridobljeno 9. aprila 2022 iz <https://siol.net/novice/slovenija/bo-umetna-inteligenca-popolnoma-nadomestila-zdravnike-491750>
  60. Križnar, P. (2019). Kazenske sankcije in algoritmi: Kaznovanje v1.01. *Revija za kriminalistiko in kriminologijo*, 70(1), 21–33.

61. Kumar, S. (2019, 25. november). Advantages and Disadvantages of Artificial Intelligence. *Towards Data Science*. Pridobljeno 7. julija 2021 iz <https://towardsdatascience.com/advantages-and-disadvantages-of-artificial-intelligence-182a5ef6588c>
62. Kundu, S. (2019, 3. julij). *Ethics in the Age of Artificial Intelligence* [objava na blogu]. Pridobljeno 12. aprila 2021 iz <https://blogs.scientificamerican.com/observations/ethics-in-the-age-of-artificial-intelligence/>
63. Kušar, P. (2018, 21. september). *Kakšno vlogo ima umetna inteligenca v našem vsakdanu?* Pridobljeno 5. februarja 2022 iz <https://mladipodjetnik.si/novice-in-dogodki/novice/kaksno-vlogo-ima-umetna-inteligenca-v-nasem-vsakdanu>
64. Lasry, B., Kobayashi, H. & Unesco. (2018). *Human decisions: thoughts on AI*. Paris: United Nations Educational, Scientific and Cultural Organization.
65. Leetaru, K. (2016, 4. januar). In Machines We Trust: Algorithms Are Getting Too Complex To Understand. *Forbes*. Pridobljeno 10. marca 2022 iz <https://www.forbes.com/sites/kalevleetaru/2016/01/04/in-machines-we-trust-algorithms-are-getting-too-complex-to-understand/?sh=5d91285633a5>
66. Lunter, J. (2021, 1. november). *Using and storing biometric data: Scrutinizing practices of banks and governments*. Pridobljeno 12. decembra 2021 iz <https://www.biometricupdate.com/202111/using-and-storing-biometric-data-scrutinizing-practices-of-banks-and-governments>
67. Madhavan, R. (2020, 4. december). *AI in Biometrics and Security – Current Business Applications*. Pridobljeno 20. junija 2021 iz <https://emerj.com/ai-sector-overviews/ai-in-biometrics-current-business-applications/>
68. Martin, K. (2018, 7. junij). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, 160, 835–850.
69. Martin, K. E. (2019, junij). Designing Ethical Algorithms. *MIS Quarterly Executive*, 18(2), 129-142.
70. Martínez-Miranda, J. & Aldea, A. (2005, 2. marec). Emotions in human and artificial intelligence. *Computers in Human Behavior*, 21(2), 323-341.
71. Mazurek, M. (brez datuma). *Future of algorithm*. Pridobljeno 2. aprila 2021 iz <https://fully-verified.com/future-of-algorithm/>
72. Mayer-Schönberger, V. & Cukier, K. (2013, januar). *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. New York: Houghton Mifflin Harcourt Publishing Company.
73. Michael. (2019, 8. maj). In the future, you may be fired by an algorithm. *Towards Data Science*. Pridobljeno 2. aprila 2021 iz <https://towardsdatascience.com/in-the-future-you-may-be-fired-by-an-algorithm-35aefd00481f>
74. Mihajlovic, I. (2019, 25. april). Everything You Ever Wanted To Know About Computer Vision. *Towards Data Science*. Pridobljeno 29. marca 2021 iz <https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e>

75. Mikalef, P., Conboy, K., Eriksson Lundström, J. & Popovič, A. (2022, 11. Februar). Thinking responsibly about responsible AI and ‘the dark side’ of AI. *European Journal of Information Systems*, 31(3), 257–268.
76. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, D. (2016, 1. december). The ethics of algorithms: Mapping the debate. *Big data & Society*, 3(2), 1-21.
77. Nalini, B. (2019, 1. maj). The Hitchhiker’s Guide to AI Ethics. *Towards Data Science*. Pridobljeno 2. aprila 2021 iz <https://towardsdatascience.com/ethics-of-ai-a-comprehensive-primer-1bfd039124b0>
78. Nexus Integra. (brez datuma). *Advantages and disadvantages of artificial intelligence*. Pridobljeno 17. decembra 2021 iz <https://nexusintegra.io/advantages-disadvantages-artificial-intelligence/>
79. Nielly, C. (2020, 12. februar). Can we let algorithm take decisions we cannot explain? *Towards Data Science*. Pridobljeno 11. decembra 2021 iz <https://towardsdatascience.com/can-we-let-algorithm-take-decisions-we-cannot-explain-a4e8e51e2060>
80. Nyholm, S. & Smids, J. (2016, 28. julij). The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem? *Ethical Theory and Moral Practice*, 19, 1275–1289.
81. Olavsrud, T. (2021, 23. november). 6 famous analytics and AI disasters. *CIO*. Pridobljeno 30. novembra 2021 iz <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>
82. Open Access Government. (2021, 3. december). *The new government standard for algorithmic transparency*. Pridobljeno 22. januar 2022 iz <https://www.openaccessgovernment.org/the-new-government-standard-for-algorithmic-transparency/125343/>
83. O'Sullivan, C. (2021, 20. maj). What is Responsible AI? *Towards Data Science*. Pridobljeno 30. avgusta 2021 iz <https://towardsdatascience.com/what-is-responsible-ai-548743369729>
84. Pandey, A. (2022, 23. Februar). *Banking the Unbanked Carries Risk (That AI Modeling Can Mitigate)*. Pridobljeno 6. marca 2022 iz <https://www.corporatecomplianceinsights.com/banking-unbanked-carries-risk-ai-modeling-can-mitigate/>
85. Pandya, J. (2019, 20. januar). Can Artificial Intelligence Be Biased? *Forbes*. Pridobljeno 17. julija 2020 iz <https://www.forbes.com/sites/cognitiveworld/2019/01/20/can-artificial-intelligence-be-biased/?sh=56815ce47e7c>
86. Pazzanese, C. (2020, 26. oktober). Ethical concerns mount as AI takes bigger decision-making role in more industries. *The Harvard Gazette*. Pridobljeno 2. aprila 2021 iz <https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/>
87. Picard, R. W. (1995). *Affective Computing*. Cambridge: MIT Media Laboratory, Perceptual Computing.



88. Podpečan, M. (2019, 8. januar). Umetna inteligenca in sojenje. *Odvetnik*, 88(5). Pridobljeno 16. decembra 2021 iz <https://www.tax-finlex.si/Publikacije/TfIGlasnik/f716f66c-0386-41bc-9da8-b6845e9de7ed>
89. Polonski, V. (2018, 10. januar). People Don't Trust AI-Here's How We Can Change That. *Scientific American*. Pridobljeno 12. oktobra 2021 iz <https://www.scientificamerican.com/article/people-dont-trust-ai-heres-how-we-can-change-that/>
90. Požnel, B. (2018, 26. april). AMZS. #Video: Šest stopenj avtonomne vožnje. Pridobljeno 9. septembra 2021 iz <https://www.amzs.si/motorevija/v-zarometu/avtomoto/2018-04-26-video-sest-stopenj-avtonomne-voznje>
91. Prabhu, R. (2022, 13. januar). *Introduction to Algorithms*. Pridobljeno 15. februarja 2022 iz <https://www.geeksforgeeks.org/introduction-to-algorithms/>
92. PwC. (2021). *PwC's Responsible AI*. Pridobljeno 22. januarja 2022 iz <https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html>
93. Quinyx. (brez datuma). *What's The Difference Between AI, ML, and Algorithms* [objava na blogu]. Pridobljeno 17. julija 2020 iz <https://www.quinyx.com/blog/difference-between-ai-ml-algorithms>
94. Rainie, L. & Anderson, J. (2017, 8. januar). *Code-Dependent: Pros and Cons of the Algorithm Age*. Pridobljeno 18. julija 2020 iz <https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/>
95. Rajšek, J. (2019, 23. oktober). *Veliki evropski brat se prebuja—"Obraza ne moremo zamenjati"*. Pridobljeno 7. junija 2021 iz <https://www.rtv slo.si/evropska-unija/veliki-evropski-brat-se-prebuja-obraza-ne-moremo-zamenjati/502913>
96. Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C. & Levin, K. (2018, 25. september). *Artificial Intelligence & Human Rights: Opportunities & Risks*. Cambridge: Berkman Klein Center for Internet & Society Research Publication.
97. Ravindran, K. (2020, 10. februar). The Two Sides of Artificial Intelligence. *IE Insights*. Pridobljeno 5. oktobra 2021 iz <https://www.ie.edu/insights/articles/the-two-sides-of-artificial-intelligence/>
98. Rock Content Writer. (2021, 28. junij). *Artificial Intelligence Algorithm: Everything You Need To Know About It* [objava na blogu]. Pridobljeno 15. oktobra 2021 iz <https://rockcontent.com/blog/artificial-intelligence-algorithm/>
99. Roos, D. (2019, 22. maj). *How Alan Turing and His Test Became AI Legend*. Pridobljeno 18. julija 2021 iz <https://electronics.howstuffworks.com/future-tech/alan-turing-test.htm>
100. Ross, M. & Taylor J. (2021, 10. november). Managing AI Decision-Making Tools. *Harvard Business Review*. Pridobljeno 7. septembra 2021 iz <https://hbr.org/2021/11/managing-ai-decision-making-tools>

101. Ruiz, R. (2020, 25. september). *Doctors use algorithms that aren't designed to treat all patients equally*. Pridobljeno 12. novembra 2021 iz <https://mashable.com/article/healthcare-algorithms-racism>
102. Ryan, M. & Stahl, B. C. (2021, 3. marec). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information Communication and Ethics in Society*, 19(1), 61-86.
103. Samek, W., Wiegand, T. & Müller, K. R. (2017, 28. avgust). Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. *arXiv*. Pridobljeno iz <https://arxiv.org/abs/1708.08296>
104. Savič, D. (2018, 27. februar). Algoritmi vam bodo snedli obraz. *Monitor*. Pridobljeno 14. junija 2021 iz <https://www.monitor.si/clanek/algoritmi-vam-bodo-snedli-obraz/184307/>
105. Savič, D. (2019, 26. junij). Geslo: vaš obraz. *Monitor*. Pridobljeno 14. junija 2021 iz <https://www.monitor.si/clanek/geslo-vas-obraz/191660/>
106. Scott, A. (2021, 7. julij). *Difference Between Algorithm and Artificial Intelligence*. Pridobljeno 18. septembra 2021 iz <https://www.datasciencecentral.com/difference-between-algorithm-and-artificial-intelligence/>
107. Schrauth, B. & Funk, W. (2021, 1. marec). Slovenci naklonjeni avtonomnim vozilom. *AMZS*. Pridobljeno 9. septembra 2021 iz <https://www.amzs.si/motorevija/v-zarometu/avto-moto/2021-03-01-slovenci-naklonjeni-avtonomnim-vozilom>
108. Schwartz, O. (2019, 1. april). Untold History of AI: Why Alan Turing Wanted AI Agents to Make Mistakes Infallibility isn't the same thing as intelligence. *IEEE Spectrum*. Pridobljeno 25. julija 2020 iz <https://spectrum.ieee.org/untold-history-of-ai-why-alan-turing-wanted-ai-to-make-mistakes#toggle-gdpr>
109. Selig J. (2022, 14. marec). *What Is Machine Learning? A Definition* [objava na blogu]. Pridobljeno 17. aprila 2022 iz <https://www.expert.ai/blog/machine-learning-definition/>
110. Sherman, S. (brez datuma). *Algorithms and transparency: leaving no one behind in the Digital Age* [objava na blogu]. Pridobljeno 18. januarja 2022 iz <https://www.objective.com.au/resources/blog-algorithms-and-transparency-leaving-no-one-behind-in-the-digital-age>
111. Smith, M. & Miller, S. (2021, 13. april). The ethical application of biometric facial recognition technology. *AI & Society*, 37, 167-175.
112. Stromboli. (2021, 25. april). Evropska unija pripravlja strožje omejitve in ukrepe za uporabo umetne inteligence. *Računalniške novice*. Pridobljeno 2. maja 2021 iz <https://racunalniske-novice.com/evropska-unija-pripravlja-strojje-omejitve-in-ukrepe-za-uporabo-umetne-inteligence/>
113. Synopsis. (brez datuma). *What is an Autonomous Car?* Pridobljeno 10. september 2021 iz <https://www.synopsys.com/automotive/what-is-autonomous-car.html>
114. Šprohar, N. (2019, 23. december). Največji izziv je ugotoviti, kako umetna inteligenca pride do rešitev. *Glas gospodarstva GG*. Pridobljeno 20. aprila 2020 iz

- [https://www.gzs.si/zdruzenje\\_za\\_informatiko\\_in\\_telekomunikacije/Novice/ArticleId/74032/najvecji-izziv-je-ugotoviti-kako-umetna-inteligenca-pride-do-resitev](https://www.gzs.si/zdruzenje_za_informatiko_in_telekomunikacije/Novice/ArticleId/74032/najvecji-izziv-je-ugotoviti-kako-umetna-inteligenca-pride-do-resitev)
115. Thales. (2021, 24. junij). *Face recognition – fascinating and intriguing*. Pridobljeno 7. julija 2021 iz <https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition>
  116. The Medical Futurist. (2019, 5. februar). *Top Smart Algorithms In Healthcare*. Pridobljeno 16. novembra 2021 iz <https://medicalfuturist.com/top-ai-algorithms-healthcare/>
  117. The Public Voice. (2018, oktober). *Universal Guidelines for Artificial Intelligence*. Pridobljeno 20. marca 2022 iz <https://thepublicvoice.org/ai-universal-guidelines/memo/>
  118. Tranberg, P. (2017, 18. februar). *Experts On The Pros & Cons of Algorithms*. Pridobljeno 5. maja 2020 iz <https://dataethics.eu/prosconsai/>
  119. Trend Micro. (2020, 19. november). *Exploiting AI: How Cybercriminals Misuse and Abuse AI and ML*. Pridobljeno 14. aprila 2021 iz <https://www.trendmicro.com/vinfo/de/security/news/cybercrime-and-digital-threats/exploiting-ai-how-cybercriminals-misuse-abuse-ai-and-ml>
  120. Triglav. (2016, december). Bi prepustili odločitve svojemu avtu? *Naložbenik*, 16. Pridobljeno 7. septembra 2021 iz <https://vsebovredn.triglav.si/tehnologija/bi-prepustili-odlocitve-svojemu-avtu>
  121. United Nations. (2021, 15. september). *Urgent action needed over artificial intelligence risks to human rights*. Pridobljeno 20. novembra 2021 iz <https://news.un.org/en/story/2021/09/1099972>
  122. Vasegaard, A. E. (2021, 9. september). The Future of Decision-Making in AI. *Towards Data Science*. Pridobljeno 10. januarja 2022 iz <https://towardsdatascience.com/the-future-of-decision-making-in-ai-17576cdb6e66>
  123. Walch, K. (2019, 29. december). Ethical Concerns of AI. *Forbes*. Pridobljeno 2. aprila 2021 iz <https://www.forbes.com/sites/cognitiveworld/2020/12/29/ethical-concerns-of-ai/?sh=13617f2f23a8>
  124. Warnock, C. (2019, 16. september). *Power, speed, and peril: welcome to the world of algorithms* [objava na blogu]. Pridobljeno 15. aprila 2021 iz <https://www.su.org/blog/welcome-to-the-future-of-algorithms-and-machine-learning>
  125. Watson, H. (2019, januar). Addressing the Growing Need for Algorithmic Transparency. *Communications of the Association for Information System*, 488-510.
  126. Weitzner, D. J. (2019, 19. februar). *AI Tool Builders and Their Users: What Should We Expect From the Tools and Who Is Responsible When They Fail?* Pridobljeno 10. oktobra 2021 iz <https://internetpolicy.mit.edu/ai-tool-builders-and-their-users/>
  127. Xu, T. (2021, 19. julij). *AI Makes Decisions We Don't Understand. That's a Problem*. Pridobljeno 7. oktobra 2021 iz <https://builtin.com/artificial-intelligence/ai-right-explanation>
  128. Završnik, A. (2017). Algoritmčno nadzorstvo: veliko podatkovje, algoritmi in družbeni nadzor. *Revija za kriminalistiko in kriminologijo*, 2017(2), 68, 135–149.

129. Zetlin, M. (brez datuma). AI Is Now Analyzing Candidates' Facial Expressions During Video Job Interviews. *Inc.* Pridobljeno 10. junija 2021 iz <https://www.inc.com/minda-zetlin/ai-is-now-analyzing-candidates-facial-expressions-during-video-job-interviews.html>

## **PRILOGE**



## Priloga 1: Tabele

Tabela 1: Značilnosti vzorca

Spol	f	f (%)
Moški	83	41,50 %
Ženski	117	58,50 %
Skupaj	200	100,00 %

Starostna skupina	f	f (%)
1 (18–24 let)	32	16,00 %
2 (25–34 let)	73	36,50 %
3 (35–44 let)	26	13,00 %
4 (45–54 let)	50	25,00 %
5 (nad 55 let)	19	9,50 %
Skupaj	200	100,00 %

Najvišja dosežena izobrazba	f	f (%)
1 (I. stopnja izobrazbe – nedokončana OŠ)	1	0,50 %
2 (II. stopnja izobrazbe – OŠ)	0	0,00 %
3 (III. stopnja izobrazbe – nižje poklicno izobraževanje (2 leti))	1	0,50 %
4 (IV. stopnja izobrazbe – srednje poklicno izobraževanje (3 letno))	3	1,50 %
5 (V. stopnja izobrazbe – gimnazijsko, srednje poklicno–tehniško izobraževanje, srednje tehniško oz. drugo strokovno izobraževanje)	45	22,50 %
6 (VI. stopnja izobrazbe – višješolski strokovni program, specializacija po višješolskem programu, visokošolski strokovni program oz. po bolonjskih programih visokošolski strokovni in univerzitetni program (1. bol. st.))	82	41,00 %
7 (VII. stopnja izobrazbe – specializacija po visokošolskem strokovnem programu, univerzitetni program oz. po bolonjskih programih magisterij stroke (2. bol.st.))	55	27,50 %
8 (VIII. stopnja izobrazbe – specializacija po univerzitetnem programu, magisterij znanosti, doktorat znanosti (3. bol. st.))	13	6,50 %
Skupaj	200	100,00 %

Smer poklica	f	f (%)
Splošne izobraževalne aktivnosti/izidi	25	11,47 %
Izobraževalne vede in izobraževanje učiteljev	25	11,47 %
Umetnost in humanistika	8	3,67 %
Družbene, poslovne, upravne in pravne vede	59	27,06 %
Naravoslovje, matematika in računalništvo	24	11,01 %
Tehnika, proizvodne tehnologije in gradbeništvo	22	10,09 %
Kmetijstvo, gozdarstvo, ribištvo, veterinarstvo	3	1,38 %
Zdravstvo in sociala	19	8,72 %
Storitve (osebne, transportne, varstvo okolja, varovanje)	13	5,96 %
Neopredeljeno	20	9,17 %
Skupaj	218	100,00 %

Sektor dela	f	f (%)
Javni sektor	66	32,67 %
Zasebni sektor	120	59,41 %
Drugo	16	7,92 %
Skupaj	202	100,00 %

Vir: lastno delo.

*Tabela 2: Poznavanje izraza umetna inteligenca*

Poznavanje izraza umetna inteligenca	f	f (%)
Da	220	97,78 %
Ne	5	2,22 %
Skupaj	225	100,00 %

*Vir: lastno delo.*

*Tabela 3: Zavedanje o nezavedni uporabi umetne inteligence*

Zavedanje o nezavedni uporabi umetne inteligence	f	f (%)
Da	198	88,00 %
Ne	27	12,00 %
Skupaj	225	100,00 %

*Vir: lastno delo.*

*Tabela 4: Razmišljanje o odločanju algoritmov umetne inteligence namesto ljudi*

Razmišljanje o odločanju algoritmov umetne inteligence namesto ljudi	f	f (%)
Da	133	59,38 %
Ne	91	40,63 %
Skupaj	224	100,00 %

*Vir: lastno delo.*

*Tabela 5: Pogled na trditve*

Pogled na trditve	f	f (%)
Možno je, da računalniški programi sprejemajo odločitve brez človeške pristranskosti.	103	46,19 %
Računalniški programi bodo vedno odražali pristranskost ljudi, ki so jih zasnovali.	120	53,81 %
Skupaj	223	100,00 %

*Vir: lastno delo.*

*Tabela 6: Mnenje o etični pravičnosti tovrstnega programa*

Mnenje o etični pravičnosti tovrstnega programa	f	f (%)
Zelo pravičen	6	2,76 %
Pretežno pravičen	62	28,57 %
Pretežno nepravičen	100	46,08 %
Zelo nepravičen	49	22,58 %
Skupaj	217	100,00 %

*Vir: lastno delo.*



*Tabela 7: Mnenje o učinkovitosti tovrstnega programa*

Mnenje o učinkovitosti tovrstnega programa	f	f (%)
Zelo učinkovit	8	3,67 %
Pretežno učinkovit	97	44,50 %
Pretežno neučinkovit	88	40,37 %
Zelo neučinkovit	25	11,47 %
Skupaj	218	100,00 %

*Vir: lastno delo.*

*Tabela 8: Mnenje o etični sprejemljivosti tovrstnega programa*

Mnenje o etični sprejemljivosti tovrstnega programa	f	f (%)
Da	64	29,36 %
Ne	153	70,18 %
Skupaj	218	100,00 %

*Vir: lastno delo.*

*Tabela 9: Etična sprejemljivost/nesprejemljivost*

Etična sprejemljivost/nesprejemljivost	f	f (%)
Prihrani podjetjem čas pri pregledovanju velikega števila prošenj	72	15,13 %
Podjetja lahko zaposlujejo na kakršenkoli način želijo	53	11,13 %
Podjetja lahko zaposlijo kogarkoli želijo po svojih kriterijih	73	15,34 %
Gre samo za dodaten korak v procesu intervjuja	46	9,66 %
Bolj objektivni način	31	6,51 %
Algoritmi so v redu dokler ne posegajo v celoten proces izbire	120	25,21 %
Pristranskost do nekaterih prosilcev (na primer žensk)	74	15,55 %
Drugo	7	1,47 %
Skupno	476	100,00 %

*Vir: lastno delo.*

*Tabela 10: Mnenje o etični pravičnosti tovrstnega programa*

Mnenje o etični pravičnosti tovrstnega programa	f	f (%)
Zelo pravičen	12	5,61 %
Pretežno pravičen	66	30,84 %
Pretežno nepravičen	102	47,66 %
Zelo nepravičen	34	15,89 %
Skupaj	214	100,00 %

*Vir: lastno delo.*

*Tabela 11: Mnenje o učinkovitosti tovrstnega programa*

Mnenje o učinkovitosti tovrstnega programa	f	f (%)
Zelo učinkovit	9	4,23 %
Pretežno učinkovit	81	38,03 %
Pretežno neučinkovit	99	46,48 %
Zelo neučinkovit	24	11,27 %
Skupaj	213	100,00 %

*Vir: lastno delo.*

*Tabela 12: Mnenje o etični sprejemljivosti tovrstnega programa*

Mnenje o etični sprejemljivosti tovrstnega programa	f	f (%)
Da	76	35,68 %
Ne	137	64,32 %
Skupaj	213	100,00 %

*Vir: lastno delo.*

*Tabela 13: Etična sprejemljivost/nesprejemljivost*

Etična sprejemljivost/nesprejemljivost	f	f (%)
Smiselno je uporabiti kot dodatno orodje in ne kot odločilni faktor	141	28,48 %
Izboljšanje trenutnega sistema, ki ni najboljši	33	6,67 %
Ljudje si zaslužijo drugo priložnost	49	9,90 %
Vsako orodje, ki bi prepoznalo možnost ponovnega kaznivega dejanja je dobro za varnost ljudi	47	9,49 %
Ljudje so se morda spremenili in algoritmi nimajo trenutnih informacij	76	15,35 %
Vsak človek in situacija je specifična	126	25,45 %
Pristransko	22	4,44 %
Drugo	1	0,20 %
Skupaj	495	100,00 %

*Vir: lastno delo.*

*Tabela 14: Mnenje o etični pravičnosti tovrstnega programa*

Mnenje o etični pravičnosti tovrstnega programa	f	f (%)
Zelo pravičen	14	6,73 %
Pretežno pravičen	89	42,79 %
Pretežno nepravičen	87	41,83 %
Zelo nepravičen	18	8,65 %
Skupaj	208	100,00 %

*Vir: lastno delo.*

*Tabela 15: Mnenje o učinkovitosti tovrstnega programa*

Mnenje o učinkovitosti tovrstnega programa	f	f (%)
Zelo učinkovit	15	7,21 %
Pretežno učinkovit	118	56,73 %
Pretežno neučinkovit	64	30,77 %
Zelo neučinkovit	11	5,29 %
Skupaj	208	100,00 %

*Vir: lastno delo.*

*Tabela 16: Mnenje o etični sprejemljivosti tovrstnega programa*

Mnenje o etični sprejemljivosti tovrstnega programa	f	f (%)
Da	107	51,69 %
Ne	100	48,31 %
Skupaj	207	100,00 %

*Vir: lastno delo.*

*Tabela 17: Etična sprejemljivost/nesprejemljivost*

Etična sprejemljivost/nesprejemljivost	f	f (%)
Učinkovit način za podjetja, ki iščejo ustrezne stranke	69	14,84 %
Podjetja se lahko odločijo za načine pridobitve strank, ki jih sama želijo	72	15,48 %
Ljudje, ki so dejavni na internetu, se morajo sami zavedati, da puščajo za seboj podatke	89	19,14 %
Kršenje zasebnosti	72	15,48 %
Družbena omrežja in spletni podatki ne morejo podati prave slike o ljudeh	87	18,71 %
Diskriminatorno ravnanje	40	8,60 %
Ni možnosti vpliva na podatke oz. jih popraviti	35	7,53 %
Drugo:	1	0,22 %
Skupaj	465	100,00 %

*Vir: lastno delo.*

*Tabela 18: Mnenje o učinkovitosti tovrstne tehnologije*

Mnenje o učinkovitosti tovrstne tehnologije	f	f (%)
Zelo učinkovita	67	32,68 %
Pretežno učinkovita	118	57,56 %
Pretežno učinkovita	16	7,80 %
Zelo učinkovita	4	1,95 %
Skupaj	205	100,00 %

*Vir: lastno delo.*

*Tabela 19: Mnenje o etični sprejemljivosti tovrstne tehnologije*

Mnenje o etični sprejemljivosti tovrstne tehnologije	f	f (%)
Da	184	89,76 %
Ne	21	10,24 %
Skupaj	205	100,00 %

*Vir: lastno delo.*

*Tabela 20: Etična sprejemljivost/nesprejemljivost*

Etična sprejemljivost/nesprejemljivost	f	f (%)
Odkrivanje bolezenskih stanj	153	26,38 %
Večja natančnost in obvladovanje napak	136	23,45 %
Hitrejša diagnoza	169	29,14 %
Kršenje varovanja pravic pacientov	12	2,07 %
Zbiranje podatkov iz različnih virov	57	9,83 %
Prevzemanje odgovornosti	12	2,07 %
Izvzetje človeške presoje	40	6,90 %
Drugo:	1	0,17 %
Skupaj	580	100,00 %

*Vir: lastno delo.*

*Tabela 21: Mnenje o učinkovitosti tovrstne tehnologije*

Mnenje o učinkovitosti tovrstne tehnologije	f	f (%)
Zelo učinkovita	24	11,71 %
Pretežno učinkovita	106	51,71 %
Pretežno učinkovita	62	30,24 %
Zelo neučinkovita	13	6,34 %
Skupaj	205	100,00 %

*Vir: lastno delo.*

*Tabela 22: Mnenje o etični sprejemljivosti tovrstne tehnologije*

Mnenje o etični sprejemljivosti tovrstne tehnologije	f	f (%)
Da	103	50,49 %
Ne	101	49,51 %
Skupaj	204	100,00 %

*Vir: lastno delo.*

Tabela 23: Etična sprejemljivost/nesprejemljivost

Etična sprejemljivost/nesprejemljivost	f	f (%)
Večja varnost na cesti	82	16,60 %
Možnost človeške napake je večja kot tehnološke	76	15,38 %
Pristranskost pri odločanju vozila o tem komu bi škodilo	71	14,37 %
Vprašanje odgovornosti	84	17,00 %
Dodana vrednost	25	5,06 %
Potrebna je človeška vpetost	92	18,62 %
Večja dostopnost (na primer za invalide in starejše, ki ne morejo sami upravljati vozila)	58	11,74 %
Drugo	6	1,21 %
Skupaj	494	100,00 %

Vir: lastno delo.

Tabela 24: Zavedanje o zbiranju in obdelavi biometričnih podatkov

Zavedanje o zbiranju in obdelavi biometričnih podatkov		Da	Ne	Skupaj
S pomočjo biometričnih fotografij v potnem listu	f	143	60	203
	f (%)	70,44 %	29,56 %	100,00 %
Prstnih odtisov	f	155	48	203
	f (%)	76,35 %	23,65 %	100,00 %
Uporabo kamer s tehnologijo prepoznavanja obrazov	f	124	80	204
	f (%)	60,78 %	39,22 %	100,00 %

Vir: lastno delo.

Tabela 25: Mnenje o etični sprejemljivosti uporabe biometričnih podatkov

Mnenje o etični sprejemljivosti uporabe biometričnih podatkov	f	f (%)
Da	117	57,92 %
Ne	85	42,08 %
Skupaj	202	100,00 %

Vir: lastno delo.

Tabela 31: Etična sprejemljivost/nesprejemljivost

Etična sprejemljivost/nesprejemljivost	f	f (%)
Z vidika splošne varnosti	106	16,23 %
Zaradi prepoznavanja osumljencev kaznivih dejanj v množici ljudi, ter s tem preprečevanja kriminala in terorizma	102	15,62 % se nadaljuje

Tabela 31: Etična sprejemljivost/nesprejemljivost (nad.)

Etična sprejemljivost/nesprejemljivost	f	f (%)
Hitrejšega postopka primerjave posnetka kamere z Evidenco fotografiranih oseb na katerem je na primer osumljenec ropa	79	12,10 %
Kot pomoč pri iskanju pogošenih oseb	93	14,24 %
Hitrejši postopek izkazovanja identitete na letališčih	66	10,11 %
Kršenje varstva o osebnih podatkih	73	11,18 %
Kršenje človekovih pravic	69	10,57 %
Omejevanje svobode gibanja	60	9,19 %
Drugo	5	0,77 %
Skupaj	653	100,00 %

Vir: lastno delo.

Tabela 26: Mnenje o podanih vprašanjih

Mnenje o podanih vprašanjih		Da	Ne	Skupaj
Ali bi morala biti večja preglednost podatkov za zaupanje vredno okolje pri odločanju algoritmov?	f	182	17	199
	f (%)	91,46 %	8,54 %	100,00 %
Ali bi si morali prizadevati za rast algoritmične izobraženosti zato, da bi ljudje razumeli delovanje sprejemanja avtomatiziranih odločitev?	f	182	17	199
	f (%)	91,46 %	8,54 %	100,00 %
Ali bi moral biti večji nadzor nad uporabo algoritmov?	f	188	13	201
	f (%)	93,53 %	6,47 %	100,00 %
Ali bi vas morala država ščititi pred tovrstnimi zlorabami?	f	186	15	201
	f (%)	92,54 %	7,46 %	100,00 %
Ali se vam zdi smiselno razviti ureditev etike podatkov?	f	189	12	201
	f (%)	94,03 %	5,97 %	100,00 %
Ali bodo po vašem mnenju pozitivni vplivi algoritmov odtehtali negativne?	f	123	76	199
	f (%)	61,81 %	38,19 %	100,00 %
Ali zaupate / bi zaupali tehnologiji algoritmov, da podaja kakovostne in nepristranske podatke?	f	97	102	199
	f (%)	48,74 %	51,26 %	100,00 %

Vir: lastno delo.

Tabela 33: Mnenje o odgovornosti za neetično odločitev

Mnenje o odgovornosti za neetično odločitev	f	f (%)
Stroj	8	4,00 % se nadaljuje

Tabela 33: Mnenje o odgovornosti za neetično odločitev (nad.)

Mnenje o odgovornosti za neetično odločitev	f	f (%)
Razvijalec	73	36,50 %
Podjetje, ki uporablja omenjeno tehnologijo	119	59,50 %
Skupaj	200	100,00 %

Vir: lastno delo.

Tabela 27: Pomembnost vloge spodnjih vprašanj pri sprejemanju odločitev

Pomembnost vloge spodnjih vprašanj pri sprejemanju odločitev		Da	Ne	Skupaj
Ali je odločitev oz. dejanje zakonito?	f	176	24	200
	f (%)	88,00 %	12,00 %	100,00 %
Ali je odločitev oz. dejanje uravnoteženo (med obema stranema)?	f	158	42	200
	f (%)	79,00 %	21,00 %	100,00 %
Kakšno mnenje bom imel o sebi?	f	147	53	200
	f (%)	73,50 %	26,50 %	100,00 %

Vir: lastno delo.

Tabela 28: Srečevanje z uporabo algoritmov za odločanje pri delu

Srečevanje z uporabo algoritmov za odločanje pri delu	f	f (%)
Da	64	32,00 %
Ne	127	63,50 %
Nameravam(o) v prihodnosti	7	3,50 %
Drugo	2	1,00 %
Skupaj	200	100,00 %

Vir: lastno delo.

Tabela 29: Področja uporabe algoritmov za odločanje

Področja uporabe algoritmov za odločanje	f	f (%)
Hitrejši postopek izkazovanja identitete	35	24,14 %
Varnostna zaščita potrjevanja transakcij	42	28,97 %
Ocena osebnih financ	12	8,28 %
Splošna varnost	26	17,93 %
Zdravstvo	23	15,86 %
Samodejni pregled življenjepisov za delovno mesto	5	3,45 %
Drugo	2	1,38 %
Skupaj	145	100,00 %

Vir: lastno delo.

*Tabela 30: Uporaba tehnologije v primeru ustreznega kadra oziroma sredstev*

Uporaba tehnologije v primeru ustreznega kadra oziroma sredstev	f	f (%)
Da	79	59,40 %
Ne	54	40,60 %
Skupaj	133	100,00 %

*Vir: lastno delo.*

*Tabela 31: Izzivi pri vzpostavitvi omenjene tehnologije*

Izzivi pri vzpostavitvi omenjene tehnologije	f	f (%)
Varovanje zasebnosti podatkov	99	36,80 %
Zagotovitev finančnih sredstev	47	17,47 %
Zagotovitev upravljanja s podatki	50	18,59 %
Zagotovitev kakovostnih podatkov	67	24,91 %
Drugo	6	2,23 %
Skupaj	269	100,00 %

*Vir: lastno delo.*



## Priloga 2: Anketni vprašalnik

Sem Tina Koželj, študentka 2. letnika magistrskega študija na Ekonomski fakulteti v Ljubljani, smer Poslovna logistika in pripravljam magistrsko delo z naslovom Etični vidik odločanja s pomočjo algoritmov. Z Vašo pomočjo bi želela ugotoviti, kaj velja za etično sprejemljivo odločanje s pomočjo algoritmov. V ta namen Vas lepo prosim, da mi pomagate in izpolnite spodnji vprašalnik, ki je anonimen, vaši odgovori pa zaupni in prostovoljni. Vse podatke, ki jih bom pridobila z izpolnjenim vprašalnikom, bom uporabila izključno za namen magistrskega dela. Za Vašo pomoč se Vam že vnaprej najlepše zahvaljujem.

Q1 – 1. Ali poznate izraz umetna inteligenca?

Da.

Ne.

Q2 – 2. Ali ste vedeli, da nezavedno uporabljate umetno inteligenco v vsakdanjem življenju (na primer pri uporabi Google iskalnika, Facebooka itd.) s tem, ko vam selektivno podajajo oglase za izdelke, storitve, predloge oziroma druge relevantne rezultate?

Da.

Ne.

Q3 – 3. Algoritem je navodilo, s katerim se rešuje nek problem. Običajno je zapisan kot seznam korakov, ki pripeljejo do rešitve problema. Ali ste kdaj razmišljali, da se algoritmi umetne inteligence odločajo namesto vas?

Da.

Ne.

Q4 – 4. Katera od spodnjih trditev je najbližje vašemu pogledu, četudi nobena ni natančno pravilna?

Možno je, da računalniški programi sprejemajo odločitve brez človeške pristranskosti.

Računalniški programi bodo vedno odražali pristranskost ljudi, ki so jih zasnovali.

Q5 – Sledijo sklopi vprašanj, kjer so predstavljeni različni primeri uporabe algoritmov za zbiranje in analizo velike količine podatkov, tako javnih kot zasebnih, v resničnem svetu.

Q6 – 5. Nekatera podjetja za izboljšanje zaposlitvenega procesa uporabljajo postopek, s katerim računalnik vsakemu kandidatu dodeli samodejno oceno, glede na vsebino njegovega življenjepisa in primerjave z življenjepisom uspešnih zaposlenih. Vodja kadrovskega oddelka tako dobi samo tiste življenjepise, ki izpolnjujejo določeno oceno.

Q7 – 5.1. Ali ste vedeli, da se algoritmi umetne inteligence uporabljajo pri zgoraj opisanem primeru?

Da.

Ne.

Q8 – 5.2. Kako etično pravičen je po vašem mnenju tovrstni program do prosilcev za delovno mesto?

Zelo pravičen.

Pretežno pravičen

Pretežno nepravičen.

Zelo nepravičen.

Q9 – 5.3. Kako učinkovit je po vašem mnenju tovrstni program za prepoznavanje prosilcev, ki bi bili dobri kandidati za delo?

Zelo učinkovit.

Pretežno učinkovit.

Pretežno neučinkovit.

Zelo neučinkovit.

Q10 – 5.4. Ali menite, da je etično sprejemljivo, da podjetja uporabljajo tovrstni program pri zaposlovanju ali ne?

Da.

Ne.

Q11 – 5.5. Zakaj menite, da je etično sprejemljivo / nesprejemljivo? (Možnih je več odgovorov)

Prihrani podjetjem čas pri pregledovanju velikega števila prošenj.

Podjetja lahko zaposlujejo na kakršen koli način želijo.

Podjetja lahko zaposlijo kogar koli želijo po svojih kriterijih.

Gre samo za dodaten korak v procesu intervjuja.

- Bolj objektiven način.
- Algoritmi so v redu, dokler ne posegajo v celoten proces izbire.
- Pristranskost do nekaterih prosilcev (na primer žensk).
- Drugo:

Q12 – 6. V nekaterih primerih se uporabljajo avtomatizirani programi za izračun kazenskega tveganja za ljudi v zaporu na podlagi zbiranja informacij iz mnogih virov o osebnostnih lastnostih in preteklem vedenju osebe. Podatki se primerjajo z drugimi obsojenimi za kazniva dejanja in tako se dodeli ocena vsakemu posamezniku, ki pomaga pri odločitvi, ali je smiselno, da bi nekoga pogojno odpustili ali ne.

Q13 – 6.1. Ali ste vedeli, da se algoritmi umetne inteligence uporabljajo pri zgoraj opisanem primeru?

- Da.
- Ne.

Q14 – 6.2. Kako etično pravičen je po vašem mnenju tovrstni program do ljudi pri pogojnih zaslišanjih?

- Zelo pravičen.
- Pretežno pravičen.
- Pretežno nepravičen.
- Zelo nepravičen.

Q15 – 6.3. Kako učinkovit je po vašem mnenju tovrstni program za prepoznavanje ljudi, ki si zaslužijo pogojnega odpusta?

- Zelo učinkovit.
- Pretežno učinkovit.
- Pretežno neučinkovit.
- Zelo neučinkovit.

Q16 – 6.4. Ali menite, da je etično sprejemljivo, da pravni sistemi uporabljajo tovrstni program?

Sprejemljivo.

Nesprejemljivo.

Q17 – 6.5. Zakaj menite, da je etično sprejemljivo / nesprejemljivo? (Možnih je več odgovorov)

Smiselno je uporabiti kot dodatno orodje in ne kot odločilni faktor.

Izboljšanje trenutnega sistema, ki ni najboljši.

Ljudje si zaslužijo drugo priložnost.

Vsako orodje, ki bi prepoznalo možnost ponovnega kaznivega dejanja, je dobro za varnost ljudi.

Ljudje so se morda spremenili in algoritmi nimajo trenutnih informacij.

Vsak človek in situacija sta specifična.

Pristransko.

Drugo.

Q18 – 7. Številna podjetja nudijo avtomatizirane programe za izračun ocene osebnih financ na podlagi zbranih finančnih podatkov (na primer prihodki, bančna in davčna zgodovina) kot tudi osebnih in vedenjskih podatkov (na primer geolokacijski, dejavnost na družbenih omrežjih, elektronska pošta, zgodovina brskanja po internetu). Ta je v pomoč podjetjem pri lažji odločitvi, ali naj določeni osebi ponudi posojila, posebne ponudbe in druge storitve.

Q19 – 7.1. Ali ste vedeli, da se algoritmi umetne inteligence uporabljajo pri zgoraj opisanem primeru?

Da.

Ne.

Q20 – 7.2. Kako etično pravičen je po vašem mnenju tovrstni program do strank?

Zelo pravičen.

Pretežno pravičen.

Pretežno nepravičen.

Zelo nepravičen.

Q21 – 7.3. Kako učinkovit je po vašem mnenju tovrstni program za prepoznavanje ljudi, ki bi bili dobre stranke?

- Zelo učinkovit.
- Pretežno učinkovit.
- Pretežno neučinkovit.
- Zelo neučinkovit.

Q22 – 7.4. Ali menite, da je etično sprejemljivo, da podjetja uporabljajo tovrstni program?

- Sprejemljivo.
- Nesprejemljivo.

Q23 – 7.5. Zakaj menite, da je etično sprejemljivo / nesprejemljivo? (Možnih je več odgovorov.)

- Učinkovit način za podjetja, ki iščejo ustrezne stranke.
- Podjetja se lahko odločijo za načine pridobitve strank, ki jih sama želijo.
- Ljudje, ki so dejavni na internetu, se morajo sami zavedati, da puščajo za seboj podatke.
- Kršenje zasebnosti.
- Družbena omrežja in spletni podatki ne morejo podati prave slike o ljudeh.
- Diskriminatorno ravnanje.
- Ni možnosti vpliva na podatke oz. jih popraviti.
- Drugo.

Q24 – 8. Algoritmi umetne inteligence so prisotni v medicini za hitrejše odkrivanje bolezenskih stanj z visoko stopnjo natančnosti (na primer diagnoza možganskega tumorja) in lažjo obdelavo podatkov (na primer radiolog pred posegom in obsevanjem rakavih celic ročno pregleda ogromno količino medicinskih slik in označuje organe ter vrsto tkiva, medtem ko algoritmi delo, ki bi trajalo več ur, opravijo v nekaj minutah kot tudi z večjo natančnostjo).

Q25 – 8.1. Ali ste vedeli, da se algoritmi umetne inteligence uporabljajo pri zgoraj opisanem primeru?

Da.

Ne.

Q26–8.2. Kako učinkovita je po vašem mnenju tovrstna tehnologija za medicino?

Zelo učinkovita.

Pretežno učinkovita.

Pretežno neučinkovita.

Zelo neučinkovita.

Q27–8.3. Ali menite, da je etično sprejemljivo, da medicina uporablja tovrstno tehnologijo?

Sprejemljivo.

Nesprejemljivo.

Q28–8.4. Zakaj menite, da je etično sprejemljivo / nesprejemljivo? (Možnih je več odgovorov.)

Odkrivanje bolezenskih stanj.

Večja natančnost in obvladovanje napak.

Hitrejša diagnoza.

Kršenje varovanja pravic pacientov.

Zbiranje podatkov iz različnih virov.

Prevzemanje odgovornosti.

Izvzetje človeške presoje.

Drugo.

Q29 – 9. V prihodnosti bo popolna avtonomnost vozil s pomočjo algoritmov in umetne inteligence, ki bo povečala varnost na cestah, saj bo odpravila človeške napake, ki so vzrok za 95 % nesreč. Vendar zamislite si situacijo, ko bi zaupali vožnjo samovozečemu vozilu in bi nenadoma pritekel otrok na cesto. Zaradi hitrosti in razdalje bi bila onemogočena varna zaustavitev. Takrat se poraja vprašanje, ali bo vozilo ohranilo smer in trčilo ali zapeljalo s ceste in ogrozilo lastne potnike.

Q30 – 9.1. Ali ste vedeli, da se algoritmi umetne inteligence uporabljajo pri zgoraj opisanem primeru?

Da.

Ne.

Q31 – 9.2. Kako učinkovita je po vašem mnenju tovrstna tehnologija za večjo varnost pri vožnji?

Zelo učinkovita.

Pretežno učinkovita.

Pretežno neučinkovita.

Zelo neučinkovita.

Q32–9.3. Ali menite, da je etično sprejemljivo, da algoritmi in umetna inteligenca odločata o tem, kako bo vozilo reagiralo v določenih situacijah?

Sprejemljivo.

Nesprejemljivo.

Q33–9.4. Zakaj menite, da je etično sprejemljivo / nesprejemljivo? (Možnih je več odgovorov.)

Večja varnost na cesti.

Možnost človeške napake je večja kot tehnološke.

Pristranskost pri odločanju vozila o tem, komu bi škodilo.

Vprašanje odgovornosti.

Dodana vrednost.

Potrebna je človeška vpetost.

Večja dostopnost (na primer za invalide in starejše, ki ne morejo sami upravljati vozila).

Drugo:

Q34–10. Ali ste vedeli, da država zbira in obdeluje vaše biometrične podatke?

Da

Ne

S pomočjo biometričnih fotografij v potnem listu.

Prstnih odtisov.

Uporabo kamer s tehnologijo prepoznavanja obrazov.

Q35–10.1. Ali menite, da je etično sprejemljivo, da država uporablja biometrične podatke?

Sprejemljivo.

Nesprejemljivo.

Q36 – 10.2. Zakaj menite, da je etično sprejemljivo / nesprejemljivo? (Možnih je več odgovorov.)

Z vidika splošne varnosti.

Zaradi prepoznavanja osumljencev kaznivih dejanj v množici ljudi ter s tem preprečevanja kriminala in terorizma.

Hitrejšega postopka primerjave posnetka kamere z Evidenco fotografiranih oseb na katerem je na primer osumljenec ropa.

Kot pomoč pri iskanju pogrešanih oseb.

Hitrejši postopek izkazovanja identitete na letališčih.

Kršenje varstva o osebnih podatkih.

Kršenje človekovih pravic.

Omejevanje svobode gibanja.

Drugo.

Q37 – 11. Odgovorite z "da" ali "ne" na naslednja vprašanja glede na vaše mnenje.

Da

Ne

Ali bi morala biti večja preglednost podatkov za zaupanje vredno okolje pri odločanju algoritmov?



Ali bi si morali prizadevati za rast  
algoritmične izobraženosti zato, da  
bi ljudje razumeli delovanje  
sprejemanja avtomatiziranih  
odločitev?

Ali bi moral biti večji nadzor nad  
uporabo algoritmov?

Ali bi vas morala država ščititi pred  
tovrstnimi zlorabami?

Ali se vam zdi smiselno razviti  
ureditev etike podatkov?

Ali bodo po vašem mnenju  
pozitivni vplivi algoritmov  
odtehtali negativne?

Ali zaupate / bi zaupali tehnologiji  
algoritmov, da podaja kakovostne  
in nepristranske podatke?

Q38–13. Če stroj odloča z nenamernimi škodljivimi posledicami, kdo je po vašem mnenju  
odgovoren za neetično odločitev?

Stroj.

Razvijalec.

Podjetje, ki uporablja omenjeno tehnologijo.

Q39–14. Označite ali vam pri sprejemanju odločitev naslednja tri vprašanja predstavljajo  
pomembno vlogo?

Da

Ne

Ali je odločitev oz. dejanje  
zakonito?

Ali je odločitev oz. dejanje  
uravnoteženo (med obema  
stranema)?

Kakšno mnenje bom imel o sebi?



Q40 – 15. Ali se pri svojem delu srečujete z uporabo algoritmov za odločanje, kot so na primer hitrejši postopek izkazovanja identitete, varnostna zaščita potrjevanja transakcij, zdravstveni nameni, splošna varnost itd.?

Da.

Ne.

Nameravam(o) v prihodnosti.

Drugo:

IF (10) Q40 = [1]

Q41 – 15.1. Na katera področja ste želeli vplivati z uporabo algoritmov za odločanje? (Možnih je več odgovorov.)

Hitrejši postopek izkazovanja identitete.

Varnostna zaščita potrjevanja transakcij.

Ocena osebnih financ.

Splošna varnost.

Zdravstvo.

Samodejni pregled življenjepisov za delovno mesto.

Drugo.

IF (12) Q40 = [2, 3]

Q42 – 15.1. Ali bi uporabljali omenjeno tehnologijo, če bi imeli za to ustrezen kader oziroma sredstva?

Da.

Ne.

IF (13) Q40 = [2, 3]

Q43 – 15.2. S katerimi izzivi se soočate pri vzpostavitvi omenjene tehnologije? (Možnih je več odgovorov)

- Varovanje zasebnosti podatkov.
- Zagotovitev finančnih sredstev.
- Zagotovitev upravljanja s podatki.
- Zagotovitev kakovostnih podatkov.
- Drugo:

XSPOL – Spol:

- Moški.
- Ženski.

XSTAR2a4–V katero starostno skupino spadate?

- 18–24 let.
- 25–34 let.
- 35–44 let.
- 45–54 let.
- Nad 55 let.

XIZ1a2 – Kakšna je vaša najvišja dosežena formalna izobrazba?

- I. stopnja izobrazbe – nedokončana OŠ.
- II. stopnja izobrazbe – OŠ.
- III. stopnja izobrazbe – nižje poklicno izobraževanje (2 leti).
- IV. stopnja izobrazbe – srednje poklicno izobraževanje (3-letno).
- V. stopnja izobrazbe – gimnazijsko, srednje poklicno-tehniško izobraževanje, srednje tehniško oz. drugo strokovno izobraževanje.
- VI. stopnja izobrazbe – višješolski strokovni program, specializacija po višješolskem programu, visokošolski strokovni program oz. po bolonjskih programih visokošolski strokovni in univerzitetni program (1. bol. st.).
- VII. stopnja izobrazbe – specializacija po visokošolskem strokovnem programu, univerzitetni program oz. po bolonjskih programih magisterij stroke (2. bol.st.)

○ VIII. stopnja izobrazbe – specializacija po univerzitetnem programu, magisterij znanosti, doktorat znanosti (3. bol. st.)

Q44 – Smer poklica (Možnih je več odgovorov)

- Splošne izobraževalne aktivnosti/izidi.
- Izobraževalne vede in izobraževanje učiteljev.
- Umetnost in humanistika.
- Družbene, poslovne, upravne in pravne vede.
- Naravoslovje, matematika in računalništvo.
  - Tehnika, proizvodne tehnologije in gradbeništvo.
- Kmetijstvo, gozdarstvo, ribištvo, veterinarstvo.
- Zdravstvo in sociala.
- Storitve (osebne, transportne, varstvo okolja, varovanje).
- Neopredeljeno.

Q45 – Sektor dela (Možnih je več odgovorov.)

- Javni sektor.
- Zasebni sektor.
- Drugo: