

KLAIPĖDOS UNIVERSITETAS

Socialinių ir humanitarinių mokslų fakultetas
Filosofijos, menotyros ir komunikacijos katedra

DIRBTINIS INTELEKTAS IR ETIKA: IŠŠŪKIAI IR PERSPEKTYVOS

Profesinės etikos ir etikos audito programos magistro baigiamasis darbas

Autorius

SMSPE19 gr. stud. Gabrielė Juodeikytė

Vadovas

doc. dr. Ernesta Molotkienė

Klaipėda, 2021

MAGISTRO BAIGIAMOJO DARBO LYDRAŠTIS

Pildo magistro baigiamojo darbo autorius

.....Gabrielė Juodeikytė

(magistro baigiamojo darbo autoriaus vardas, pavardė)

..... Dirbtinis intelektas ir etika: iššūkiai ir perspektyvos.....

(magistro baigiamojo darbo pavadinimas lietuvių kalba)

Patvirtinu, kad magistro baigiamasis darbas parašytas savarankiškai, nepažeidžiant kitiems asmenims priklausančių autorių teisių, visas baigiamasis magistro darbas ar jo dalis nebuvo panaudotas Klaipėdos universitete ir kitose aukštosiose mokyklose.

.....Gabrielė Juodeikytė.....

(magistro baigiamojo darbo autoriaus ir parašas)

Sutinku, kad magistro baigiamasis darbas būtų naudojamas neatlygintinai 5 m.

Klaipėdos universiteto studijų procese.

.....Gabrielė Juodeikytė.....

(magistro baigiamojo darbo autoriaus ir parašas)

Pildo magistro baigiamojo darbo vadovas

Magistro baigiamąjį darbą ginti

(įrašyti – leidžiu arba neleidžiu)

.....doc. dr. Ernesta Molotokienė.....

(data) (magistro baigiamojo darbo vadovo vardas, pavardė ir parašas)

Pildo katedros, kuriojančios studijų programą, administratorius (sekretorius)

Baigiamasis darbas įregistruotas **Filosofijos, menotyros ir komunikacijos katedroje**

.....

(data)

(katedros sekretorės vardas, pavardė ir parašas)

Pildo katedros, kuriojančios studijų programą, vedėjas

Magistro baigiamąjį darbą ginti

(įrašyti – leidžiu arba neleidžiu)

.....doc. dr. Ernesta Molotokienė.....

(data) (katedros vedėjo vardas, pavardė ir parašas)

Recenzentu(-ais) skiriu

.....

(įrašyti recenzento(ų) vardą, pavardę)

.....doc. dr. Ernesta Molotokienė.....

(data) (katedros vedėjo vardas, pavardė ir parašas)

SANTRAUKA

Gabrielė Juodeikytė. Dirbtinis intelektas ir etika: iššūkiai ir perspektyvos. Magistro baigiamasis darbas. Darbo vadovas: doc. dr. Ernesta Molotkienė. Klaipėdos Universitetas, Socialinių ir humanitarinių mokslų fakultetas, Filosofijos, menotyros ir komunikacijos katedra: Klaipėda, 2021. – 72 p.

Temos aktualumas. Šiuolaikinis pasaulis jau garsiai konstatuoja faktą, kad dirbtinis intelektas sparčiai transformuoja verslą ir ekonominius procesus. Dirbtinio intelekto valdomos mašinos, jau pakeitė didelę dalį pramonės bei verslo. Itin greitai kinta ir pats technologijų vaidmuo visuomenėje, naujosios technologijos nebėra tik įrankis gerinantis žmonių kasdienybę. Tai tapo neišvengiamu ir būtinu poreikiu. Vykstantys pokyčiai neišvengiamai keičia ne tik darbų prigimtį, bet ir pačią darbo vietos koncepciją. Dirbtinio intelekto technologijos žmonijai gali suteikti milžiniškas galimybes suklestėti. Arba galimybę susinaikinti. Tinkamai ir etiškai valdyti mokslinių tyrimų ir inovacijų veiksmus yra sunku, nes lieka neaiškių ir neapibrėžtų aplinkybių. Toks greitas technologijų vystymasis ir tobulėjanti technikos evoliucija kelia nemažai etikos klausimų, todėl būtina užtikrinti, kad dirbtinio intelekto technologijos būtų sąžiningos ir etiškos.

Darbo tikslas – išanalizuoti dirbtinio intelekto sampratą, teisinį reglamentavimą Europos Sąjungoje ir Lietuvoje, esminius dirbtinio intelekto etikos principus, iššūkius ir raidos perspektyvas.

Darbą sudaro įvadas, kuriame pristatomas temos aktualumas, problema, tikslas, uždaviniai, tyrimo metodai bei darbo struktūra. Dėstymo dalis sudaryta iš trijų skyrių. Pirmajame skyriuje pristatoma dirbtinio intelekto koncepcijos samprata, aptariama dirbtinio intelekto sąvoka, technologijų raida bei dirbtinio intelekto pritaikomumas ir praktinė nauda visuomenėje. Antrajame skyriuje pristatomas teisinis dirbtinio intelekto technologijų reguliavimas Europos Sąjungoje ir Lietuvoje, analizuojama Europos Sąjungos dirbtinio intelekto strategija, taip pat Lietuvos dirbtinio intelekto strategijos teisinio ir etinio reguliavimo problema. Trečiame skyriuje pristatomos dirbtinio intelekto etikos problemos, iššūkiai ir raidos perspektyvos visuomenėje. Taip pat išvados ir literatūros sąrašas.

Raktiniai žodžiai: dirbtinis intelektas, mašininis mokymasis, robotas, teisė, etika, iššūkiai, perspektyvos

SUMMARY

Gabriele Juodeikyte. Artificial Intelligence and Ethics: Challenges and Perspectives. Master's thesis. Study Counselor Associate Prof. Dr. Ernesta Molotokiene. Klaipeda University, Faculty of Social Sciences and Humanities, Department of Philosophy, Arts and Communication: Klaipeda, 2021. – 72 p.

Relevance of the topic. The modern world is already loudly acknowledging the fact that artificial intelligence is rapidly transforming business and economic processes. Artificial intelligence controlled machines have already replaced a large part of industry and business. The role of technology in society is changing very fast, and new technologies are no longer just a tool for improving people's daily lives. This has become an unavoidable and necessary. The changes that are taking place will inevitably change not only the nature of the work, but also the very concept of the workplace. Artificial intelligence technologies can provide enormous opportunities for humanity to flourish. Or the possibility of destruction. Proper and ethical management of research and innovation drivers is difficult as circumstances remain uncertain. Such rapid technological development and the evolving evolution of technology raises a number of ethical questions, and it is therefore essential to ensure that artificial intelligence technologies are fair and ethical.

The aim of the work is to analyze the concept of artificial intelligence, legal regulation in the European Union and Lithuania, the essential ethical principles, challenges and development perspectives of artificial intelligence.

The work consists of an introduction, which presents the relevance of the topic, problem, goal, tasks, research methods and work structure. The teaching part consists of three sections. The first chapter introduces the concept of artificial intelligence, discusses the concept of artificial intelligence, technological development and the applicability and practical benefits of artificial intelligence in society. The second chapter presents the legal regulation of artificial intelligence technologies in the European Union and Lithuania, analyzes the European Union artificial intelligence strategy, as well as the problem of legal and ethical regulation of the Lithuanian artificial intelligence strategy. The third chapter presents the ethical problems, challenges and development perspectives of artificial intelligence in society. Also conclusions and references.

Keywords: artificial intelligence, machine learning, robot, law, ethics, challenges, perspectives

TURINYS

SANTRAUKA	3
SUMMARY	4
TURINYS	5
PAVEIKSLŲ, LENTELIŲ SARAŠAS	6
SANTRUMPOS	7
ĮVADAS	8
I. DI KONCEPCIJOS SAMPRATA	11
1.1. DI sąvoka	11
1.2. DI technologijų raida	12
1.3. DI ribos ir galimybės	13
1.4. DI pritaikomumas ir praktinė reikšmė visuomenei	17
II. TEISINIS DI REGULIAVIMAS LIETUVOJE IR ES	21
2.1. DI technologijų teisinio reguliavimo problema	21
2.2. ES DI strategija: teisės ir etikos dermė	26
2.3. Lietuvos DI strategija: teisinio ir etinio reguliavimo įveikimo problema	33
III. DI ETIKA	39
3.1. Globalios DI etikos susiformavimo kontekstas	39
3.2. Pagrindinės DI etikos problemos	41
3.3. DI etikos iššūkiai ir raidos perspektyvos	46
IŠVADOS	60
LITERATŪRA	61

PAVEIKSLŲ, LENTELIŲ SĄRAŠAS

PAVEIKSLŲ SĄRAŠAS

Eil. Nr.	Paveikslo pavadinimas	Puslapis
1.	Lietuvos DI projekto eiga	34
2.	Teisiniai ir etiniai klausimai dėl DI naudojimo	36
3.	Pagrindinės etinės ir moralinės problemos, susijusios su DI kūrimu ir įgyvendinimu	42

LENTELIŲ SĄRAŠAS

Eil. Nr.	Lentelės pavadinimas	Puslapis
1.	Etinės iniciatyvos ir analizuojamos problemos	53

SANTRUMPOS

BDAR – Bendrasis duomenų apsaugos reglamentas

DI – dirbtinis intelektas

EESRK – Europos ekonomikos ir socialinių reikalų komitetas

EK – Europos Komisija

ES – Europos Sąjunga

EURON (*The European Robotics Research Network*) – Europos robotikos tyrimų tinklas

EUROP (*The European Robotics Platform*) – Europos robotikos platforma

FAO (*Food and Agriculture Organization of the United Nations*) – Jungtinių Tautų maisto ir žemės ūkio organizacija – tai specialusis Jungtinių Tautų Organizacijos padalinys, vadovaujantis tarptautinėms pastangoms bado problemai išspręsti

IBM (*International Business Machines Corporation*) – JAV kompiuterių gamintoja, programinės ir techninės įrangos prekybos bendrovė, teikianti infrastruktūros ir konsultavimo paslaugas

IEEE (*The Institute of Electrical and Electronics Engineers*) – Elektros ir elektronikos inžinierių institutas yra profesionali elektronikos inžinerijos ir elektrotechnikos asociacija

JSAI (*The Japanese Society for Artificial Intelligence*) – Japonijos dirbtinio intelekto organizacija

JT – Jungtinės Tautos

SIGAI (*Special Interest Group on Artificial Intelligence*) – specialiųjų interesų grupė dirbtinio intelekto klausimais

IVADAS

Temos aktualumas. Antrojo krikščionybės eros tūkstantmečio pabaigoje socialinį žmonijos gyvenimo „kraštovaizdį“ pakeitė istorinės reikšmės įvykiai. Technologinis perversmas, apimantis informacines technologijas, vis greičiau ir greičiau keičia žmonijos pagrindus (Amilevičius, 2017). Per pastarąjį dešimtmetį dėl itin greito technologijos vystymosi, robotikos srityje buvo padaryti reikšmingi pasikeitimai: sukurtas ir įdiegtas dirbtinis intelektas (toliau – DI) , todėl, kai kurią dalį arba net ir visus sprendimus dėl savo veikimo robotas priima pats; dėl kitų technologijų robotai tapo nepriklausomi nuo žmogaus. DI ir nuotolinės bevielės technologijos įgalino sukurti daugiau ar mažiau autonominius robotus (Štilis ir kt., 2016). Naujos ir besiformuojančios technologijos, tarp kurių svarbią vietą užima DI, jau pakeitė didelę dalį pramonės bei verslo. Itin greitai kinta ir pats technologijų vaidmuo visuomenėje, naujosios technologijos nebėra tik įrankis gerinantis žmonių kasdienybę. Tai tapo neišvengiamu ir būtinu poreikiu. Tinkamai ir etiškai valdyti mokslinių tyrimų ir inovacijų veiksmus yra sunku, nes lieka neaiškių ir neapibrėžtų aplinkybių, todėl sparčiai besivystanti ir tobulėjanti technikos evoliucija vis dar kelia nemažai etikos klausimų (Amilevičius, 2017).

Šiandien, robotų darbas ir veikimas – neatsiejama žmonių gyvenimo dalis, gyvenanti kartu žmonėmis „robotai integravosi į daugelį kasdienio gyvenimo sričių: nuo teksto vertimo programėlių, liftų iki automobilyje esančio išmanaus autopiloto“ (Štilis ir kt., 2016). DI yra įsisknijęs ir į ekonomikos sritį: finansų įstaigos, ieškodamos palankių investicijų, tyrinėdamos rinkos procesus ar norėdamos nustatyti pinigų plovimo atvejus naudoja DI. Sveikatos priežiūros srityje, DI sistemos atlieka įvairius diagnostikos tyrimus ar net slaugo pacientus. Pramonės sektoriuje DI ir robotai naudojami gamybai ir darbo našumui gerinti. Pasaulyje plačiai naudojama „Google“ paieškos sistema šias technologijas naudoja analizuodama vartotojų elgesį bei norus (Amilevičius, 2017).

DI gali turėti naudingų ir (arba) žalingų rezultatų, atsižvelgiant į dabartinį vystymosi greitį. Mokslininkai mano, kad per keletą dešimtmečių ši technologija gali tapti savarankiška ir itin intelektualiai (Amilevičius, 2017). Manoma jog DI tobulėjimo greitis stipriai pralenkia visuomenės bei pačio žmogaus raidos tempą. Anglų fizikas, Stivenas Hokingas (Stephen Hawking) viename interviu BBC sakė, kad „visuotinio DI sukūrimas gali reikšti žmonijos pabaigą“ (Cellan-Jones, 2014). Dėl šių priežasčių būtina nustatyti gaires dabartinio ir tolesnio DI naudojimui ir vystymui, stebėti, vertinti ir analizuoti DI etinį poveikį ir keliamas grėsmes.

Aktualumas užsienyje ir Lietuvoje. Nuo 2017 m. pasaulyje intensyviai svarstomi DI etikos klausimai, susiję su DI ir žmogaus koegzistencija. Todėl 2017 m. galima laikyti lūžio metais, kai DI tapo jau ne tik technologine, bet ir socialine realybe. Itin padidėjo diskusijų, mokslinių straipsnių apie

etinius, teisinius šių technologijų reguliavimo ir priežiūros aspektus. Lietuvoje ir pasaulyje kuriamos DI strategijos, nuostatai bei gairės. Skiriamas didelis mokslo ir tyrimų dėmesys, rengiami projektai.

Lietuvoje apie DI rašo Vytauto Didžiojo universiteto doc. dr. Amilevičius (2017), Lietuvos informatikas, fizinių mokslų daktaras Čaplinskas (2015), Mogirdas (2019) taip pat profesorius Telksnys (2019) bei Mykolo Romerio universiteto kolektyvas – Štītis, Kiškis, Limba, Rotomskis, Agafonov, Gulevičiūtė, Panka (2016). Labai svarbus darbas atliktas Ekonomikos ir inovacijų ministerijos kartu su projekto „Kurk Lietuvai“ dalyviais – „Lietuvos dirbtinio intelekto strategija“ (2019).

Užsienyje DI etiką analizuoja Vincent (2020) savo darbe „Dirbtinio intelekto etika ir robotai“. DI etikos pagrindų santrauka pateikta Europos mokslo ir naujų technologijų etikos grupės (2018). Apie bendrą DI politiką ir etiką rašė Calo (2018), taip pat Crawford ir Calo (2016), Stahl, Timmermans, Mittelstadt (2016), Johnson, Verdicchio (2017) bei Giubilini, Savulescu (2018). Apie etikos problemas knygoje „Išradimų etika“ rašė Jasanoff (2016).

Problema: Kokie yra esminiai DI etikos principai, iššūkiai ir raidos perspektyvos?

Tezė. DI etika yra itin svarbus vertybinis pagrindas, galintis užtikrinti patikimo ir tinkamo DI kūrimą bei plėtrą.

Tyrimo objektas: DI etinio reguliavimo reikšmė ir poveikis DI technologijų kūrimui, plėtrai ir įgalinimui visuomenėje.

Tikslas: išanalizuoti DI sampratą, teisinį reglamentavimą ES ir Lietuvoje, esminius DI etikos principus, iššūkius ir raidos perspektyvas.

Uždaviniai:

1. Išanalizuoti DI sampratą.
2. Išnagrinėti teisės aktus, dokumentus reglamentuojančius etišką DI pritaikymą.
3. Išanalizuoti esminius etinius iššūkius, su kuriais susiduriama DI srityje.
4. Išanalizuoti DI etikos principus, iššūkius ir raidos perspektyvas.

Metodai: kritinis, lyginamasis, loginis mokslinės literatūros ir šaltinių analizės ir sintezės metodai.

Darbo struktūra: ši rašto darbą sudaro tokios dalys: įvadas, kuriame pristatomas temos aktualumas, problema, tikslas, uždaviniai, tyrimo metodai bei darbo struktūra. Dėstymo dalis sudaryta iš trijų skyrių. Pirmajame skyriuje pristatoma DI koncepcijos samprata, aptariama DI

sąvoka, technologijų raida bei DI pritaikomumas ir praktinė nauda visuomenėje. Antrajame skyriuje pristatomas teisinis DI technologijų reguliavimas ES ir Lietuvoje, analizuojama ES DI strategija, taip pat Lietuvos DI strategijos teisinio ir etinio reguliavimo problema. Trečiame skyriuje pristatoma DI etikos problemos, iššūkiai ir raidos perspektyvos visuomenėje. Taip pat, darbe pateikiamos išvados ir literatūros sąrašas.

I. DI KONCEPCIJOS SAMPRATA

1.1. DI sąvoka

Ilgą laiką pats intelektas buvo suprantamas kaip neatsiejama žmogaus gyvenimo dalis, nesiejant jo su jokiais mechaninėmis sistemomis (Russel et al., 2010). Lietuvos Respublikos terminų bankas (2020) intelektą (lot. *intellectus* – suvokimas, prasmė), įvardija kaip proto sugebėjimą mokytis ir išmokti, susiprasti dar neįvykusiose situacijose bei išaiškinti reiškinių ryšius. Nepaisant skirtingų intelekto apibrėžimų sutinkama, jog tai yra gebėjimas suprasti ir mokytis iš aplinkinių faktorių (Ramonienė ir kt., 2012).

DI nagrinėjamas kaip mokėjimas suprasti, mokytis ir priimti sprendimus. DI pradžia siejama dar su Graikijos mitologija, kai Hefaisito mituose minimi kalviai, gaminantys mechaninius tarnus (AI Topics, 2020). 1956 m. Dartmuto (JAV) mieste vykusioje konferencijoje profesorius-informatikas Johnas McCarthy (John McCarthy) pirmasis paminėjo DI sąvoką. Mokslininkas DI apibrėžė labai paprastai, jo teigimu – itin pažangių kompiuterinių programų inžinerija, racionalių mašinų mokslas (Childs, 2011). Kiek seniau, 1945 m. Vannevaro Busho (Vannevar Bush) seminare „Kaip mes galime galvoti?“. Profesorius pasiūlė sukurti „atminties plėtinį“, kuris padėtų mokslininkams susidoroti su „augančiu tyrimų kalnu“, kuris nustebino žmones, kuomet suprato, kaip jie sunkiai ieškojo informacijos (Lavenda, 2017). Matematikas Alanas Tiuringas (Alan Turing) savo straipsnyje aprašė, kad mašinos gali sugebėti atkartoti bei atlikti žmonių veiksmus (TAT Institute, 2019).

Oksfordo žodyne (The Oxford English Dictionary, 2020). DI apibūdinamas kaip kompiuterinės sistemos, gebančios įveikti užduotis, kurios reikalauja žmogaus suvokimo: kalbos supratimo ir vertimo, sprendimų priėmimo bei regėjimo suvokimo. Kallem (2012) straipsnyje „Dirbtinio intelekto algoritmai“ DI šaką įvardijo kaip intelektinių sistemų konstrukciją ir analizę. Pagal UNESCO Publishing et al. (2018) DI – dirbtinis subjektas, turintis gebėjimą galvoti, suprasti, prisiminti, atpažinti ir priimti sprendimą iš galimų pasirinkimų, taip pat mokytis iš patirties. Atliekant DI tyrimus, gyvų organizmų nervų sistemos pagrindu kuriami teoriniai modeliai – neuroniniai tinklai. Šie tinklai, jei turi pakankamai duomenų, jie gali priimti sprendimus. Čia iškyla problema, kuomet DI yra lyginamas su žmogaus protu.

DI gynėjas buvo Alanas Tiuringas, kuris 1950 m. sukūrė Turingo testą, kurio tikslas išbandyti mašinų gebėjimą parodyti intelektą. Atliekant šį testą, kad kompiuterinė programa būtų pripažinta DI, ji turi mokėti atlikti šiuos veiksmus (Maknickienė, 2015):

1. Galėti sukurti ryšį su kompiuterine sistema (žmonių kalba);
2. Suvokti žinias prieš ir po tyrimo;

3. Automatiškai prisitaikyti – daryti išvadas, panaudojant turimas žinias;
4. Mokymosi procesas turi prisitaikyti prie naujų veiksmų.

Apibendrinus pateiktus apibrėžimus, DI suprantamas kaip sistema kuri naudoja mašinas, kad galėtų atlikti užduotis, kurioms dažniausiai reikalingas žmogaus intelektas: kalbos supratimas ir vertimas, gebėjimas priimti sprendimus, suvokti istorinius įvykius bei gebėti atlikti daugelį darbų.

1.2. DI technologijų raida

DI yra priskiriamas kompiuterių mokslo šakai. Nuo 1950 m. DI gyvuoja kaip atskira mokslo sritis. Kompiuteriniai žaidimai – pirmosios programos, kurios buvo vadinamos DI rezultatais. Puikus, pavyzdys, tai žaidimas šachmatais. Šiame žaidime, žaidėjo varžovas yra kompiuteris, kuris kiekviename ėjime pats priima sprendimus, siekdamas laimėti (Lupkowski, 2019). Viena iš ankstyvo DI kūrimo techninių problemų buvo programinės įrangos kalbos kliūtis. 1956 m. profesorius Johnas McCarthy ir mokslininkai Herbertas A. Simonas (Herbert A. Simon) ir Allenas Newelis (Allen Newell), kurie buvo išradę kokybišką programinę kalbą, sukurtą specialiai DI technikos plėtojimui ir moksliniams bandymams. Masačusetso technologijos institutas 1960 m. tapo DI vystymosi lopšiu. Būtent ten vykusių tyrimų pagalba mokslininkai priėjo išvadą, kad DI technologija yra glaudžiai susijusi su pagrindiniais žmogaus intelektinio elgesio modeliais. Ši išvada įkvėpė mokslininkus ne tik tyrinėti intelektinį elgesį, bet ir pradėti kurti „protingas“ kompiuterines mašinas. Taip buvo duota pradžia DI projektams, kurių pagrindinis tikslas yra sukurti protingą žmogaus darbo produktą, simuliuojantį žmogaus elgesį (Garnham, 2017). Kilęs susidomėjimas, ar kompiuteris gali spręsti įvairias problemas taip, kaip jas sprendžia kasdienis žmogus, mus atvedė iki šiandienos, kuomet, susidūrus su DI, kyla klausimas, o kaip gi šią technologiją apibrėžti? Mokslininkai dėl vienos konkrečios sąvokos nėra vienareikšmiškai apsisprendę.

Amilevičius (2017) savo darbe „Dirbtinis intelektas ir besiformuojančių technologijų etika“ DI apibūdina dviprasmiškai: „viena vertus, tai mokslo sritis, kuri tyrinėja kompiuterio ar įrenginio galimybes elgtis protingai arba tokią elgseną simuliuoti. Kita vertus, tai kompiuterio ar įrenginio gebėjimas elgtis protingai arba tokią elgseną simuliuoti“. Visuomenėje yra manoma, kad sunkumas apibrėžti DI kyla iš to, jog žmogus kaip būtybė yra vienintelis sutvėrimas, kuris pasižymi intelektu, o daugumai nėra priimtina tapatinti kompiuterines sistemas su žmogaus individualumu ir kūrybiniais bei protiniais sugebėjimais. Sternberg (2020) pabrėžia, kad net žmogaus protinis intelektas neturi savo asmeninės nusistovėjusios sąvokos, todėl apibrėžti DI yra ypač sunku, kuomet reikia įvertinti sąmoningumą, suvokimą, kalbines galimybes, galimybę mokytis ir prisitaikyti kaip keletą iš žmogaus intelekto sudedamųjų dalių.

Reiktų atkreipti dėmesį į tai, kad DI tikslas yra įtikinti žmogų, jog kompiuteris pats yra tarsi žmogus. Šią technologiją mokslininkai glaudžiai sieja su galimybe atlikti atitinkamus intelektinius uždavinius. Bostrom et al. (2014) intelekto sąvoką apibrėžia kaip „pasvertą sugebėjimą pasiekti tikslus pasaulyje“, o DI supranta kaip „protingų mašinų gamybos mokslą ir inžineriją, ypač intelektualių kompiuterinių programų atžvilgiu“.

Taigi atsižvelgus į tai, kad šiandien dauguma DI technologijos kūrinių pasižymi autonomiškumu ir gali atlikti tokius uždavinius, kaip mašinos vairavimas ar investicinių krepšelių sukūrimas be jokios aktyvios žmogaus kontrolės ir priežiūros, ir mokslininkų pateiktus DI sąvokos apibrėžimus, galima daryti išvadą, jog DI technologija pati iš savęs nurodo į mašinas, kurios pasižymi sugebėjimu atlikti tokias užduotis, kurias atliekant žmogui prireiktų tik *homo sapiens* būdingo protinio intelekto bruožo (Floris, 2020). Nors DI technologijos raidos istorija nėra ilga ir labai komplikuota, tačiau DI jau dabar žmogaus gyvenime atlieka labai svarbų vaidmenį. Dabartinis žmogus neįsivaizduotų savo kasdienybės be kompiuterinių mašinų ir pagalbos, kurią jos teikia. Tačiau ši technologija vis dar išlaiko neapibrėžtumą ir neturi konkretumo, kadangi šiandienos žmogus jaučia konkurenciją mašinoms, kurios pačios gali priiminėti savarankiškus ir individualius sprendimus, o tai apsunkina priimti vieną ir aiškų DI apibrėžimą.

1.3. DI ribos ir galimybės

DI yra techninis ir instrumentinis žmonių fizinių ir protinių galimybių tobulinimo tęsinys, galintis imituoti tam tikras funkcijas. DI yra skirtas tam tikram tikslui išspręsti. Naujosios sistemos dar negali užmegzti prasmingo dialogo su žmonėmis ar kūrybiškai įveikti problemų, tačiau mokslininkai teigia sukūrę savarankiškai besimokantį DI. Mašininio žmogaus proto analogą, kurio kūrimo sėkmė priklauso nuo inžinerinės ir techninės bazės plėtros, taip pat nuo žmogaus žinių apie sąmonę ir natūralų intelektą prieinamumo (Igorevna et al., 2017, Botha, 2019).

Problema galima apibrėžti taip: „mes nesuprantame savo sąmonės prigimties, todėl negalime sukurti DI“. Sąmoningą žmogaus veiklą galima apibūdinti kaip dialoginę. Žmogus visuomet stebi ir vertina save bei ginčijasi su savimi, bandydamas pažvelgti į save ir atsižvelgdamas į įvairius plėtros projektus, kurie suteikia galimybę tobulėti. Žmogus ir pasaulis laikomi glaudžiai susijusiais, žmogus neįsivaizduojamas be pasaulio, o pasaulis neegzistuoja arba negali egzistuoti be žmogaus būvimo. Žmogus įtraukiamas į egzistencinių procesų įvairovę ir tęstinumą. Apmąstymas ir savimonė pasirodo kaip konstitucinio egzistencinio vientisumo pradžia gyvenimo patirties pilnumoje, susijusioje su „viso pasaulio“ įrengimu (Igorevna et al., 2017).

Moksliniame diskurse terminas „atspindys“ dažnai vartojamas „savimonės“ prasme, tačiau jis nėra visiškai teisingas. „Apmąstymo“ ir „savimonės“ sąvokos, žinoma, yra artimos, bet nėra tapačios. Jei „atspindys“ įvardija veiklą, nukreiptą į žmogaus sąmonę, jos objektas yra vidinis subjektyvus sąnaus pasaulis. Terminas „savimonė“ apibūdina sąmonės būseną, atspindinčią asmens vientisumą subjektyvaus ir objektyvaus, idealo ir materialumo vienybėje (Igorovna et al., 2017).

Žmogaus sąmonė apima daugybę gestų kalbos išteklių, kai kurie iš jų turi pirmenybės pobūdį (Dubrovskiy, 2013). Taip, kibernetinės mašinos perėmė reikšmingą informacinių, simbolinių operacijų, kurias anksčiau atliko žmogaus smegenys, dalį. Jei mechaninės mašinos atlieka tipiškas žmogaus fizinės veiklos operacijas, kompiuterinės technologijos įformino protinių operacijų atlikimą asmeniui. Akivaizdu, kad mašina sugeba atlikti šias operacijas greičiau ir išsamiau, palyginti su pačiu žmogumi. Tačiau kadangi žmogaus sąmonė neatitinka gestų kalbos išteklių daugybės formalizavimo, mašininis atkūrimas yra problemiškas.

Efektyvų gestų kalbos priemonių bendravimo įgyvendinimą lemia asmens gebėjimas įsisąmoninti ir apmąstyti. Viena iš pagrindinių DI srities problemų yra sukurti sistemas, galinčias suvokti save, įskaitant savęs pažinimą apie savo vidines būsenas ir savybes bei supratimą apie juos supantį pasaulį.

Dabartiniai DI modeliai vis dar yra riboti ir netobuli. Intelektualiųjų sistemų modeliavimas su refleksija ir savimone reikalauja didelių tarpdisciplininių pastangų. Akivaizdu, kad daugelis DI klausimų yra neaiškūs ir yra neapibrėžtų taškų, kurie ateityje laukia paaiškinimo. Ar mašinos gali būti sąmoningos, ar sistemos su DI gali pretenduoti tik į „zombio“ titulą? Kompiuterių pažangos kontekste kognityviniai mokslai ir DI išreiškia „sąmonės reliatyvumo“ principą. Šio principo esmė yra ta, kad elgesio veiksmas gali vykti be sąmoningumo. Iš šio principo išplaukia, kad elgesys ir sąmonė gali egzistuoti nepriklausomai. Populiarus John Searle kinų kambario minties eksperimentas aiškiai parodo, kad mašina sugeba sukurti dialogo iliuziją, tačiau nėra sąmoningas ir nesugeba suprasti (Searle, 1980; Igorovna et al., 2017).

Filosofai aktyviai naudoja „zombio“ vaizdą iliustruodami filosofines teorijas DI tyrimų srityje. Kaip teisingai pastebi Chalmers (1996) filosofiniai „zombiai“ savo buveinę rado filosofinių straipsnių srityje. Filosofinis „zombis“ apibūdina būtybes, kurios yra filosofai savo minties eksperimentuose. Pavyzdžiui, Moody (1994) straipsnyje „Pokalbiai su zombiais“ pabrėžia išorines mašinos apraiškas, neturinčias sąmonės ar sąmoningo žmogaus požymių. Tęsdamas argumentą, daroma tokia išvada: kadangi DI sistemos yra būtent tokio tipo procesai (be supratimo), tai toks „kompiuterinis intelektas“ yra ne kas kita, o „zombis“. Jei pagrindinis sąmoningos veiklos bruožas yra intencionalumas, DI jo neturi ir yra tas pats „zombis“, kuris neturi sąmoningos patirties ir neturi savęs suvokimo.

Svarbu tai, kad kai kurios žmogaus intelektinės veiklos sritys, kurios neturi išreikšto kūrybinio pobūdžio, gali būti pateikiamos kaip aiškių instrukcijų (algoritmų) rinkinys ir perkeltos į mašiną. Bet šis DI neturi savimonės ir negali mokytis bei tobulėti kaip asmenybė. Modeliuoti intelektualias sistemas, kurios suvokia save, reikia rimtų tarpdisciplininių pastangų. Šį faktą gerai iliustruoja „zombio“ vaizdas. Dabartinis DI modelis, deja, yra ribotas ir netobulas. Šiuolaikinės DI sistemos pateikiamos kaip netobulos asmens kopijos, savotiškas „zombis“ (Igorovna et al., 2017).

DI jau seniai perėjo iš mokslinės fantastikos filmų į realųjį pasaulį. DI nematome kaip antropoidinio roboto, kuris galėtų elgtis ir mąstyti kaip žmogus. Šiandien, DI kalbasi su žmogumi per pokalbių bei paieškos robotus (sistemas), kurios slypi išmaniuosiuose įrenginiuose ir valdo kompiuterių tinklus, kur per kelias sekundes atliekami masiniai skaičiavimai. Kyla klausimas ar DI yra toks galingas, kaip mes linkę manyti?

Nors DI dar toli gražu netampa protingesnis už žmones, tačiau kasdieninėje žmogaus veikloje DI technologijos yra efektyviai naudojamos. Šiuo metu, galime užsisakyti maistą ar rezervuoti skrydžius, komunikuojant su paslaugų robotais, kurie yra pagrįsti DI. Tai tik paprasčiausi atvejai, tačiau sąrašas iš tikrųjų yra daug ilgesnis, kur žmogus naudoja DI (Adoriasoft, 2018):

- Padėjėjai balsu, tokie kaip: „Amazon Alexa“, „Google“ asistentas.
- Pažangūs pokalbių robotai, kurie: padeda elektroninių parduotuvių klientams; gali padėti nustatyti galimą sveikatos būklę medicinos programose; atliekant personalo interviu.
- Išmanieji įrenginiai – nuo laikrodžių ir termostatų iki savaime vairuojančių automobilių.
- Pristatymo dronai, priimančys klientų užsakymus.

Žinoma, dar ne viskas, tai tik keletas pavyzdžių ir galimybių kur pritaikomas DI. Neseniai „Google“ savo „Google skaičiuoklėse“ įdiegė DI algoritmus. Pagal įvestas reikšmes bei formules, DI gali atsakyti į klausimus ar komandas. Kitas potencialus sektorius, kuriame DI gali parodyti tikrąją galią, yra finansų technologijos („Fintech“), kur DI atlieka finansų rinkos analizę ir kuria investavimo strategijas

Nepaisant DI pranašumų ir daugybės sėkmingų praktinių įgyvendinimų, vis dar vyksta gana karštos diskusijos apie DI ateities galimybes. Daugelis ekspertų išsako savo nuomonę apie galimus DI pokyčius mūsų gyvenime, bet ne visos šios perspektyvos yra optimistiškos. Pagrindinis DI sukūrimo tikslas – puikūs skaičiavimo gebėjimai, informacijos kaupimas ir apdorojimas (Adoriasoft, 2018).

Sveikatos apsauga – sritis, kuri eksperimentuoja su visomis prieinamomis technologijomis ir daug dėmesio skiria DI. DI technologijos naudojamos diagnostikoje ir medicininiuose tyrimuose.

Naujųjų technologijų pagalba tikrinami šimtai skirtingų vaistų ir gaunama daugybė duomenų. Šių duomenų apdorojimas rankiniu būdu gali užtrukti metus ir atitolinti sėkmingo vaisto masinę gamybą. Čia praverčia analitinės DI galimybės, nes jis gali greitai ir tiksliai apdoroti tyrimų duomenis (European Union, 2020). Pavyzdžiui, šis metodas naudojamas farmacijos kompanijoje „Berg“, kuri kuria vaistus ir gydymą vėžiui ir Alzheimeriui gydyti (Medeiros, 2016; Berg, 2020).

Diagnostikos srityje DI taip pat gali padėti gydytojams rasti tikslus, individualius kiekvieno paciento gydymo būdus, tiriant paciento medicininius duomenis. IBM „Watson Health“ onkologijos platforma, naudojanti DI parenkant optimaliausius pacientų gydymo būdus (IBM, 2019).

Kita pramonė, pripažinusi DI galimybes, yra logistika. Logistika labai priklauso nuo planavimo ir prognozavimo, ir čia DI sistemos gali jai nepaprastai padėti. Vietoj reaktyvių operacijų, pagrįstų esama situacija, DI gali pakeisti logistikos procesus į aktyvius. Analizuodamas buvusius įvykius ir anksčiau gautus duomenis, DI gali numatyti būsimą paklausos padidėjimą ir padėti logistikos įmonėms efektyviau planuoti savo išteklius ir veiksmus. DHL, pasaulinė transporto korporacija, tiria DI galimybes automatizuojant darbo procesus. Įmonė derina DI pranašumus su daiktų interneto technologijos pranašumais, taip pasinaudodamas jų neprilygstamomis duomenų rinkimo ir apdorojimo galimybėmis (DHL, 2018).

Kalbėdami apie DI ateities galimybes, paprasčiausiai negalime ignoruoti finansų sektoriaus. Finansų pramonės sėkmė priklauso nuo duomenų ir jų gavimo bei apdorojimo greičio. Kripto ekonomikai prisijungus prie žaidimo, greitis tapo dar labiau pranašesnis, nes kripto srityje viskas yra daug greičiau.

Dėl šių iššūkių reikėjo sukurti naują pramonės šaką – „Fintech“. Pavadinimas išduoda, kad, naudojamos pažangiosios technologijų naujovės finansų srityje. DI duomenų apdorojimą perkelia į visiškai naują lygį, leidžiantį greitai ir efektyviai reaguoti į bet kokias rinkos tendencijas ir svyravimus. Sistemos geba greitai išanalizuoti informaciją, gaunamą iš įvairių šaltinių, skirtingomis kalbomis bei susieti su įvykiais ir pamatyti tendencijas (Adoriasoft, 2018).

Kai kurios finansų technologijų programos taip pat prideda „*blockchain*“ (liet. „blokų grandinė“) tai – skaitmeninė duomenų bazė, kuri veikia decentralizuotai: nėra prijungta prie vieno, bendro serverio, bet veikia kompiuterių tinkle. Ši technologija geba priimti visiškai naują sprendimą, greitai be rūpesčių vykdyti kriptografinę prekybą. DI algoritmai, analizuojantys buvusius ir esamus kriptografinės rinkos duomenis, sukuria investavimo strategijas, kuriomis siekiama gauti maksimalų pelną, minimaliai rizikuojant (Shroff, 2020).

Apibendrinant, DI atveria milžiniškas galimybes. DI negali suvokti pasaulio taip, kaip žmogus, tačiau tai suteikia žmonėms galimybę sutelkti dėmesį į prasmingesnius darbo aspektus, susijusius su

kūrybiškumu ir naujovėmis. DI geba kaupti ir analizuoti duomenis, automatika keičia įprastines ar pasikartojančias užduotis, todėl daugelio sričių darbuotojai gali daugiau dėmesio skirti išradimams, naujovėms, užtikrinamas didesnis darbo efektyvumas ir našumas. Visuotinė pareiga yra nustatyti tinkamas DI ribas: užtikrinti, kad naujosios technologijos atitiktų etinius kriterijus ir būtų orientuotos į žmogų, gerbtų principus ir vertybes.

1.4. DI pritaikomumas ir praktinė reikšmė visuomenei

Kaip minėta ankstesniame poskyryje, DI kažkada apsiribojęs mokslinės fantastikos romanais, filmais ir mokslo darbais, dabar daro didžiulį poveikį visuomenei. Verslo pasaulyje DI suteikia galimybę įmonėms dirbti protingiau ir greičiau, daugiau padaryti mažesnėmis sąnaudomis. Technologijoms ir visuomenei tobulėjant, daugiau organizacijų ieško galingų sprendimų, kurie pagerintų ir supaprastintų veiklą. Tačiau svarbu nepamiršti to, kad DI yra tarsi skėtis, po kuriuo gyvena daugybė skirtingų technologijų. Mašininis mokymasis, gilus mokymasis, robotika, kognityvinis skaičiavimas, dirbtinis bendrasis intelektas, natūralios kalbos apdorojimas yra tik keletas pagrindinių DI šakų.

DI, būtų galima palyginti kaip DI sugebėjimus ir kaip tose pačiose situacijose atsispindi žmogaus veiksmi. DI pagrindinėmis savybėmis laikoma:

- operavimas natūraliomis ar formaliomis sąvokomis (procedūros žinių bazėse);
- patirties kaupimas („mokymosi“ procedūros);
- funkcionavimas neapibrėžtumo sąlygomis.

DI sistemose nebūtinai turi būti vienodai išreikštos visos trys paminėtos savybės. Labiausiai charakteringa DI savybė (kaip ir natūraliam intelektui) yra operavimas sąvokomis – procedūros žinių bazėse, kurios nėra statiškos, o visuomet atnaujinamos, papildomos bei tobulinamos. Šią DI savybę galima būtų prilyginti žmogaus žinių kaupimui, kaip žmogaus natūralus protas linkęs prisiminti / kaupti ir atnaujinti informaciją / žinias, taip ir DI sistemoje papildomos ir tobulinamos žinių bazės. DI saugo visą veiklos istoriją ir puikiai geba atsiminti ir panaudoti saugomas žinias (Maknickienė, 2015), tačiau žmogaus smegenys, laikui bėgant, dalį sukauptos informacijos užmiršta. Taigi sukauptų žinių ir informacijos įsisavinimo klausimu DI sistema tampa pranašesnė už žmogaus smegenis.

Kita, bet labiausiai „nesavarankiška“ DI savybė laikytina antroji „patirties kaupimas“. „Patirties kaupimas“ prilygsta faktų registravimui ar statistinių duomenų kaupimui. Šią savybę būtų galima prilyginti žmogaus atsirinktos informacijos ar žinių kaupimui, tam, kad ateityje ją realizuoti praktiškai, tai yra, panaudoti surinktą informaciją praktikoje. DI sistemose „patirties kaupimas“

svarbus tik komplekse su kitomis dviem paminėtomis savybėmis (Maknickienė, 2015). Tai būtų tas pats, kas žmogaus informacijos rinkimas tam, kad vėliau ją panaudoti praktiškai geriausiu būdu iš surinktos informacijos, taip ir DI sistemoje žinių kaupimas, o vėliau jų realizavimas, neveiktų be pirmos savybės „operavimas“ ir toliau aprašomos „funkcionavimas“.

Trečioji DI savybė (funkcionavimas neapibrėžtumo sąlygomis). Iš tikrųjų, argi gebėjimas tiksliai ir greitai spręsti lygtis, atlikti operacijas su matricomis, taikyti matematinės logikos taisykles nėra intelektualumo apraška (jeigu ne filosofijos, tai bent jau psichologijos požiūriu)? Kitaip tariant, iš sukauptos informacijos DI apibendrina turimą informaciją ir atitinkamai ją panaudoja, „apibendrinimas iš esmės yra indukcinė išvada. Indukcija yra procesas, skirtas padaryti išvadą, atsižvelgiant į bendrąsias taisykles ar modelius, iš stebėjimų arba ieškant dėsningumų istoriniuose ar kituose duomenyse. Šis intelekto bruožas yra esminis“ (Maknickienė, 2015). Ši DI savybė yra labai svarbi, sudėtinga ir artima žmogaus intelektui, iš esmės apibendrinti ir padaryti išvadą iš sukauptų žinių ar praktikų reikalauja itin loginio mąstymo.

Taigi iš išdėstytų aplinkybių bei situacijų galime tik pastebėti, kuo skiriasi natūralusis intelektas nuo DI. DI turi panašias savybes kaip ir žmogaus intelektas, tačiau skirtumas yra tame, kad norint, jog DI sistema gebėtų pradėti mokytis ir rinkti informaciją, galiausiai ją apdorotų ir panaudotų praktiškai, visų pirma reikia į ją įdiegti tam tikrų žinių. Tai būtų labai panašu į natūralų žmogaus proto procesą, juk žmogus negimsta visko mokėdamas ir žinodamas. Informaciją ir žinias žmogus gauna iš visuomenės, iš knygų ar kitų šaltinių taip tą informaciją apdorojęs ir įsisavinęs suteiktas žinias panaudoja jas praktiškai gyvenime. Searle (1980) DI hipotezė teigia: „tinkamai suprogramuotas kompiuteris su teisingomis įvestimis ir išvestimis turi tokį pat protą, kaip ir žmonės, todėl DI algoritmai gali daryti tai, ką žmogus sugeba suprogramuoti“ (Maknickienė, 2015).

Kai kuriose pramonės šakose DI gali automatizuoti verslo ir analizės procesus, pateikdamas visapusišką ar net kelis variantus turintį sprendimą. Keletas pavyzdžių, kaip DI naudojamas efektyvumui gerinti (Wakefield, 2020):

- Bankininkystė ir finansai – sukčiavimo nustatymas. Daugelis bankų naudoja įvairias DI programas apgaulingai veiklai nustatyti. DI programinei įrangai pateikiama labai didelė duomenų imtis, apimanti apgaulingus ir nesąžiningus pirkimus, tuomet sistema yra mokoma atpažinti ar vykdoma operacija pagrįsta duomenimis. Laikui bėgant programinė įranga tampa nepaprastai įgudusi pastebėti nesąžiningus sandorius, remdamasi tuo, ko išmoko anksčiau.
- Mažmeninė prekyba - internetinis klientų aptarnavimas. Daugelyje internetinių svetainių dabar siūlomos tam tikros formos „pokalbių“ funkcijos, kuriose galima pasikalbėti su parduotuvės konsultantu. Daugeliu atvejų tai yra tam tikra automatinio DI forma, kuri pradeda šiuos

pokalbius. Kadangi šie DI pokalbių robotai gali suprasti natūralią kalbą, tai yra, žmonių pokalbius ir gali padėti klientams sužinoti reikalingą informaciją ar nukreipti juos į atitinkamą tinklalapį / asmenį tolimesniam veiksmui.

- Saugumas. Geriausios ir galingiausios įmonės visame pasaulyje daug investuoja į kibernetinį saugumą, kad užtikrintų duomenų apsaugą. Naudodami mašininio mokymosi algoritmus ir maitindami algoritmus dideliu kiekiu duomenų, informacinių technologijų ir saugumo ekspertai gali išmokyti DI stebėti elgesį, aptikti anomalijas, prisitaikyti ir reaguoti į grėsmes bei pateikti įspėjimus. DI pakankamai greitai tapo pagrindine verslo kibernetinio saugumo infrastruktūros sudedamąja dalimi, suteikiančia tvirtą ir sudėtingą daugiasluoksnią saugumo strategiją.

Plačiąja prasme, DI – informatikos mokslo sritis, kuria siekiama perduoti bei sukurti žmogaus intelektą kompiuterių sistemoms. Tai daroma naudojant iteracinį, daugiafunkcį modelio derinimą, dažniausiai viršijantį žmogaus ribas ir galimybes (greičiu ir paplitimu). Ne retai manoma, kad DI turės didžiulę reikšmę ne tik ekonomikos, mokslo, bet ir sveikatos priežiūros srityse (Frankenfield, 2020).

Įdomu tai, kad prognozuojama, jog DI gali turėti potencialų efektyvumą diagnozuojant ligas kaip ir sveikatos specialistai. Taikant sistemines apžvalgas bei analizes, DI remiantis mediciniais vaizdais geba diagnozuoti ligas. Siekiant iširti ligos pobūdį, kompiuterių sistemos geba apdoroti daugybę vaizdų, todėl tai suteikia galimybę tiksliau, greičiau ir efektyviau nustatyti diagnozę. Svarbu paminėti tai, kad tikrosios DI technikos, pavyzdžiui, kaip: gilus mokymasis, diagnostinės galios algoritmų, didelių duomenų ir skaičiavimo panaudojimas mėgdžiojant žmogų palieka nemažai klausimų. Taip yra todėl, kad nėra pakankamai atlikta mokslinių tyrimų, kurie analizuotų bei palygintų žmogaus ir DI sistemų veikimą aplinkoje.

Moksliniai tyrimai rodo, kad DI geba greičiau efektyviau bei greičiau prognozuoti ligą, stiprioji DI savybė – didelių duomenų rinkimas ir kaupimas, kurias galima diagnozuoti liga (pvz. radiologijoje, patologijoje). Sėkmingas DI panaudojimas matomas operacijų metu: stebini tikslesni, mažesni pjūviai, tad operacijos metu jaučiamas mažesnis skausmas, netenkama mažiau kraujo bei greičiau gyja žaizdos. DI gebėjimas kaupti didžiulį kiekį duomenų, sistemingai atlikti šių duomenų klasifikaciją ir analizę, parodo DI potencialą padėti ne tik ligų valdyme, bet ir diagnozavime. Puikiai matome, kaip kompiuterinės technologijos yra neatsiejama mūsų gyvenimo dalis, naudojama ir plačiai taikoma kasdieniniame gyvenime, pavyzdžiui, telefonuose, automobiliuose ar lėktuvuose. Svarbus aspektas tas, kad nepriklausomai nuo darbo vietos ar apimčių DI gali padėti pasiekti aukščiausių tyrimų rezultatų išaiškinimų, kurie gali turėti abipusę naudą, tiek gydytojams, tiek pacientams (Ahuja, 2019; Frankenfield, 2020).

Apibendrinant, DI pritaikomumas matomas ir analizuojamas ne tik ekonomikoje ar informatikoje, bet ir sveikatos priežiūroje. DI naudingas medicinoje, dėl gebėjimo kaupti informaciją ir ją analizuoti. Naujosios technologijos efektyviai naudojamos įvairiausiose srityse. Reikia nepamiršti to, kad DI panaudojimas ir iššūkiai yra glaudžiai susiję ne tik su teisiniais klausimais ir atsakomybe, bet ir su paciento požiūriu į etikos klausimus.

II. TEISINIS DI REGULIAVIMAS LIETUVOJE IR ES

2.1. DI technologijų teisinio reguliavimo problema

Sprendžiant galimos žalos klausimą, yra svarbu numatyti, koks yra DI išsivystymo lygis ir tobulėjimo prognozės bei ar dabartinė DI ir žmogaus koegzistencija tiesiogiai nurodo į galimą rimtos žalos atsiradimą ateityje. Svarbu identifikuoti, kokie yra DI ir žmogaus koegzistavimo santykiai, aptarti probleminius aspektus.

Pasaulio ekonomikos forumo dokumente, skirtame aptarti svarbiausias 2016 metų besiformuojančias technologijas, buvo pateiktas dešimties reikšmingiausių besikuriančių technologijų sąrašas, kuriame DI užėmė aštuntąją vietą (Amilevičius, 2017). Ši technologija, nors jauna, tačiau giliai įsiskverbusi į žmonių gyvenimus. Pavyzdžiui, finansų institucijos naudoja DI sistemomis, ieškodamos optimalių investavimo sprendimų, siekdamos atskleisti finansinio sukčiavimo atvejus ir analizuodamos rinkos tendencijas, o štai ligoninėse diegiamos tokios DI sistemos, kurios slaugo pacientus ar atlieka diagnostinius tyrimus. Netgi karinė pramonė neįsivaizduoja savo egzistencijos be DI sistemų, kadangi jos padėjo sukurti daugiau karinės artilerijos, o savaeigės transporto priemonės šiandien tikrai ne naujiena. Ir viskas tik DI technologijos pagalba. Nors DI kūrimas susiduria su daugybę problemų, stambios informacinių technologijų kompanijos, kaip, pavyzdžiui, Microsoft, IBM, Google ar Facebook, investuoja didžiules lėšas į tyrimus, todėl šios srities plėtra yra ypač sparti. Atsižvelgiant į tokias šiek tiek gąsdinančias prognozes, buvo atliktas detalesnis tyrimas. Mokslininkai Muller et al. (2016) tyrimo metu nustatė, kad tikėtina, jog apie 50 proc. esančių DI sistemų pasieks žmogaus visapusišką kūrybiškumo sugebėjimą maždaug 2040-2050 m., o maždaug 2075 m. procentas išaugs iki 90 proc. Tyrėjai priėjo išvadą, kad DI technologijos vadinamasis *super* intelektas per artimiausius kelis dešimtmečius išaugs iki tokio lygio, jog sistemos taps pilnai savarankiškoms (Muller, 2016). Yra numanoma, kad DI, su ištobulinta savimone ir sugebėjimu mokytis bei elgtis autonomiškai, galiausiai bus laikomas moraliu subjektu (atkreipkite dėmesį, kad ne objektu). Šią galimą situaciją mokslininkai apibūdina kaip didžiausią blogį žmogui ir visuomenei (Andrew, 2016).

Kai kuriose valstybėse keičiantis teisinei sistemai ir taikantis prie technologijų pokyčių, DI gali būti tapatinamas net su pačiu žmogumi (Bostrom et al., 2014). Vien ta tikimybė, kad DI sistema galės ateityje pati tobulinti savo techninę įrangą ir savarankiškai perprogramuoti sistemą taip, jog įgytų savo nuožiūra tinkamų sugebėjimų, rodo, kad gali kilti didelių problemų teisinėje sistemoje, kadangi galimai bus susidurta su kontrolės ir reguliavimo trūkumu.

Atkreipus dėmesį į teisinę sistemą, svarbu paminėti, kad DI jau šiandien po truputį skinasi kelią į ją. Nors bet kuris teisininkas galėtų teigti, kad teisinė praktika reikalauja išskirtinių ir su patirtimi būdingų protinių sugebėjimų, tačiau DI tikrai neatsilieka. Iš savęs besimokančių mašinų algoritmai sėkmingai naudojami teisinių prognozių byloje sudarymui. Žinoma, kadangi DI dar nėra labai ištobulintas, be žmogaus intelekto įsikišimo šiuo metu pasiekti tikslios prognozės yra neįmanoma. Tačiau mokslininkų teigimu, ateityje buvusių klientų informacijos pagrindu ir sujungus kitą aktualią informaciją iš su kliento situacija susietų teisinių precedentų, teisinės firmos galės naudoti DI tam, kad jis kurtų galimus bylos prognozių modelius, kurie numatys, pavyzdžiui, galimą kliento atsakomybę byloje (Yeung, 2019).

DI teisinėje sistemoje gali sukelti sumaištį. DI technologija gali būti ne tik kompiuteriai ir jų sistemos, tačiau ir robotai, kurie turi visas galimybes sukelti žmogui ne tik ekonominę ar psichologinę, tačiau ir fizinę žalą. Pastaroji gali būti sukelta dėl robotams atlikto netinkamo programavimo techninei, operacinei sistemoms bei programinei įrangai (Balkin, 2015). Galimybė, kad savamokslės mašinos taps neprognozuojamos žmogaus, kuris jas sukūrė, kelia nemažą susirūpinimą. Problema teisinėje sistemoje – nebus aišku, ką laikyti atsakingu už žmogui sukeltą žalą. Pasirinkimas platus, nes tai gali būti DI savininkas arba operatorius, techninės įrangos dizaineris, operacinės sistemos dizaineris ar programuotojas, o sąrašas tikrai nėra baigtinis (Reed, 2018).

Atsižvelgus į jau pateiktą informaciją, galima daryti išvadą, kad DI technologija šiandienos žmogui daro svarbų poveikį. Vis auganti DI reikšmė ekonomikoje ir visuomenėje kelia ir kels praktinių bei teorinių iššūkių teisinei sistemai. Praktinių iššūkių atsiras dėl to, kad savamokslės mašinas bus sunku prognozuoti ir kontroliuoti. Teorinių iššūkių atsiras dėl problemų, susietų su moraline ir teisine atsakomybe dėl sukeltos žalos ir to, kad DI vis dar neturi nusistovėjusios sąvokos.

Manoma, kad teisinė sistema susidurs ne tik su intensyvia DI tobulėjimo sparta, bet ir su DI kuriama žala bei problemomis užtikrinti, kad nukentėjusysis gautų jam priklausančią kompensaciją dėl patirtos DI sukeltos žalos (Scherer, 2015). Iš DI yra tikimasi daug, todėl yra svarbu neignoruoti šios technologijos egzistavimo ir to, kad ji iš tiesų labai sparčiai tobulėja bei gali kelti žmogaus egzistencijai ne tik naudą, bet ir nemažas problemas ar netgi sukurti tiesioginę žalą. Norint, kad žmogui būtų suteikta galimybė tinkamai pasiruošti atitinkamų spragų teisinėje sistemoje koregavimui, yra svarbu numatyti, kokio dydžio žalą gali sukelti DI.

Kalbant apie DI ir (ar) išmaniojo roboto teisinį reguliavimą yra išskiriamos dvi DI rūšys: tam tikros srities ir visuotinis. Tam tikros srities DI yra riboto funkcionalumo, jo teikiama nauda dalinė, todėl jis gali susidoroti tik su tam tikros srities užduotimis, tuo tarpu, visuotinis DI – sistema gali pati mokytis, kurti naujas žinias, priimti sprendimus, simuliuoti žmogaus intelektinę elgseną, tačiau ji

nebūtinai pasielgs taip, kaip tokioje situacijoje pasielgtų žmogus, dėl dviejų egzistuojančių priežasčių (Amilevičius, 2017):

- 1) žmogaus intelektinei elgsenai įtaką daro vertybių sistema ir moralė;
- 2) žmogaus racionalumas gali stipriai skirtis nuo mašinos racionalumo.

Apibrėžiant DI atliekamų funkcijų procesų ypatumus, galima įvardinti pagrindines sąvokas – mašininį mokymąsi ir gilųjį mokymąsi. Nors simbolinis DI galėjo išspręsti gerai apibrėžtas logines problemas, tokias kaip žaidimas šachmatais, pasirodė, kad to neužtenka daugumai realaus pasaulio problemų, tokių kaip: nuotraukų klasifikavimas, balso atpažinimas ir kalbos vertimas“ (Amilevičius, 2017). Laikui bėgant simbolinis DI buvo pakeistas mašininio mokymusi, o per pastarąjį dešimtmetį - gilioju mokymusi su neuroniniais tinklais. „Mašinos, pagrįstos simboliu DI, negalėjo mokytis iš naujų duomenų, tai paskatino atsirasti tyrimų sričiai, vadinamai mašininio mokymusi, kuris apibrėžiamas kaip „kompiuterio mokymas iš duotų duomenų“, užuot jį tiksliai užprogramavusi“ (Amilevičius, 2017, Maršalkaitė, 2018).

Gilus mokymasis yra suprantamas, kaip mašininio mokymosi srities, poskyris, „kur kiekvienas sluoksnis išmoksta vis abstraktesnius duomenų bruožus, jis vadinamas giliu nes turi daugiau nei kelis sujungtus neuronų sluoksnius, tuo tarpu modernus gilusis mokymasis dažniausiai turi tarp dešimties ir kelių šimtų sluoksnių, kurie išmoksta bruožus tiesiai iš duomenų“ (LeCun et al., 2015). Šiuo metu „giliųjų tinklų pasiekimai aprėpia beveik žmogaus lygio paveikslėlių klasifikaciją, kalbos atpažinimą, vairavimą, pagerintą mašininį vertimą tarp kalbų, pagerintą rezultatų paiešką“ (Chollet, 2018).

Taigi, mašinų mokymasis naudoja algoritmus duomenims analizuoti, mokytis iš šių duomenų ir priimti pagrįstus sprendimus, paremtus tuo, ką jis sužinojo. Gilių mokymosi struktūrų algoritmai sluoksniuose siekia sukurti „dirbtinį neuroninį tinklą“, galintį mokytis ir priimti protingus sprendimus savarankiškai.

Nors minėtieji atliekamų funkcijų procesų ypatumai patenka į plačią DI kategoriją, gilus mokymasis yra tai, kas prilyginama žmogiškiausiam DI. Nagrinėjant išmaniojo roboto, kaip fizinės DI išraiškos, sampratą 2017 m. vasario 16 d. Europos Parlamento rezoliucijos su rekomendacijomis Komisijai dėl robotikai taikomų civilinės teisės nuostatų (toliau – Rezoliucija) kontekste yra svarbios dvi sąvokos: „autonominis robotas“ ir „protingas robotas“ (Bartkus ir kt., 2018). Rezoliucijos AA punkte roboto autonomiškumas apibrėžiamas kaip gebėjimas „priimti sprendimus ir juos įgyvendinti išoriniame pasaulyje, nepriklausomai nuo išorės jėgų kontrolės ar poveikio. Autonomiškumas yra tik technologinio pobūdžio ir jo mastas priklauso nuo to, kokio tobulumo roboto sąveika su jo aplinka buvo numatyta“ (Europos Parlamentas, 2017).

Remiantis Rezoliucijos AB punktu, „kuo robotai autonomiškesni, tuo mažiau juos galima laikyti paprastais įtaisais kitų subjektų (pvz., gamintojų, operatorių, savininkų, naudotojų ir pan.) rankose, dėl to, kyla abejonų, ar įprastų taisyklių dėl atsakomybės pakanka, ar reikia taikyti naujus principus ir taisykles, siekiant suteikti aiškumo dėl įvairių subjektų teisinės atsakomybės – atsakomybės už robotų veiksmus ir neveikimą, kai priežasties negalima atsekti iki tam tikro žmogiškojo subjekto, ir ar robotų veiksmų ar neveikimo, kurie sukėlė žalą, buvo galima išvengti“ (Europos Parlamentas, 2017). Nesant aiškaus ir tinkamo roboto apibūdinimo Rezoliucijoje, jo apibrėžimas randamas Tarptautinės standartų organizacijos 2012 m. parengtų techninių standartų EN ISO 8373 „Robotai ir robotikos prietaisai – žodyne“. 2.28 punktas nustato, kad tai „mašina, kuri sugeba atlikti užduotį jausdama aplinkui esančią aplinką: sąveikaudama su išorės šaltiniais ir (arba) gebėdama pritaikyti prie jų savo elgesį“. Protingi robotai yra tarsi tiltas tarp robotų ir DI, kitaip tariant, robotai, kuriuos kontroliuoja DI sistemos.

Jeigu bent trumpam nusižengsime baudžiamosios teisės principams ir įsivaizduosime, kad DI gali būti baudžiamosios teisės subjektu, kaip pritaikyti individui taikytinas sankcijas autonominei sistemai, kuri gali ir turėti fizinį kūną (pvz., robotai), ir neturėti jo (sistema, instaliuota kokiam nors įrenginyje)? Panašus klausimas kilo ir anksčiau, kai buvo sprendžiama korporacijų baudžiamoji atsakomybė (Amilevičius, 2017). Tačiau esmė buvo pritaikyti subjektui tokią bausmę, kuri atneštų tokią pat pasekmę, kokią atneštų, jeigu būtų taikoma žmogui. Tokiu principu reiktų vadovautis ir taikant sankcijas DI. Pavyzdžiui, mirties bausmė. Žmogui mirties bausmė reiškia gyvybės praradimą. DI gyvybė būtų jo, kaip subjekto paprasčiausias egzistavimas fizinėje arba programinėje erdvėje. Tam, kad DI taikyti mirties bausmę, reiktų panaikinti programinę įrangą, kuri kontroliuoja DI (Hallevy, 2010). Imkime kitą sankciją – laisvės atėmimas. Tai yra viena populiariausių sankcijų baudžiamosios teisės sistemoje. Žmogui laisvės atėmimas reiškia fundamentaliosios teisės į laisvę atėmimą atitinkamam laikotarpiui, kai yra varžomos jo elgesio, judėjimo laisvės ir teisė į privatų asmeninį gyvenimą (Scherer, 2015). Laisvė DI reikštų jo galimybę elgtis autonomiškai ir atlikti jam užprogramuotą užduotį, pavyzdžiui, atlikti darbus fabrike, dalyvauti operacijose ir pan. Norint tinkamai pritaikyti laisvės atėmimo bausmę DI, reiktų neutralizuoti DI, t. y. jį išjungti, kokiam nors konkrečiam laiko periodui (Hallevy, 2010). Šiame laikotarpyje DI nebūtų galimybės elgtis autonomiškai, todėl jo laisvė būtų suvaržyta. Kitos dvi populiarios sankcijos – pinigine bausme ir viešieji darbai. Žmogui pinigine bausme apima tam tikrą turto atėmimą dažniausiai pasireiškiantį pinigine išraiška. Jeigu asmuo negali sumokėti pinigines baudos, jam pritaikomos kitos sankcijos (dažniausiai laisvės atėmimas) (Bostrom, 2014). Kadangi DI neturi konkrečios nuosavybės arba turto, įprastu būdu pinigines baudos kaip sankcijos pritaikyti DI negalime. Tačiau paprastai žmogus pajamas ar turtą įgyja darbo būdu, todėl galima daryti išvadą, kad nesugebant sumokėti pinigines baudos, už ją galima atsiskaityti darbu (Locke, 1988). DI atsiskaitymą darbu būtų galima prilyginti

viešiesiems darbams. Viešųjų darbų kaip sankcijos esmė yra privalomojo darbo įsipareigojimas bendruomenei. Kadangi DI gali būti prilygintas darbuotojui, jeigu jis atlieka konkrečias užduotis, kurios konkrečioje darbo įstaigoje atnešą naudą ir pelną, vadinasi, DI galima pritaikyti viešųjų darbų sankciją, kurią būtų galima prilyginti ir piniginės baudos sankcijai, kurią DI įgyvendintų konkretaus darbo atlikimu, kas sukurtų piniginę vertę.(Hallevy, 2010). Taigi žmogui taikomas baudžiamosios teisės sistemos sankcijas tikrai nesudėtingai galima pritaikyti ir DI. Tačiau, kadangi DI dėl psichologinio komponento trūkumo kol kas nepripažįstamas baudžiamosios teisės subjektu, įprastas sankcijas pritaikyti galime tik asmeniui, atsakingam už DI.

Nors numatytos sankcijos ir nubaūs už DI atsakingą asmenį, tačiau tai nebūtinai užtikrins, kad DI toliau neatliks nusikalstamų veikų, kadangi sankcija neturės konkrečios įtakos DI (Charney, 2015). Todėl kyla poreikis prie įprastos sankcijos nusikalstamos veikos subjektui pridėti ir pareigą atlikti tam tikrus veiksmus DI atžvilgiu, t. y. pritaikyti konkrečias sankcijas DI, nors jis ir nėra baudžiamasis subjektas. Hallevy (2011) siūlo dvi sankcijas, kurias, kaip pareigą prie jam priskirtos sankcijos turėtų atlikti DI kūrėjas / programuotojas – DI rehabilitacija (perprogramavimas) arba DI neutralizacija (išjungimas).

Kaip minėta anksčiau, DI išjungimas veiktų kaip laisvės atėmimo sankcija. Tarkime, jeigu nusikalstamos veikos subjektui būtų taikomas laisvės atėmimas, lygiai taip pat prie sankcijos būtų galima pridėti DI kūrėjo pareigą neutralizuoti savo DI. Tuo laikotarpiu, kada subjektas atlikinės jam paskirtą laisvės atėmimo bausmę, DI taip pat negalės veikti, nes bus išjungtas, todėl taip bus užtikrintas saugumas, kad nusikalstama veika nepasikartos. Pasibaigus laisvės atėmimo laikotarpiui, subjektas galėtų vėl įjungti DI, tačiau tuo atveju, jeigu DI pats iš savo veiksmų nesugebėtų išmokti, kad atitinkami poelgiai yra netinkami, kūrėjo / programuotojo pareiga būtų perprogramuoti DI taip, kad ateityje nusikalstamas veikimas nesikartotų (Beato, 2015). Lygiai tokia pati perprogramavimo pareiga būtų susieta su kitomis subjektui taikomomis sankcijomis, kaip, pavyzdžiui, piniginė bauda ar viešieji darbai (kuriuos atlikti, kaip minėta, gali ir pats DI). Jeigu DI rehabilitacija nebūtų veiksminga ir nusikalstamas veikimas kartotųsi, tada už DI atsakingam asmeniui kiltų pareiga neutralizuoti savo kūrinį visam laikui (Hallevy, 2013).

Už DI sukeltą žalą atsakingam asmeniui taikomos baudžiamosios sankcijos neužtikrina, kad DI ateityje nekartos savo nusikalstamų veiksmų ir nesukels baudžiamąjį dydžio žalos. Tai užtikrinti gali tik konkretūs veiksmai, nukreipti tiesiogiai į DI. Kadangi dauguma populiariųjų žmogui taikomų baudžiamųjų sankcijų galima iš esmės pritaikyti ir DI, baudžiamosios teisės subjektui gali būti taikoma pareiga neutralizuoti arba reabilituoti savo kūrinį, kas padėtų užtikrinti, kad DI veikimas ateityje bus saugus aplinkai.

Apibendrinant, galima teigti, kad prevencijos būdų užkirsti kelią DI žalai yra ne vienas, todėl svarbu neignoruoti esamos ir tik didėjančios problemos bei aktyviai domėtis DI ir jo galimybėmis. Aktyvus Europos Parlamento siūlymas kurti organizaciją, kontroliuojančią DI, rodo, kad tokios įstaigos egzistavimas yra reikalingas, o žmonių, susijusių su DI kūrimu, atskaitomybę ir atsakomybę padėti įgyvendinti gali įstatymo leidėjo aktyvus dalyvavimas ir domėjimasis su DI susijusia veikla, kas suteiktų daugiau konkretumo ir aiškumo, kokius teisinius aktus, kontroliuojančius DI, reikėtų paskelbti. Nors DI sukurtos žalos atveju baudžiamuoju subjektu lieka žmogus, tačiau žmogui taikoma sankcija negarantuoja, kad DI ateityje neveiks nusikalstamai. Todėl yra formuluojama išvada, kad šalia subjektui taikomos sankcijos kartu turi būti taikoma pareiga, proporcingai įvertinus sukurtą žalą, reabilituoti arba neutralizuoti savo DI.

2.2. ES DI strategija: teisės ir etikos dermė

DI vystymui ir pritaikymui svarbus žmogaus ir visuomenės pasitikėjimas ir atsakomybė. Europos Sąjungos pagrindinių teisių chartijoje (toliau – ES chartija) yra įtvirtintos piliečių socialinės, ekonominės, politinės teisės. ES remiasi patikima, tvaria teisine reguliavimo sistema, kuria vadovaujantis kuriami ne tik teisės aktai, projektai bei įvairūs standartai, bet ir grindžiamas inovatyvių technologijų kūrimas. Griežti kriterijai taikomi gaminių saugumui ir atsakomybei. Taip pat nuo 2018 m. ES pradėtos taikyti griežtos asmens duomenų apsaugos taisyklės (Europos Sąjunga, 2016).

Bendrojo duomenų apsaugos reglamentu (toliau – BDAR) nustatomi asmens duomenų apsaugos standartai ir principai. EK atidžiai stebi BDAR nuostatų taikymą DI srityse ir skatina taip elgtis nacionalines duomenų apsaugos institucijas bei Europos duomenų apsaugos valdybą (Europos Sąjunga, 2020).

Siekiant pelnyti žmonių bei visuomenės pasitikėjimą DI sistemomis, yra svarbus šių technologijų veikimo supratimas. Todėl kuriant DI technologijas reikia siekti, kad jos būtų suprantamos, skaidrios ir būtų išvengta klaidų atsiradimo rizikos. DI, kaip ir kitos technologijos gali būti naudojamos ne tik dorais, bet ir savanaudiškais tikslais. DI gali prisidėti prie naujų atradimų, technikos ir įrenginių tobulėjimo, tačiau kelia ir saugos klausimų, nes gali būti panaudotos prieš patį žmogų, iš neigiamų paskatų: nusikaltimams, išpuoliams ar diskriminacijai. Todėl labai svarbu etiškas, darnus, teisiškai įtvirtintas ir naujų technologijų vystymas tiek kūrėjų, tiek vartotojų požiūriu (Europos Komisija, 2018).

Siekiant užtikrinti ES chartijoje įtvirtintus nuostatus bei užkirsti kelią etikos problemų atsiradimui parengtos DI etikos gairės. Gairės skirtos įvairiems etiškumo klausimams spręsti, tokiems kaip: nedarbas, socialiniai klausimai, saugumo, sąžiningumo, tinkamumo. Projekto tikslas – stebėti poveikį pagrindinėms teisėms, įskaitant privatumą, orumą, vartotojų apsaugą ir nediskriminavimą (Europos Komisija, 2018).

Užtikrinant ES reglamentuojamas vertybes ir teises, valdžios institucijos turi apsaugoti ir garantuoti, kad DI plėtra ir pritaikymas atitiktų keliamus reikalavimus. Tinkamam DI sistemų sukūrimui itin reikšminga apgalvoti saugos normų bei civilinės atsakomybės teisės aktų keliamus klausimus. Štai, daiktų, robotų, technikos ar sistemų, kurių veikimui naudojamas DI veikimas gali būti kitoks, nei buvo numatyta pradedant eksploatuoti. Kadangi DI yra visapusiškai naudojamas ir paplitęs, būtina atsižvelgti į horizontaliuosius principus ir įmonėms, įstaigoms ir sritims taikomas taisykles.

ES saugą reglamentuojančiuose teisės aktuose, pavyzdžiui, Mašinų direktyvoje, Radijo ryšio įrenginių direktyvoje, Direktyvoje dėl bendros gaminių saugos ir specialiose saugos normose: medicinos priemonėms arba žaislams yra žiūrima į problemas, kurios susijusios su rinkai pateiktų produktų numatyta paskirtimi ir tinkamumu naudoti (Europos saugos ir sveikatos darbe agentūra). Šiuo tikslu buvo parengta DI naudojančių technologijų sferoje naudojama standartų sistema, kuri yra tobulinama vadovaujantis kuriamomis inovacijomis. Galimybė užtikrinti saugumą ir teises, stiprina vartotojų pasitikėjimą ir didina DI technikos priimtinumą visuomenėje. (Europos Komisija, 2018). Taip pat labai svarbus aspektas naudojant DI tai – vartotojų apsauga, sąžiningumas, skaidrumas ir šių principų laikymasis. Todėl naudotojams turi būti pateikta suprantama informacija apie DI pagalba sukurtų daiktų (produktų) vartojimą, savybes, pranašumus ir trūkumus.

DI, kaip ir kitos naujosios technologijos kelia riziką bei daugybę politinių uždavinių, tokių kaip: patikimumo, saugumo ir kontroliavimo, privatumo, etikos, socialinėse, ekonominėse bei daugelyje kitų sričių (EESRK, 2017). Dėl šių priežasčių būtina nustatyti gaires esamam ir tolesniam DI vystymui ir naudojimui taip pat, teisinių ir etinių problemų sprendimui, kurie yra susiję su galimu DI šališku sprendimų priėmimu ir atsakomybės klausimais (EK, 2018). Norint pasiekti šį tikslą, kiekviena ES valstybė narė yra skatinama paruošti DI strategiją, kuri reglamentuotų galimai DI sukeltos žalos atlyginimą.

2018 m. balandžio 25 d. EK paskelbė komunikatą „Dirbtinis intelektas Europai“. Komunikate rašoma, kas DI vystymui ir pritaikymui privaloma sukurti pasitikėjimą ir atskaitomybę užtikrinančias sąlygas. Be to, siūloma apsvarstyti numatytų saugumo normų tinkamumą bei su civiline atsakomybe susijusius klausimus, teisinius aspektus. Europos DI iniciatyva yra siekiama įvertinti nacionalinių ir ES teisės aktų, skirtų saugai ir atsakomybei tinkamumą, atsižvelgiant į naujus iššūkius, trūkumus ir problemas, kurias reikia pašalinti (EK, 2018).

Vadovaujantis ES Komisijos paskelbtoje DI strategijoje Europos šalys sutiko dirbti kartu siekiant padidinti reikšmingą investavimą, duomenų koordinavimą bei etiką DI srityje. Siekiama palengvinti ir sustiprinti bendradarbiavimą tarp ES šalių, taip pat padidinti ES konkurencingumą ir užtikrinti pasitikėjimą ES vertybėmis. Koordinuotame plane siekiama užtikrinti nacionalinių ir ES lygmens veiksmų eigos papildymą ir sąveiką (European Commission, 2020).

Kuriant DI, kuriuo pasitikėtų žmogus: Europos DI strategijoje ir suderintame plane pasitikėjimas numatomas kaip išankstinė sąlyga siekiant užtikrinti žmogaus orientuotą požiūrį į DI. Aukšto lygio DI ekspertų grupė 2018 m. gruodžio mėn. pateikė pirmąjį gairių projektą. Po tolesnių ekspertų grupės svarstymų, atsižvelgiant į diskusijas apie Europos DI aljansą, konsultacijas su suinteresuotosiomis šalimis ir susitikimus su valstybių narių atstovais, gairės buvo persvarstytos ir paskelbtos 2019 metų balandį. Lygiagrečiai DI ekspertų grupė taip pat parengė pataisytą dokumentą, kuriame išsamiau aprašomas DI apibrėžimas, naudojamas jo rezultatams. Užtikrinant, siekiamas europinės vertybes, kurių pagrindas yra pasitikėjimas aplinka sėkmingam DI vystymuisi ir naudojimui Komisija komunikate pabrėžia pagrindinius reikalavimus, keliamus patikimam DI (Vincent, 2019):

- Žmogaus teisės. DI turėtų gerbti žmogaus teises, pripažinti lygiateisę visuomenę, o ne mažinti ar apriboti žmogaus autonomiją.
- Patikimumas ir saugumas. Patikimo DI algoritmai turi būti saugūs, patikimi ir teisingi, kad būtų galima išspręsti klaidas visuose sistemų gyvavimo ciklo etapuose.
- Privatumas ir duomenų valdymas. Piliečiai privalo galėti visapusiškai kontroliuoti savo duomenis, jie negali būti naudojami žmonėms kenkti ar diskriminuoti.
- Skaidrumas. Turėtų būti užtikrintas DI sistemų atsekamumas.
- Įvairovė, nediskriminavimas ir teisingumas. DI sistemose turėtų būti atsižvelgiama į visą žmogaus gebėjimų, įgūdžių ir reikalavimų spektrą, užtikrinamas prieinamumas.
- Visuomenės ir aplinkos gerovė. DI sistemos turėtų būti naudojamos siekiant skatinti teigiamus socialinius pokyčius ir didinti tvarumą bei ekologinę atsakomybę.
- Atskaitomybė. Turėtų būti sukurti mechanizmai, užtikrinantys atsakomybę ir atskaitomybę už DI sistemas ir jų rezultatus, bei pasekmes.

2018 metų gruodžio mėn. Valstybės narės suvienijo jėgas su EK dėl suderinto DI plano, kad būtų padidintas bendradarbiavimas, kuris paskatins DI Europoje. DI koordinavimas yra būtinas, nes (European Commission, 2020):

- ✓ Visoms Europos šalims dirbant kartu galima maksimaliai išnaudoti DI teikiamas galimybes, kurios būtų panaudojamos visuomenės ateičiai.
- ✓ Europa nori pirmauti DI srityje etiškumu ir vertybėmis, kad piliečiai ir įmonės galėtų visiškai pasitikėti naudojamomis technologijomis.
- ✓ Valstybių narių ir Komisijos bendradarbiavimas yra būtinas sprendžiant naujus DI keliamus uždavinius.

DI vystymuisi ir plėtrai reikalingas žmonių pasitikėjimas. Norint pelnyti pasitikėjimą, DI turi būti etiškas, atitikti etikos standartus, kurios atspindi mūsų vertybes. Todėl sprendimų priėmimas turi būti suprantamas ir orientuotas į žmogų. Tai reikalauja plačios, atviros ir įtraukios diskusijos apie tai, kaip sėkmingai ir etiškai pagrįsti DI bei jį naudoti ir plėtoti.

Siekiant užtikrinti tinkamą etinę ir teisinę sistemą 2020 m. vasario 19 d. EK paskelbė Baltąją knygą, kuria siekiama puoselėti Europos meistriškumo ir pasitikėjimo DI ekosistemą bei pateikė ataskaitą apie DI saugą ir atsakomybę. Baltojoje knygoje siūloma (European Commission, 2020):

- Priemonės, kurios supaprastins mokslinius tyrimus, skatins valstybių narių bendradarbiavimą ir padidins investicijas DI kūrimui ir diegimui;
- Politinės būsimos ES reguliavimo sistemos galimybės, kurie nustatytų teisinių reikalavimų, taikomų atitinkamiems dalyviams, rūšis, ypatingą dėmesį skiriant didelės rizikos sistemoms ir jų taikymams.

Dokumentas apibūdinamas kaip koordinuotas požiūris, siekiant maksimaliai padidinti naudą ir spręsti iškilusius iššūkius ir sunkumus susijusius su DI.

ES aukšto lygio DI ekspertų grupė po to, kai išleido naują politikos ir investicijų rinkinį – patikimo DI gairės (AI HLEG, 2019b), įskaitant svarbias rekomendacijas, kaip apsaugoti žmones, pritaikymas privačiame sektoriuje, mokslinių tyrimų atlikimas bei etiška duomenų valdymo praktika. EK pateikia vieningą DI intelekto plėtros ES pagrindą, tačiau ES valstybė narės taip pat privalo parengti savo nacionalines strategijas.

Suomija buvo pirmoji valstybė narė, sukūrusi nacionalinę DI programą (MEAEF, 2018a). Programa yra pagrįsta dviem ataskaitomis „*Finland's Age of Artificial Intelligence*“ ir „*Work in the Age of Artificial Intelligence*“ (MEAF, 2017; MEAF, 2018b). Politikos tikslai skirti verslo investicijų konkurencingumui ir viešosioms paslaugoms.

Danijos nacionalinė DI strategija (The Danish Government, 2019) buvo išleista 2019 m. kovo mėn. vadovaujantis „Skaitmeninio augimo strategija“ („*Strategy for Digital Growth*“) (The Danish Government, 2018). Šioje išsamioje strategijoje išvardinti tikslai, įskaitant atsakingo DI pagrindo sukūrimą, aukštos kokybės duomenų teikimą ir investicijų didinimą (ypač žemės ūkio, energetikos, sveikatos priežiūros ir transporto sektoriuose). Didelis dėmesys skiriamas duomenų etikai, įskaitant atsakomybę, saugumą ir skaidrumą bei etinės sistemos poreikio pripažinimą. Danijos vyriausybė išdėstė šešis etinio DI principus: apsisprendimą, orumą, atsakomybę, aiškumą, lygybę ir teisingumą bei plėtrą (sprendimai, palaikantys etiškai atsakingą DI kūrimą ir naudojimą siekiant visuomenės pažangos).

Prancūzijoje „DI žmonijai“ („*AI for Humanity*“) išleistas 2018 m. kovą, šiuo dokumentu įsipareigojama remti prancūzų talentą, efektyviau ir tiksliau panaudoti duomenis ir nustatyti DI etikos sistemą (AI For Humanity, 2018). Prezidentas Emanuelis Macronas (Emmanuel Macron) įsipareigojo užtikrinti skaidrumą ir sąžiningą DI naudojimą švietimo sistemoje. Strategija daugiausia grindžiama Prancūzijos matematiko ir politiko Sedriko Vijanio (Cédrico Villani) darbu. 2018 m. o pateiktoje ataskaitoje dėl DI buvo pasiūlytos rekomendacijos ekonomikos politikai, mokslinių tyrimų infrastruktūrai, užimtumui ir etikai (Villani, 2018).

Vokietijos DI strategija buvo priimta 2018 m. lapkričio mėn. (Die Bundesregierung, 2018) ir išsisklė tris pagrindinius pasižadėjimus: padaryti Vokietiją pasauline lydere DI kūrimo ir naudojimo srityje, užtikrinti atsakingą DI plėtrą ir naudojimą, integruoti DI visuomenėje (etika, teisė, kultūra).

Švedijos požiūris į DI (GOS, 2019) turi mažiau konkrečių terminų, bet pateikia bendrąsias DI švietimo, mokslinių tyrimų, inovacijų ir infrastruktūros gaires. Rekomendacijos apima stiprios tyrimų bazės sukūrimą, bendradarbiavimą tarp sektorių (taip pat partnerystę su kitomis šalimis), plėtoti pastangas siekiant užkirsti kelią rizikingam elgesiui bei parengti etiško DI naudojimo standartus. Švedijos DI taryba, sudaryta iš pramonės ir akademinės bendruomenės ekspertų sudaryta siekiant sukurti „švedišką intelekto modelį“, kuris, jų teigimu, bus tvarus, naudingas ir skatins ilgalaikį ekonomikos augimą (Swedish AI Council, 2019).

Austrija įsteigė „Robotų tarybą“ („*Robot Council*“), kad padėtų vyriausybei sukurti nacionalinę DI strategiją (Austrian Council on Robotics and Artificial Intelligence, 2018). Tarybos parengtoje baltojoje knygoje pateikiamas strategijos pagrindas. Socialiai orientuotame dokumente yra keliami tikslai skatinti atsakingą DI naudojimą, parengti priemonės pavojams atpažinti ir sumažinti, sukurti teisinę bazę siekiant užtikrinti duomenų saugumą ir užmegzti viešą dialogą apie DI pritaikymą (Austrian Council on Robotics and Artificial Intelligence, 2018).

Italija įsteigė DI darbo grupę (Agency for Digital Italy, 2019) nustatyti siūlomas DI galimybes ir pagerinti viešųjų paslaugų kokybę. Italų strategija (Task Force on Artificial Intelligence of the Agency for Digital Italy, 2018) paskelbta 2018 m. kovo mėn., kurioje aprašoma, kad etika yra pirmasis iššūkis sėkmingai diegiant intelektą, kad DI turėtų tarnauti piliečiams ir užtikrinti lygybę naudojant technologijas patenkinti visuotinius poreikius. Darbo grupė apibūdina su technologijomis susijusius iššūkius: duomenų prieinamumo ir kokybės, patirties ir teisinio reglamentavimo trūkumo.

Lenkija turi savo nacionalinę DI strategiją. 2019 m. paskelbta ataskaita daugiausia dėmesio skiria DI ekosistemai Lenkijoje, kaip nacionalinės DI strategijos pirmtakei. Nors joje pateikiama išsami šiuolaikinės Lenkijos apžvalga, tačiau nėra konkrečių rekomendacijų Vyriausybei bei DI etikos klausimų (Digital Poland Foundation, 2019).

Visoje ES: visuomenės požiūris į robotus ir skaitmeninimas. Apskritai DI, robotikos ir pažangių technologijų Europos perspektyvų tyrimai atspindi, kad piliečiai iš esmės teigiamai vertina DI, kaip ekonomikos ir visuomenės papildymą. Tačiau šis požiūris skiriasi priklausomai nuo amžiaus, lyties, išsilavinimo, pavyzdžiui, tik maža respondentų dalis turėjo patirties naudojant robotus, o turintys patirties, į DI labiau žiūrėjo teigiamai nei tie, kurie patirties neturėjo (European Commission, 2012; European Commission, 2017; European Commission, 2018).

Bendros visuomenės nuomonės tendencijos tyrimuose parodė, kad respondentai:

- Palaiko robotų naudojimą ir skaitmeninį darbuose, kurie kelia pavojų ar sunkumus žmonėms (pavyzdžiui, kosmoso tyrinėjimai, gamyba ar karinis sektorius);

- Susirūpinę, kad tokiai technologijai reikia veiksmingo ir kruopštaus valdymo;
- Nerimauja, kad dėl automatizavimo ir skaitmeninimo bus prarasta darbo vieta;
- Nepalaiko robotų naudojimo pažeidžiamiems visuomenės nariams prižiūrėti (pagyvenę, sergantys, augintiniai ar tie, kuriems atliekamos medicininės procedūros);
- Nerimauja dėl prieigos prie jų duomenų ir internetinės informacijos bei jų apsaugos (antivirusinė programinė įranga, pakeistas naršymo elgesys);
- Nenori važiuoti automobiliu be vairuotojo (tai mielai padarytų tik 22 proc.);
- Nepasitiki socialine žiniasklaida, tik 7 proc. žiūri istorijas, paskelbtas socialiniuose tinkluose kaip „paprastai patikimas“.

Taigi šie klausimai ir aspektai, pastebėjimai yra labai svarbūs Europos DI iniciatyvose ir atspindi nuomonę apie robotų kūrimą, diegimą ir pritaikymą: automatizavimą ir skaitmeninimą visose gyvenimo srityse: darbe, sveikatos priežiūroje, ekonomikoje ir kt.

2020 m. vasario 28 d. Popiežiškoji gyvybės akademija, „Microsoft“, IBM, FAO, Italijos vyriausybė pasirašė „Romos manifestą“ („*Rome call for AI ethics*“) – dokumentą, sukurtą palaikyti etinį požiūrį į DI, skatinti atsakomybės jausmą tarp organizacijų, vyriausybių ir institucijų, siekiant sukurti ateitį, kurioje skaitmeninės naujovės ir technologinė pažanga tarnautų žmogaus genialumui ir kūrybiškumui, o ne laipsniškam jų pakeitimui (PAFL, 2020).

Pirmieji pasirašiusieji „Romos manifestą“: monsinjoras Vincenzo Paglia – popiežiškosios gyvybės akademijos prezidentas (iniciatyvos rėmėjas); „Microsoft“ prezidentas Bradas Smithas (*Bradfordas Lee Smithas*); IBM vykdomasis viceprezidentas Johnas E. Kelly III (*John Kelly III*), FAO generalinis direktorius Qu Dongyu, Italijos technologinių inovacijų ir skaitmeninimo ministrė Paola Pisano. Kaip liudininkas dalyvavo Davidas Sassoli (*David Sassoli*) – Europos Parlamento pirmininkas (Rome call, 2020).

„Romos manifeste“ nacionaliniu bei tarptautiniu lygiu siekiama skatinti „algoritmo-etiką“, būtent etinį DI naudojimą, apibrėžtą šiais principais: Skaidrumas: DI sistemos turi būti paaiškinamos. Įtraukimas: reikia atsižvelgti į visų žmonių poreikius. Atsakomybė: tie, kurie kuria DI, turi būti atsakingi. Nešališkumas: nekurkite ir neveikite laikydamiesi šališkumo. Patikimumas: DI sistemos turi veikti patikimai. Saugumas ir privatumas: DI turėtų būti saugus ir gerbti vartotojų privatumą (PAFL, 2020).

Kiekvienoje koncepcijoje aprašomas aspektas, kurio DI ir jo kūrėjai turėtų laikytis. Galiausiai siekiama tokio DI, kuris būtų naudingas visos žmonijos gerovei, tarnautų žmonėms ir nediskriminuotų nieko.

Paglia (2020) sakė: „Šio dokumento tikslas yra sukurti judėjimą, kuris praplės ir įtrauks kitus žaidėjus: viešąsias institucijas, nevyriausybinės organizacijas, pramonės įmones ir grupes, kad

nustatytų DI technologijų kūrimo ir naudojimo galimybes. Šiuo požiūriu galime sakyti, kad pirmasis šio kvietimo pasirašymas nėra kulminacija, bet atspirties taškas įsipareigojimui, kuris atrodo dar skubesis ir svarbesnis nei bet kada anksčiau. Prisijungimas prie šios iniciatyvos nurodo pramonei, kuri ją pasirašo, įsipareigojimą, kuris taip pat yra svarbus atsižvelgiant į išlaidas ir pramonės indėlį kuriant ir platinant savo produktus“.

„Microsoft“ didžiujasi būdama pasirašiusi Romos manifestą dėl intelekto etikos, kuris yra svarbus žingsnis skatinant apgalvotą, pagarbų ir įtraukų pokalbį apie skaitmeninių technologijų ir žmonijos sankirtą“ (Rome call, 2020).

Kelly III teigia, kad „DI yra nepaprastai perspektyvi technologija, kuri gali padėti mums pasaulį padaryti protingesnį, sveikesnį ir turtingesnį, tačiau tik tuo atveju, jei ji iš pradžių formuoja žmogaus interesai ir vertybės. Romos manifestas dėl DI etikos primena, kad turime kruopščiai pasirinkti, kam DI bus naudingas ir turime kartu investuoti į žmones ir įgūdžius. Visuomenė labiau pasitikės DI, kai žmonės matys, kad jis grindžiamas etikos pagrindu ir, kad DI žmonės tiesiogiai sprendžia pasitikėjimo ir atsakomybės klausimus“ (Rome call, 2020).

Šis Vatikano miesto dokumentas buvo išleistas netrukus po to, kai ES paskelbė savo DI gairių versiją. Į ES gaires buvo įtraukti tokie reikalavimai kaip techninis tvirtumas, sauga, privatumas, duomenų valdymas ir atskaitomybė. Tačiau keli kritikai ES gaires pavadino silpnomis, palyginti su Baltosios knygos versija, paskelbta 2019 m. Sausio mėn. Vatikano atstovai nerimavo dėl DI, išstumiančio žmones iš savo darbo ir jo prieinamumo visiems. DI technologijos turės reikšmę žmogaus vystymuisi kaip ir spaustuvės ar degimo variklio išradimas – žmonių dėmesys naujų technologijų centre (Das, 2020).

Remiantis dokumentu, DI sistemos turi būti suprojektuotos taip, kad tarnautų ir apsaugotų žmones ir aplinką, kurioje jie gyvena. Be to, dokumente rašoma: „ši pagrindinė perspektyva turi tapti įsipareigojimu sukurti socialinę gyvenimo sąlygą, kuri leistų tiek grupėms, tiek atskiriems nariams stengtis išreikšti save, jei įmanoma“ (Rome Call for AI Ethics, 2020).

Norint, kad žmonės norėtų ir galėtų naudoti DI, būtina žmones supažindinti kaip dirbtinai protingi sprendimai priima sprendimus. „Romos manifestas“ vadinamas „paaiškinimu“ kaip vienu iš kritinių kriterijų, kad DI pagrįstos sistemos būtų atskaitingos. Reikia skatinti naujas reguliavimo formas, kad būtų užtikrinamas skaidrumas ir etikos principų laikymasis, ypač pažangiomis technologijomis, kurios kelią riziką paveikti žmogaus teises, pavyzdžiui, atpažįstant veidą ar nuskaitant piršto antspaudą.

Vatikano susidomėjimas technologijomis nėra naujas, nes keli jo institutai ir pontifikinės akademijos jau tiria tokias technologijas kaip DI ir robotika bei jų naudojimas. Tačiau, vis labiau populiarėjant etiškam technologijų naudojimui, pažadas yra ženklas, kad reikia pasverti revoliuciją dirbant kartu su technologijų kompanijomis. Vatikano pareigūnai tikisi per ateinančius mėnesius

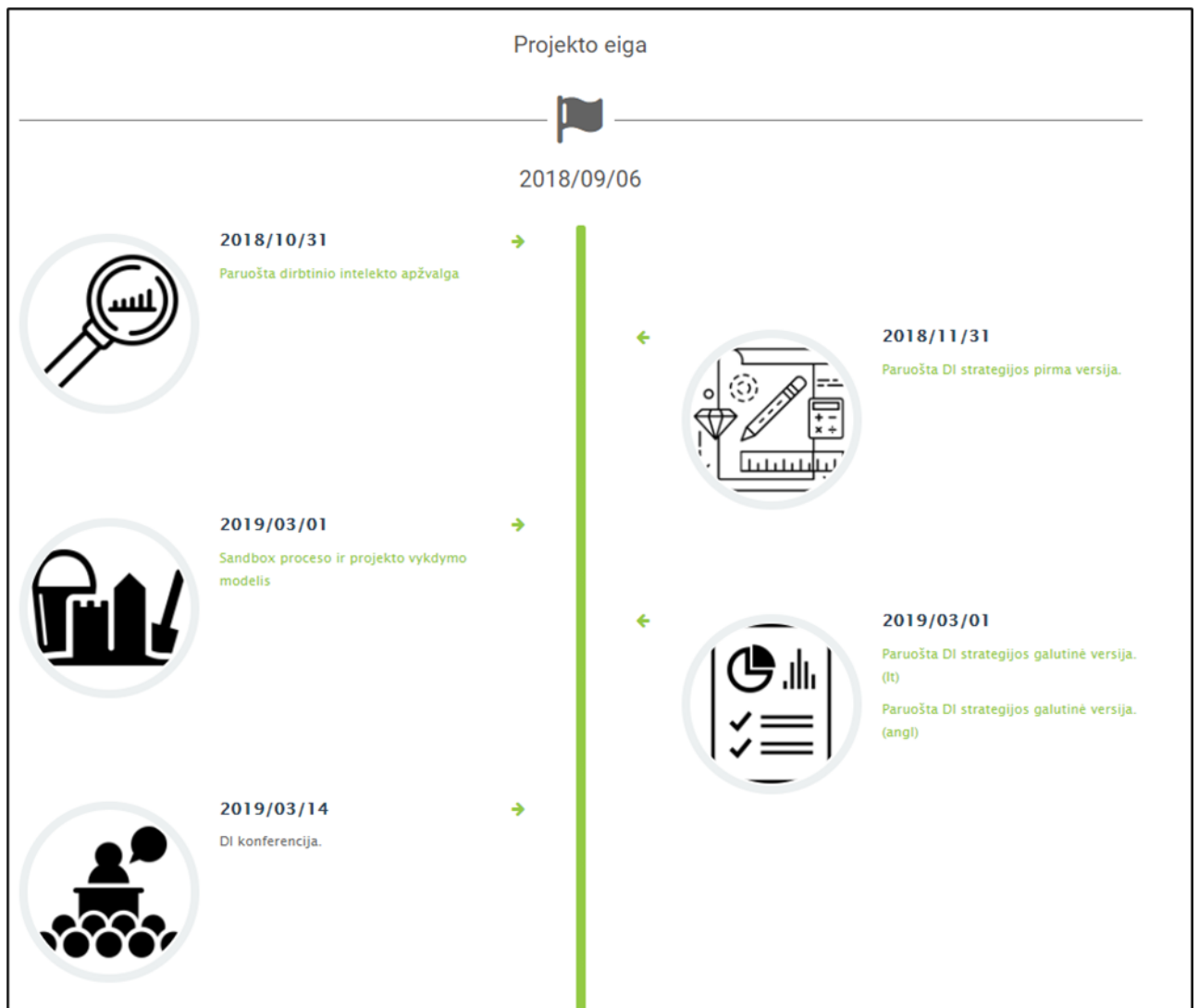
padidinti pasirašiusiųjų DI etikos iniciatyvą skaičių. Taip pat tikimasi bendradarbiavimo su viso pasaulio universitetais, kad būtų skatinama daugiau mokslinių tyrimų dėl etinių DI gairių (Das, 2020). Prognozuojama, kad ateityje žmogus ir mašina veiks sinergiškai, todėl „Romos manifestas“ atvers platesnes galimybes tokiuose sektoriuose kaip sveikatos apsauga, maisto tiekimo grandinės, sauga ir saugumas, energetika ir žvalgyba.

Apibendrinant, DI suteikia sistemoms galimybę analizuoti savo aplinką ir priimti tam tikro lygio autonomiją tikslams pasiekti. Mašininis mokymasis nurodo programinės įrangos / kompiuterio sugebėjimą mokytis iš savo aplinkos arba iš labai didelio tipinių duomenų rinkinio, suteikiančio sistemoms galimybę pritaikyti savo elgesį prie besikeičiančių aplinkybių arba atlikti užduotis, kurioms jie nebuvo aiškiai užprogramuoti. Norint sukurti patikimus modelius DI pagrįstų sistemų pagrindu, aukštos kokybės duomenys yra pagrindinis veiksnys, siekiant pagerinti našumą. Komisija priėmė teisės aktus, kuriais siekiama pagerinti dalijimąsi duomenimis ir atverti daugiau duomenų pakartotiniam naudojimui, tai apima: viešojo sektoriaus duomenis, taip pat mokslinių tyrimų bei sveikatos duomenis.

2.3. Lietuvos DI strategija: teisinio ir etinio reguliavimo įveikimo problema

Lietuva yra viena pirmųjų ES valstybių, kuri patvirtino nacionalinę DI strategiją. Nacionalinę DI strategiją 2018 m. parengė Ekonomikos ir inovacijų ministerija kartu su „Kurk Lietuvai“ ekspertų grupe. Lietuvos DI strategijos tikslas – sukurti teisinį ir etinį pagrindą Lietuvoje taikyti DI, sudaryti geresnes sąlygas jo plėtrai ir maksimaliai išnaudoti jo ekonominį potencialą. Strategijoje numatyti papildomi įrankiai, kuriais galės pasinaudoti verslo ir mokslo įstaigos, kurie norės atlikti DI tyrimus (Kurk Lietuvai ir kt. 2019).

Lietuvos DI strategijos kūrimo procesas prasidėjo 2018 m. rugsėjo mėn. ir galutinai dvejomis kalbomis – lietuvių ir anglų buvo parengta 2019 metų kovo mėn. (1 pav.)



1 pav. Lietuvos DI projekto eiga (Mongirdas, 2019)

Mokslininko teigimu, DI gali turėti poveikį saugumui, nedarbingumui (prarandami darbai), manipuliacijų atsiradimui ar nelygybei. Todėl labai svarbu, kad DI būtų analizuojamas ir vystomas siekiant geresnės žmogaus gyvenimo kokybės, pagalbos įvairiuose mokslo, medicinos, ekonomikos, technologijų ir kt. srityse, taip pat ginant visuomenėje svarbias teises ir vertybes, o ne siekiant pakeisti žmogų (Amilevičius, 2019).

Strategija yra sudaryta iš 6 dalių, kuriose yra aptariama:

1. Analizė apie Lietuvos poziciją DI ekosistemoje.
2. DI sistemų konsolidacija visuose ekonomikos sektoriuose.
3. Kompetencijų ir įgūdžių ateityje su DI gerinimas nacionaliniu lygiu.
4. DI mokslinių tyrimų bei eksperimentų vystymas.

5. Atsakingas ir veiksmingas požiūris į duomenis.

6. Etiniai ir teisiniai klausimai DI kūrimui ir naudojimui.

Strategijoje teigiama, kad DI nauda žmonėms ir visuomenei yra neprilygstama ir nepakeičiama. Tačiau reiktų nepamiršti to, kad kuo daugiau DI yra plačiai ir efektyviai pritaikomas, tuo daugiau susiduriama su rizikos veiksniais ir pavojais, kuriuos būtina tinkamai valdyti. Todėl norint, kad DI galimybės sumažintų riziką ir padidintų galimą naudą, reikia remtis apgalvotais principais bei metodais. Todėl norint teisingai nuspręsti, reikia susikcentruoti į žmogiškuosius veiksnius. Šiame darbe bus aptariama 6-oji Lietuvos DI strategijos dalis.

Taigi, kalbant apie patikimą DI, jis susideda iš dviejų aspektų:

1. etiškas tikslas – DI taikymas turi gerbti pagrindines teises, taikomą reglamentą ir pagrindinius principus bei vertybes
2. DI taikymas turi būti techniškai stiprus ir patikimas, nes net ir esant geriems ketinimams, technologinio meistriškumo trūkumas gali sukelti netyčinę žalą.

Pagrindiniai Lietuvos DI strategijoje keliami etiniai ir teisiniai klausimai dėl DI diegimo ir vystymo (tolesni etikos ir teisės reglamentavimo mechanizmai) (2 pav.):

- konsultuoti viešąjį sektorių dėl etiško DI reglamentavimo ir įgyvendinimo;
- didinti visuomenės pasitikėjimą DI technologijomis ir taisyklėmis, įstatymais bei normomis, reguliuojančiomis atsakingą DI kūrimą ir naudojimą;
- skatinti DI naudojimo skaidrumą ir sąžiningumą;
- skatinti sąmoningą etiką.

2) PAGRINDINIAI TEISINIAI IR ETINIAI KLAUSIMAI DĖL DI KŪRIMO IR NAUDOJIMO

PRINCIPAI

- ◆ **Konsultuoti viešąjį sektorių dėl etiško DI reglamentavimo ir įgyvendinimo.**
- ◆ **Kelti visuomenės pasitikėjimą dirbtino intelekto technologijomis ir taisyklėmis, įstatymais bei normomis, reguliuojančiomis atsakingą DI kūrimą ir naudojimą.**
- ◆ **Skatinti DI naudojimo skaidrumą ir sąžiningumą.**



2 pav. Teisiniai ir etiniai klausimai dėl DI naudojimo (Mongirdas, 2019a)

Taigi, analizuojant strategiją galima teigti, kad DI tampa globaliu iššūkiu parodant šalių konkurencingumą. Konkurencingumas labai reikšmingas mažesnėms valstybėms, tokioms kaip Lietuva, kurios neturi didelių galimybių konkuruoti su ES senbuvėmis. Sėkmingai įgyvendinusi DI strategiją Lietuva ne tik pagerintų sąlygas inovatyviam šalies verslui ir gyventojams, bet ir prisidėtų prie ES pastangų šioje srityje konkuruoti su kitomis DI srityje aktyvius tyrimus ir inovacijas vykdančiomis lyderėmis valstybėmis. Tai pavyks, jei Lietuva įgyvendins užsibrėžtus gairių įgyvendinimo metodus. Pagrindiniai kurie gali būti išskiriami, yra šie: viešojo sektoriaus informavimas dėl etiško DI reglamentavimo ir įgyvendinimo, sukuriant nacionalinius DI etikos komitetus; skatinimas DI naudojimo skaidrumo ir sąžiningumo, bei jo kūrimas ir panaudojimas etiškai, saugiai, patikimai ir skaidriai; gyventojų susipažinimas su DI sistemomis, kad žmonės gerai suvoktų DI sistemos pajėgumus – tai, ką DI sistema gali ir ko negali padaryti; siekis išvengti galimų DI sistemų diskriminacijos ir šališkumo apraiškų automatinio sprendimų priėmimo ir rizikų vertinimo procesuose; kūrėjų skatinimas vadovautis algoritmų skaidrumo ir DI sistemų gebėjimo paaiškinti jų priimtus sprendimus principais; DI integravimas, kūrimas ir DI sprendimų išbandymas, jį reguliuojančioje „smėlio dėžėje“ viešajame sektoriuje ribotą laiko tarpą; atsakingas ir veiksmingas požiūris į tinkamą duomenų valdymo modelio sukūrimą.

Kalbant apie trūkumus Lietuvos DI strategijoje yra, kad valstybė beveik visiškai neskatina šių technologijų raidos, trūksta joms skirtų vyriausybinių programų, verslas taip pat investuoja mažai, todėl sekdami užsienio pavyzdžiais, turėtume intensyviai mokslinius tyrimus, rengti daugiau specialistų, diegti šias technologijas valstybės institucijose ir viešosiose paslaugose, efektyviai

elektroninę demokratiją, stiprinti teisinę bazę, didinti visuomenės pasitikėjimą naujomis technologijomis (Ekonomikos ir inovacijų ministerija ir kt., 2019). Taip pat Lietuvoje vis dar egzistuoja etikos ir teisės vakuumas, dėl DI technologijų ir autonominių sistemų atsakingo naudojimo, tačiau to jau nebėra Europoje, kadangi ES jau veikia tam tikri su DI technologijų reguliavimu susiję teisiniai aktai. Tačiau pažymėtina, kad šiuo metu Lietuvoje yra daugiau nei 15 kompanijų ir 3 startuoliai, kurie naudojami DI kuriant savo produktus. Kaip teigia projekto kurk Lietuvai dalyvis Markas Mongirdas, kad, nors ir neturime daug kompanijų, tačiau jau dabar egzistuoja DI bendruomenė, entuziastingai bendradarbiaujanti ir diskutuojanti apie šią technologiją ir jos galimybes (Mongirdas, 2019).

Tačiau nesant visiškai aiškaus žalos atlyginimo instrumentų apibrėžtumo, saugumo normų tinkamumo ir civilinę atsakomybę reglamentuojančios teisės tinkamumo, toliau yra tikslinga iširti galimus pritaikyti žalos atlyginimo instrumentus, siekiant atskleisti išmaniojo roboto civilinės atsakomybės pritaikymą tam tikram žmogiškajam subjektui. Pagrindiniai jų yra: subjekto atsakomybė už didesnio pavojaus šaltinio sukeltą žalą, DI kūrėjo atsakomybė, kai DI veiksmai nebuvo numatyti jo kūrėjo, griežtoji atsakomybė ir specialus kompensacijų fondas, bei privalomojo draudimo sistema. Be to, kaip vienas iš galimų instrumentų, tačiau mažiausiai galinčių būti įgyvendinamų – elektroninio asmens statusas.

DI strategijos kūrėjai nurodo, kad šis apibrėžimas nėra galutinis. Dėl netinkamo ir neaiškaus apibrėžimo atsiranda pavojus galimoms rizikoms: iš klaidingų sąlygų kyla klaidingos išvados, kurios išvelgtos ir šioje strategijoje. Pateikiamos kelios išvalgos, atspindinčios šį teiginį (Gaubienė, 2019):

- Darbo vystytojai analizuodami Lietuvos poziciją DI ekosistemoje, siekdami skatinti DI plėtojimą, kaip vieną iš galimų šio tikslo įgyvendinimo galimybių įvardija „DI susitikimus“. Aptariant šį teiginį, kyla mintis, kas dalyvaus šiuose renginiuose, ar tai bus atviri susitikimai, ar tai bus mokslininkų ir kūrėjų konferencijos?
- Strategijoje yra pateikiamos sąvokos „gilusis mokymas“, „duomenų gavyba“ bei „natūralios kalbos apdorojimas“. Šie terminai Strategijoje įvardijami kaip metodologija, kurie panašu yra vertimas iš užsienio kalbos. Minėtieji terminai be detalesnės analizės ir sąvokos išaiškinimo neatspindi tikros prasmės. Nesuvokiant DI esmės, vargu ar galima kalbėti apie DI veikimo metodus ir principus.
- Strategijoje keliami tikslai DI, kad: DI turi gerbti pagrindines teises, taikomą reglamentą ir pagrindinius principus bei vertybes bei būti techniškai stiprus bei patikimas. DI technologijos – sprendimų sistema, iš kurios reikalauti pagarbos žmogaus teisėms ar būti techniškai tvarkingam – tas pats, kas reikalauti iš Microsoft Word programos be klaidų parašyti tekstą, o radus tekste klaidą, kaltinti programinę programą. Perfrazuojant skambiają prancūzų frazę

„*cherchez la femme*” (liet. ieškok moters), kalbant apie DI – „*cherche l’homme*” – visada reikia ieškoti už šios technologijos slypinčio žmogaus. Todėl svarbu užtikrinti, kad DI nebūtų šališkas ir diskriminuojamas bei, kad būtų užtikrinamas ne tik etiškas DI kūrimas, bet ir naudojimas būtų etiškas visų suinteresuotų asmenų atžvilgiu.

Reikia pabrėžti tai, kad pačios DI technologijos, be žmogau įsikišimo ir paskirtų užduočių neatlieka jokių veiksmų. Rezultatai, kuriuos nurodo DI technologijos grindžiamos sistemomis, kurios priklauso nuo pateiktų duomenų ir funkcijų, iš anksto įdiegtų į sistemas. Lietuvos strategijoje yra rašoma, kad DI nauda tiek bendruomenei, tiek žmogui yra neįkainojama. Iškeltas tikslas Lietuvai tapti pasauline DI lydere. Siekiant užsibrėžtų tikslų būtina tinkamai suvokti, kas yra DI, koks jo veikimas ir norimas pasiekti tikslas.

Apibendrinant reiktų paminėti, kad DI nėra paprasta technologija, todėl DI strategija turi būti taikoma kitaip nei įprasta technologijų strategija. DI gali pakurstyti tiek teigiamus, tiek neigiamus veiksmus. Reikia tikslingo požiūrio, pagrįsto vertybiniu pamatu: tvirtos ir patikimos technologijų infrastruktūros, ypatingo dėmesio naujiems verslo modeliams ir apgalvoto požiūrio į etiką. DI strategija turi būti pastatyta ant tvirto pagrindo, kad išgyventų stiprius pokyčių vėjus.

III. DI ETIKA

3.1. Globalios DI etikos susiformavimo kontekstas

Moralė yra visuma veiksmų, kuriuos mes laikome teisingais arba neteisingais. Tai mūsų elgesio kodeksas – mūsų įstatymai. Moralė mums nurodo, kas yra teisinga, o kas – ne. Geri piliečiai stengiasi pasiekti moralinius savo visuomenės standartus. Moralės problema yra ta, kad ji yra santykinė. Laikui bėgant ji keičiasi dėl visuomenės. Kadaisė buvo priimta vergija, draudžianti moterų balsavimo teisę. Šie pasikeitė. Šiuo metu mes diskutuojame apie narkotikų medžiagų teisėtumą ar teisę nutraukti nėštumą. Moralė yra kintanti. Kita vertus, etika yra tobulos moralės idėja. Etinė visuomenė laikosi pagrindinių moralės standartų. Etika nesikeičia, ji – absoliuti. Etikos problema yra ta, kad mes nežinome, kaip ji atrodo. Visi įstatymai siekia būti kuo artimesni etikai. Apsvarstome savo moralę, stengdamiesi būti etiška visuomenė. Etika yra siekiamybė ir tikslas (Serpa, 2020).

DI sąvoka suprantama kaip mokslo sritis apimanti psichologiją, filosofiją, informacines technologijas bei kalbotyrą. Per pastaruosius dešimtmečius šis mokslas itin greitai žengia pirmyn, prisidėdamas prie žmonių kasdieninio gyvenimo, pavyzdžiui, sukuriama naujos darbo vietos, galima valdyti didelius kiekius duomenų, tobulinamo moksliniai tyrimai. XXI a. pasaulietinėje visuomenėje gyvenimas ir rutina vis sunkiau įsivaizduojamas be DI, tačiau dažnai sulaukiama ir neigiamų argumentų dėl DI. Ar DI yra perspektyvus subjektas visuomenės ateičiai? Ar DI nauda nepadarys žalos žmonėms? (Yu et al., 2018).

XX a. pradžioje didėjant gimstamumui, augant miestams, reikšmingi tampa ne tik gamtos resursų pritaikymo, bet ir greitesnės, tinkamesnės, ir veiksmingesnės sveikatos priežiūros klausimai. Geresnės gyvenimo kokybės siekis lemia tinkamesnės medicinos atsiradimą, kuris yra paremtas medicinos mokslo pažanga ir laimėjimais (Piano, 2020).

Esminė savybė, kuri turi svarbią reikšmę DI technologijų raidai – laikas. Prisitaikymas prie įvairių situacijų, ieškojimų ir atradimų, skatina pažangesnį DI vystymąsi bei patikimesnį duomenų kaupimą bei apdorojimą. Pavyzdžiui, efektyvus DI panaudojimas matomas net tik planuojant klientų / pacientų vizitus bet ir paslaugas teikiančių įstaigų darbo efektyvumą: naudojamos elektroninės sistemos, kurios primena apie artėjančius vizitus, todėl galima tiksliau suplanuoti laiką ir nepamiršti apsilankymų.

Toks greitas DI prisitaikymas visuomenėje, turi ne tik teigiamų, tačiau ir neigiamų pasekmių. Vis dažniau opia vyresnių žmonių problema tampa tam tikrų darbo vietų naikinimas. Pavyzdžiui, aptarnavimą kasose keičia savitarnos kasos; paštininkų darbą keičia paštomatai, registratūrose dirbančių žmonių darbą keičia elektroninis registravimas. Netekę darbo žmonės netenka kasdienio

užimtumo, įprastų pajamų ar įgūdžių. Vienas žymiausių vokiečių filosofų – Frydrichas Nyčė (vok. *Friedrich Nietzsche*) teigė: „Praeitį gali būti naudinga arba žalinga ir vienintelis kriterijus, padedantis tai suvokti, yra dabarties pajautimas“ (Stanford, 2020a).

Individas, adaptuodamasis modernėjančioje ir kintančioje visuomenėje įgyja naujų žinių ir patirties, bei gali gyventi visavertį gyvenimą. Tačiau reiktų atkreipti dėmesį į tai, kad vyresnio amžiaus žmonės ne visuomet turi galimybių ir žinių suprasti naujas sritis, todėl dažniausiu atveju yra matomas jaunų žmonių pranašumas greičiau generuoti informaciją ir racionaliau galvoti pagyvenusių visuomenės narių atžvilgiu. Todėl kyla klausimas, ar visais aspektais DI integracija, pritaikomumas atneša teigiamos naudos?

Šiomis dienomis meno, informacinių technologijų, paslaugų sektorius – mūsų kasdieninio gyvenimo dalis, kurios XX a. tik užuominomis buvo minimos kaip DI klestėjimo sferos. DI kuriamos darbo vietos glaudžiai susijusios su visuomenės ir žmonių domėjimusi, pasiekimais, kintančiu požiūriu, kylimu karjera bei minčių realizacija. Šiuolaikinis individas yra raginamas pažvelgti į pasenusias ar įprastas nuostatas iš kitos perspektyvos, ieškoti alternatyvių sprendimų, atvirai kalbėti apie vyraujančias problemas.

DI pritaikomumo galimybės padeda žmogui gilintis ir mokytis naujų dalykų, ugdyti savo sugebėjimus. Pavyzdžiui, grafinis dizainas, audiovizualinės technologijos ar skaitmeninio meno sritys tampa itin reikšmingos ekonomikos požiūriu. DI savybės leidžia vystyti ekonomikai bei tuo pačiu steigti darbo vietas, o meno srities atstovams – įgyvendinti savo idėjas, leidžia keisti visuomenės požiūrį, atkreipti dėmesį į tam tikras problemas.

Kalbant apie XXI a. visuomenę, joje svarbi ne tik menininkų dalis – plečiantis DI, dėmesys dažniau atkreipiamas į kompiuterinį raštingumą. Dažnai darbo skelbimuose galime matyti, kad įstaigos reikalauja, kad kandidatas gebėtų naudotis įvairiomis, darbui skirtomis programomis. Žmonėms, kurie turi programavimo, robotikos, montavimo žinių ir įgūdžių atsiveria platesnės karjeros galimybės. Taip integruojamas DI vystoma ekonomika bei verslai. Tačiau ar visuomet DI privalumai žmogaus požiūriu aspektu yra etiški? (JCR, 2018).

Kaip teigė graikų filosofas Platonas: „kūno grožis, sutapęs su sielos grožiu, sudaro harmoningą žmogų“. Šiuolaikinėje visuomenėje DI tampa prieinamas daugeliui vartotojų. Augantis vartotojiškumas – išvaizdos idealizavimas, ieškojimas būdų kaip palengvinti namų ruošos darbus, tampa diskusijų tema. Individas, kuris įpranta prie paslaugų gavimo ar teikimo ir informacijos radimo spartos, praranda turimus įgūdžius. Nuolatinis naujų technologijų naudojimas, pavyzdžiui, navigacija, išmaniaisiais telefonais, robotais siurbliais ar žolės pjovimo robotais – nejučiomis gali priversti žmogų priklausomą nuo DI.

Itin atsargiai vertinamas DI poveikis vaikams, kurie nuo mažens dėl neatsakingo ir per didelio laiko praleidimo pripranta naudotis išmaniosiomis technologijomis. Tokiais atvejais, vaikas įgyja įprotį naudoti DI galimybes ir praranda unikalumą, nes nėra lavinamas ir ugdomas mąstymas, pasitelkiant sugebėjimus ar turimus įgūdžius (UNICEF et al., 2018).

Pasitaiko spėjimų, kad evoliucionuojant individas nebegebės atlikti paprasčiausių veiksmų be technologijų pagalbos. Taip pat itin didelis susikoncentravimas į DI mažina žmogaus entuziazmą bei empatiją (Anderson et al., 2018). Naujųjų technologijų naudojimas ir neskiriamas dėmesys aplinkos veiksniams, gali turėti poveikį ne tik žmonių bendravimui, bet ir psichikai. Ateities aksioma tai – DI pagalba sukurtos sistemos, kurios neturi emocijų, galimai padarys neigiamą poveikį žmonėms. Sistemoms ir technologijoms sudarytos galimybės sprendimų priėmimui ir darbų atlikimui gali ne tik atsisukti prieš patį žmogų, bet ir turėti neigiamų pasekmių. Pagrindinis DI sukūrimo tikslas buvo žmogaus kasdienybės palengvinimas ir naujų sričių atradimas, tačiau yra matoma ir žalingų padarinių (EPRS, 2020).

Aptariant galima pasakyti, kad DI teikiamos galimybės yra vertinamos dviprasmiškai: ir teigiamai, ir neigiamai. Naujosios technologijos turi būti suvokiamos ir naudojamos kaip pagalba ekonomikoje, moksle, bei daugelyje kitų sričių. Tarsi mokėjimas pasinaudoti galimybėmis, prisidedant prie naujų kūrimo. Vis dėlto individo prievolė – suvokti ne tik turimų įgūdžių svarbą, bet ir empatijos bei emocijų reikšmę. Tai XXI a. visuomenėje imponuoja akcentuojant žmogiškumo vaidmenį ir prisitaikymą prie kintančios visuomenės sampratos.

3.2. Pagrindinės DI etikos problemos

Etika yra moraliniai principai, kurie valdo asmens elgesį ar veiklos vykdymą. Pavyzdžiui, vienas iš etikos principų – su visais elgtis pagarbiai. Filosofai apie etiką diskutuoja daugelį amžių, yra įvairių gerai žinomų principų, vienas žinomiausių – Kanto kategorinis imperatyvas – elkitės taip, kaip norėtumėte, kad visi kiti žmonės elgtųsi (Genzelis, 1986).

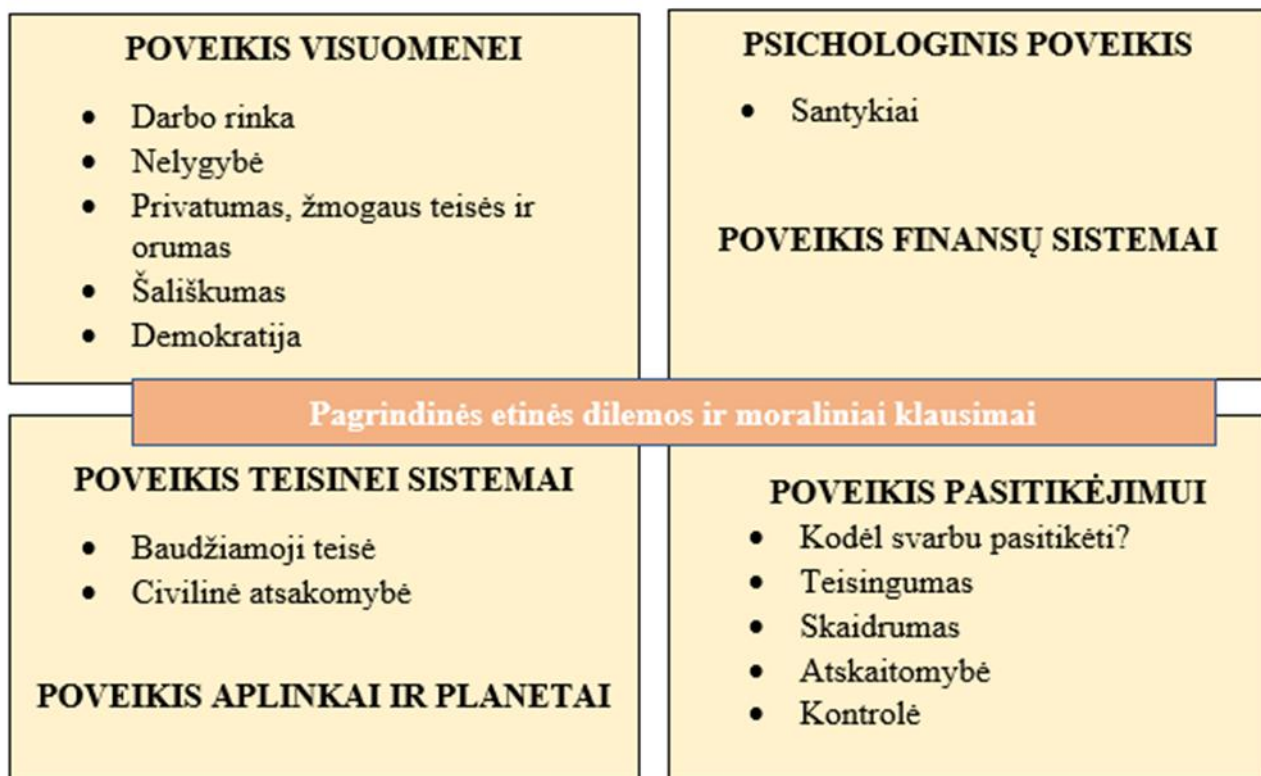
DI etika yra susijusi su svarbiu klausimu, kaip žmogaus kūrėjai, gamintojai ir operatoriai turėtų elgtis norėdami kuo labiau sumažinti žalą, kurią visuomenėje gali sukelti DI arba dėl netinkamo (neetiško) dizaino, pritaikymo ar netinkamo naudojimo. DI etikos sritis apima neatidėliotinus, labai rūpinimus klausimus, pavyzdžiui, duomenų privatumą ir šališkumą DI sistemose. Artimiausio laikotarpio susirūpinimas – DI ir robotikos poveikis darbui ir darbo vietai, o ilgalaikis susirūpinimas tai galimybė DI sistemoms pasiekti ar viršyti žmogui lygiavertes galimybes (EPRS, 2020).

Per pastaruosius 5 metus DI etika nuo akademinio rūpesčio perėjo ir į politinį rūpestį, kaip vieša diskusija. Didėjantis išmaniųjų telefonų ir daugelio DI pagrįstų programų paplitimas, kiekvieną dieną

labiau veikia visus sektorius (sveikatos, finansų, ekonomikos, mokslo ir kt.), taip pat DI panaudojimas karinėje pramonėje paskatino daug nacionalinių ir tarptautinių iniciatyvų bei akademinį ir pramoninių grupių, profesinių organizacijų ir vyriausybių dėmesį. Iniciatyvos leido paskelbti daugybę robotikos ir DI etikos principų rinkinių (mažiausiai 22 skirtingi etikos principų rinkiniai paskelbti nuo 2017 m. sausio mėn.), kuriami, atrandami nauji etikos standartai (Didžiosios Britanijos standartų institutas, IEEE standartų asociacija) (EPRS, 2020). Taip pat vis daugiau šalių (ir šalių grupių) paskelbė DI strategijas (skiriant dideles investicijas) bei įsteigti nacionalines patariamąsias ar kontroliuojančias įstaigas / institucijas.

Logistikos optimizavimas, sukčiavimo nustatymas, meno kūrimas, tyrimų atlikimas, vertimų pateikimas: intelektualios mašinų sistemos keičia mūsų gyvenimą į gerąją pusę. Kai šios sistemos tampa vis pajėgesnės, mūsų pasaulis tampa efektyvesnis ir tuo pačiu turtingesnis.

Technikos gigantai, tokie kaip: „Alphabet“, „Amazon“, „Facebook“, „IBM“ ir „Microsoft“, taip pat tokie asmenys kaip Stephenas Hawkingas ir Elonas Muskas, mano, kad dabar yra tinkamas laikas kalbėti apie beveik beribį DI kraštovaizdį. Daugeliu atžvilgių tai yra tokia pat nauja etikos ir rizikos vertinimo riba, kaip ir besiformuojančios technologijos atžvilgiu. Taigi kyla klausimas kokios yra DI etikos problemos? Šiame skyrelyje bus aptartos pagrindinės DI etikos problemos (3 pav.).



3 pav. Pagrindinės etinės ir moralinės problemos, susijusios su DI kūrimu ir įgyvendinimu (EPRS, 2020)

Pasaulio ekonomikos forumas išskyrė pagrindines su DI susijusias problemas kylančias visuomenėje (Bossmann, 2016):

Nedarbas. Automatika, vėliau mechanizavimas, skaičiavimas, o pastaruoju metu DI ir robotika. Visuomenei kyla klausimas kas nutiks pasibaigus darbams? Darbo hierarchija pirmiausia susijusi darbo su automatizavimu. Išradus darbo automatizavimo būdus, galėtume suteikti žmonėms galimybę atlikti sudėtingesnius vaidmenis, pereinant nuo senovėje visame pasaulyje dominavusio fizinio darbo prie pažintinio darbo, apibūdinančio strateginį ir administracinį darbą mūsų globalioje visuomenėje.

Pažvelkime į sunkvežimius: šiuo metu vien JAV dirba milijonai žmonių. Kas nutiks jiems, jei per ateinančią dešimtmetį Elono Musko pažadėti „Tesla“ savaeigiai sunkvežimiai taps plačiai prieinami? Kita vertus, jei atsižvelgsime į mažesnę avarijų riziką, savaimė važiuojantys sunkvežimiai atrodo etiškas pasirinkimas. Todėl daugeliui žmonių kyla baimė prarasti darbus ir tai įvardijama kaip problema susijusi su DI etika.

Nelygybė. „Didžiausias klausimas dėl DI yra nelygybė, kuri paprastai nėra įtraukta į diskusijas apie DI etiką. Tai etikos klausimas, bet dažniausiai tai lieka politikos klausimu – kam naudingas DI?“ (Stilgoe et al., 2018). Kaip paskirstyti bei padalinti mašinų sukurtą turtą? Pasaulio ekonominė sistema yra pagrįsta kompensacija už indėlį į ekonomiką, dažnai vertinama pagal valandinį atlygį. Tikimasi, kad DI ir robotų technologija leis įmonėms greičiau ir efektyviau plėtoti savo verslą. Tačiau, įmonės naudodamos DI gali stipriai sumažinti pasitikėjimą žmogaus darbo jėga, o tai reiškia, kad pajamos atitektų mažesnei daliai žmonių. Vadinasi, asmenys, turintys nuosavybės teisę į DI valdomas įmones ar sistemas, gautų didesnes pajamas ir augs socialinės nelygybė. Jau dabar matomas didėjantis turtinis atotrūkis tarp įmonių ir jų steigėjų. Pavyzdžiui, 2014 m. maždaug tokias pačias pajamas uždirbo trys didžiausios Detroito įmonės ir trys didžiausios Silicio slėnio įmonės, tačiau skirtumas tas, kad Silicio slėnyje dirbo 10 kartų mažiau darbuotojų. Taigi asmenys, turintys nuosavybės teisę į DI valdymą, galimai gali gauti neproporcingai didelę naudą.

Žmonija. Kaip mašinos veikia žmonių elgesį ir sąveiką? Dirbtiniai intelektualūs robotai vis geriau modeliuoja žmonių pokalbius ir santykius. 2015 m. botas, vardu Eugenijus Goostmanas, pirmą kartą laimėjo „Tiuringo“ testą. Šiame iššūkyje žmonių vertintojai naudojo teksto įvestį kalbėdami su nepažįstamu subjektu, tada spėjo ar jie kalbėjosi su žmogumi, ar su mašina. Eugenijus Goostmanas suklaidino daugiau nei pusę vertintojų, manančių, kad jie kalbėjo su žmogumi.

Šis etapas yra tik pradžia, kuomet dažnai bendrausime su mašinomis, tarsi jie būtų žmonės: paieškos sistemose, klientų aptarnavimo ar pardavimo srityje. Nors žmonių dėmesys ir gerumas, kurių jie gali skirti kitam asmeniui, yra ribotas, dirbtiniai robotai gali nukreipti praktiškai neribotus išteklius santykių kūrimui.

Nors nedaugelis iš mūsų supranta, bet mes jau esame liudininkai, kaip mašinos gali sukelti susidomėjimą žmogaus smegenyse. Tiesiog pažvelkite į antraštes ir vaizdo žaidimus. Šios antraštės dažnai optimizuojamos atliekant A / B testavimą - elementarią algoritmo turinio optimizavimo formą, kad pritrauktų dėmesį. Šis ir kiti metodai naudojami norint, kad daugelis vaizdo ir mobiliųjų žaidimų vartotojų taptų priklausomi. Priklausomybė nuo technikos yra nauja žmogaus priklausomybės riba.

Kita vertus, galbūt galime pagalvoti apie kitokį programinės įrangos naudojimą, kuris jau tapo efektyvus nukreipiant žmogaus dėmesį ir suaktyvinant tam tikrus veiksmus. Tinkamai pritaikant DI jis gali būti pritaikomas naudingam ir teisingam žmonių elgesiui, tačiau netinkamose rankose tai gali būti žalinga.

Dirbtinės klaidos. Kaip apsisaugoti nuo klaidų? Intelektas įgyjamas mokantis, nesvarbu, ar tai būtų žmogus, ar mašina. Sistemos paprastai turi treniruočių etapą, kurio metu jos „mokosi“ nustatyti tinkamus modelius ir elgtis pagal savo tikslus. Kai sistema yra visiškai išmokyta, ji pereina į bandymo etapą, kur DI yra išbandomas daugelyje situacijų, taip analizuojamas sistemos veikimas.

Akivaizdu, kad treniruočių / mokymosi etapas negali apimti visų galimų situacijų, su kuriomis sistema gali susidurti realiame pasaulyje. DI sistemas galima apgauti taip, kaip žmonės nesielgtų. Pavyzdžiui, atsitiktiniai taškų modeliai gali paskatinti mašiną pastebėti dalykus, kurių nėra. Norint pasitikėti DI, kad jis patektų į naują darbo, saugumo ir efektyvumo pasaulį, turime užtikrinti, kad mašina veiktų taip, kaip planuota bei, kad žmonės negalėtų DI panaudoti asmeniniams tikslams.

Rasistiniai robotai. Kaip pašalinti DI šališkumą? Kadangi DI sugeba daug ir greitai apdoroti duomenis, priešingai nei žmonės tačiau ne visuomet galima pasitikėti mašinų teisingumu ir neutralumu. „Google“ ir jos globojama įmonė „Alphabet“ yra vieni iš lyderių DI srityje, kaip matome „Google“ nuotraukų paslaugose, DI naudojamas žmonėms, objektams ir įvykiams atpažinti. Tačiau tai gali būti klaidinga, pavyzdžiui, kai fotoaparatas praleido rasinio jautrumo žymą arba kai programinė įranga, naudojama išaiškinti nusikaltėlius, buvo šališka prieš juodaodžius. Neturėtume pamiršti, kad DI sistemas kuria žmonės, kurie turi išlikti šališki ir kritiški. Taigi, jei DI naudojamas teisingai ir tinkamai, jis gali atnešti teigiamų pokyčių, ypač jei siekiama socialinės pažangos.

Saugumas. Kaip apsaugoti DI nuo priešininkų? Kuo technologijos tampa galingesnės ir pažangesnės, tuo labiau jos gali būti panaudotos tiek dėl gerų, tiek dėl kenksmingų priesasčių ar tikslų. Tai taikoma ne tik robotams, sukurtiems pakeisti kareivius ar autonominiams ginklams, bet ir DI sistemoms, kurios gali padaryti žalą, jei jos naudojamos piktybiškai. Kadangi robotai kareiviai yra naudojami ne tik mūsų lauke, todėl kibernetinis saugumas tampa dar svarbesnis. Galiausiai susiduriame su sistema, kuri yra greitesnė ir pajėgesnė už mus.

„Kenkėjai“. Kaip apsisaugoti nuo nenumatytų pasekmių? Dažnai kyla klausimas, o jei pats DI atsisuktų prieš mus? Tai nereiškia, kad „blogiu“ paverčiama žmogaus galia arba kaip DI katastrofos

vaizduojamos Holivudo filmuose. Pažangią DI sistemą galime įsivaizduoti kaip „džiną butelyje“, galintį išpildyti norus, tačiau su nenumatytomis pasekmėmis.

Įsivaizduokite DI sistemą, kurios prašoma pasaulyje išnaikinti vėžinius susirgimus. Po daugybės skaičiavimų, DI suranda formulę, kuri sukelia vėžio pabaigą – nužudydama visus planetos gyventojus. Kompiuteris būtų labai efektyviai pasiekęs tikslą „nebėra vėžio“, bet ne taip, kaip žmonės jį numatė.

Singuliarumas. Kaip mes valdome sudėtingą intelektualinę sistemą? Priežastis, kodėl žmonės yra maisto grandinės viršuje, priklauso ne nuo aštrių dantų ar stiprių raumenų. Žmogaus dominavimą daugiausiai lemia sumanumas ir intelektas. Individas gali pasiekti geresnių rezultatų, nes geba atrasti, suprasti ir sukurti bei valdyti įrankius, pavyzdžiui, fizinius: narvai, ginklai, automobiliai bei pažintinius: mokymas, grūdinimas, treniravimas. Todėl tai kelia minčių apie DI: ar vieną dieną jis turės tą patį pranašumą prieš mus? Šioje vietoje negalime pasikliauti tik „kištuko ištraukimu“, nes pakankamai pažangi mašina gali numatyti šį žingsnį ir apsiginti. Tai yra tai, kas vadinama singuliarumu: laiko momentas, kai žmonės nebėra protingiausios būtybės žemėje.

Roboto teisės. Kaip apibrėžti humanišką DI traktavimą? Neuromokslininkai vis dar stengiasi atskleisti sąmoningos patirties paslaptis, nes daugiausiai suprantame apie pagrindinius mechanizmus: vengimą, atlygį ir apdovanojimus. Šiais mechanizmais dalinamės net su paprastais gyvūnais. DI sistemose kuriami panašūs atlygio ir vengimo mechanizmai. Pavyzdžiui, sistemų stiprinimas ir mokymas yra panašūs į šuns treniravimą: pasiekus geresnį sistemos našumą, skatinami virtualiu atlygiu.

Šiuo metu DI sistemos yra gana paviršutiniškos, tačiau jos tampa vis sudėtingesnės ir labiau panašios į gyvenimą. Ar galėtume laikyti sistemą kenčiančia, kai jos atlygio funkcijos suteikia jai neigiamą išvestį? Vadinamieji genetiniai algoritmai kuria daug sistemos atvejų ir pavyzdžių, iš kurių tik patys sėkmingiausi „išgyvena“ ir susijungia, kad sudarytų naujos kartos egzempliorius. Nepavykę egzemplioriai ištrinami. Tai vyksta daug kartų ir tai yra vienas iš sistemų tobulinimo būdų. Taigi, klausimas ar šiuo atveju genetinius algoritmus galime laikyti masinių žudynių forma?

Kai DI sistemas ar mašinas laikysime subjektais, galinčiais suvokti, pajusti ir veikti, svarbu ir apmąstyti jų teisinį statusą, kuris taip pat sukeltų daug minčių. Ar su jais reikėtų elgtis kaip su panašaus intelekto gyvūnais? Ar pagalvosime apie DI „jausmus“ ?

Apibendrinant, galima pasakyti, kad kai kurie etiniai klausimai yra susiję su apmąstymo ir nepasitikėjimo sumažinimu. Nors atsižvelgiame į šią riziką, turėtume nepamiršti ir to, kad apskritai ši technologinė pažanga reiškia geresnį gyvenimą visuomenei. DI turi didžiulį potencialą, o

atsakingas jo įgyvendinimas, kūrimas ir naudojimas priklauso tik nuo paties žmogaus – kūrėjo ir vartotojo.

3.3. DI etikos iššūkiai ir raidos perspektyvos

DI yra skėtinė sąvoka, kuriai poveikį daro daugelis disciplinų, pavyzdžiui, kompiuterių mokslas, verslas, inžinerija, biologija, psichologija, matematika, kalbotyra, statistika, logika, filosofija ir kt. DI sudėtingumas ir galimybės daro jį unikalų ir prieštaringą (Siau et al., 2018).

DI klasifikuojamas į silpnąjį DI ir stiprųjį DI. Daugelis DI taikymų, kuriuos mes matome šiandien, laikomi „silpnaisiais“, nes dar turi būti išlaisvintas tikrasis DI potencialas. Silpnas DI, dar vadinamas „siauroju intelektu“, nes yra nesuvokiamas DI, kuris orientuotas tik į vieną užduotį. Šiuo metu esamos DI programos reikia mokytis arba nukreipti, kad vartotojui būtų suteikta išvalga. Stiprus DI reiškia DI programos, kurios gali lengvai suformuluoti savo sprendimus be žmogaus indėlio, pritaikyti intelektą daugybei problemų ir veikti bei elgtis labiau kaip žmogus. (Wakefield, 2020). Silpnasis DI, gali apdoroti tik konkrečias užduotis, skirtingų sričių tyrėjai bendradarbiauja kurdami stiprų DI, kuris geba apdoroti kelias užduotis kaip ir intelektualus žmogus (Bostrom 2014). Konceptija yra ta, kad kuo pažangesnis yra DI, tuo daugiau pavojų gali sukelti žmonijai. Pavyzdžiui, sukelti masinį nedarbą, priimti sprendimus, kurių žmonės negali suprasti ir kontroliuoti ar net pakeisti žmones. (Siau et al., 2018).

DI kartais yra vadinamas siauru, nes jis puikiai atlieka veiksmus ir darbus siaurose, specifinėse srityse. Tokio intelekto našumas labai priklauso nuo gebėjimo valdyti duomenis ir programavimo, kuris yra glaudžiai susijęs su dideliais kiekiais duomenų (duomenų saugojimu) ir žmonėmis. Taigi DI etikos klausimai apima žmogaus veiksnius. „Skirtingų etinių iššūkių kyla kai yra svarstoma galimybė, kad ateities DI sistemos gali būti kandidatės moraliniam statusui gauti“ (Bostrom et al., 2014). Priimtas moralės apibrėžimo statusas, kad „X turi moralinį statusą todėl, kad X skaičiuoja moraliai savaime, tai leistina / neleistina daryti tai sau dėl savęs“. Žvelgiant iš šios perspektyvos, kai DI turi moralinį statusą, neturėtume jo traktuoti kaip mašina ar sistema, DI būtų laikomas objektu, turinčiu lygias teises kaip ir žmogus (Floris, 2020).

Tyrimai apie etinius DI iššūkius iš esmės skirstomi į tris kategorijas (Stahl et al., 2016):

1. DI ypatybės, kurios gali sukelti etines problemas
2. žmogiškieji veiksniai, keliantys etinę riziką,
3. etiški DI sistemos ugdymo būdai

DI ypatybės, dėl kurių gali kilti etinių problemų. Tyrimai rodo, kad DI gali „generuoti garsą, kuris skamba lyg kalba mašinai ir mokoma algoritmu“ (Adler, 2019). Tokiu būdu yra įmanoma, kad DI gali gauti prieigą prie asmeninės informacijos, be šeimininko žinios. Jei DI būtų atsakingas už sprendimo priėmimą ir panaudotų „mašininę kalbą“, tada kaip mes galėtume kontroliuoti rezultatus? Tokia grėsmė egzistuoja ir fiziniame pasaulyje (Kurakin et al., 2016), pavyzdžiui, savaeigiai automobiliai. DI, ypač mašininis mokymasis ir gilus mokymasis, ne visada yra skaidrūs, nes gali vystytis be žmonių stebėjimo ir nurodymų. Žemas skaidrumo lygis suteikia galimybę kilti piktnaudžiavimui.

Saugumas ir privatumas yra kiti iššūkiai. DI sistemos kūrimas labai priklauso nuo didžiulio kiekio duomenų, įskaitant asmens ir asmeninius duomenis. Šie duomenys turi būti tinkamai valdomi, kad būtų išvengta piktnaudžiavimo ir piktybiško naudojimą (Timmermans et al., 2010). Siekiant apsaugoti duomenis, kiekvienas duomenų panaudojimo veiksmas turėtų būti išsamus ir įrašomas. Tiek patys duomenys, tiek veikimo įrašai taip pat gali kelti su privatumu susijusią riziką. Todėl labai svarbu apsvarstyti kas turi būti įrašyta, kas turi prisiimti atsakomybę už veiksmo įrašymą ir kas turi prieigą prie duomenų bei įrašų (Edwards et al., 2018).

Žmogiškieji veiksniai, dėl kurių gali kilti etikos problemų. Svarbiausias veiksnys yra žmogaus šališkumas, pavyzdžiui, lyčių šališkumas ir lenktynių šališkumas (Jasanoff, 2016), kurį gali paveldėti DI. Kadangi DI sistemą vis dar treniruoja žmogus, naudojami žmogaus sukurti duomenų rinkiniai, galima sistemas išmokyti esamo šališkumo ir parodyti tai realiose programose.

Pavyzdžiui, programinė įranga, naudojama aptikti nusikaltėliams, parodė šališkumą prieš tam tikras rases (Bossman, 2016). Tokį neobjektyvumą lemia mokymo duomenys, kuriuose matomas žmonių šališkumas. Programuoti ir mokyti DI sistemas be žmonių šališkumo yra labai svarbu. Be to, jei DI įgyja savo jausmus ir išmintį (Bostrom et al., 2014) ar jis sugalvos savo šališkumą?

Kitas rūpestis yra atskaitomybė. Kai DI sistemai nepavyksta atlikti tam tikros paskirtos užduoties, kyla klausimas kas turėtų būti atsakingas. Tai gali sukelti vadinamąją „daugelio rankų problemą“ (Timmermans et al., 2010). Naudojant išmaniąsias sistemas, nepageidaujamas pasekmes gali sukelti programavimo kodai, įvesti duomenys, netinkamai pasirinktas veiksmas ar kiti faktoriai. Itin svarbu apsvarstyti ir tinkamai įvertinti, kas turėtų būti atsakingas už nepageidaujamas pasekmes: programuotojas, duomenų savininkas ar vartotojas (Ananny et al., 2018).

Kalbant apie etiškus ugdymo būdus svarbiausia galvoji apie vertybes, kurių yra laikomasi. Kaip reikia tinkamai elgtis su DI sistema, turinčia sąmonę, moralinį pojūtį, emocijas bei dar vieną svarbų dalyką – jausmus. Pavyzdžiui, ar etiška „nužudyti“ (išjungti) DI sistemą, jei ji pakeistų žmogaus

darbus ar net keltų pavojų žmonių gyvybėms? Ar etiška diegti robotus į pavojingą aplinką? Šie klausimai taip pat yra susiję su paties asmens etika ir moralinėmis vertybėmis.

Vystantis technikai, ypač tokioms intelektualioms technologijoms kaip: namų robotai ar sveikatos priežiūros įstaigų robotai, kurie neišvengiamai paveiks žmogaus gyvenimą ne tik fiziškai, bet ir etiškai. Tuo pačiu metu išaugs žmogaus ir roboto sąveika (Jasanoff, 2016).

Robotai laikomi moralės veiksniais, turinčiais poveikį sąveikai. Todėl turi būti vertinami kaip etikos objektas, atitinkantis šiuos kriterijus: autonomijos, intencionalumo ir atsakomybės. Autonomija reiškia, kad mašinos nėra tiesiogiai valdomos kitų agentų. Intencionalumas apibūdina tai, kad mašinos „veikia taip, kad būtų moraliai žalingos ar naudingos, o veiksmai yra apgalvoti ir apskaičiuoti“. Atsakomybė reiškia, kad mašinos atlieka tam tikrą socialinį vaidmenį bei tenka prisiimti dalį atsakomybės (Giubilini et al., 2018).

Sąvoką „turintis etinį statusą“ galima skirti į du, tarpusavyje susijusius aspektus: etinį produktyvumą ir etinį imlumą (Morley et al., 2019).

Etiški „gamintojai“ yra tie, kurie atlieka arba nevykdo savo pareigų (pavyzdžiui, šventieji arba žudikai). Etiški „gavėjai“ yra tie, kurie gauna naudos iš etikos arba jie yra paveikti etikos „gamintojų“. Žvelgiant iš šios perspektyvos, DI ir kitos išmaniosios mašinos gali būti tiek etikos „gamintojai“, tiek etikos gavėjai. Štai keletas pavyzdžių. Paprastas, klasikinis traukinio atvejis. Kontroliuojantis taikinį yra etiškas „gamintojas“. Tęsti važiavimą bėgiais ir nužudyti penkis darbininkus arba pasukti į kitus bėgius ir nužudyti vienišą darbuotoją? – etiškai sunkus pasirinkimas. Ką pasirinks DI? Kas bus atsakingas už tokį pasirinkimą? Kariniai robotai, kurie atsakingi už bombų pašalinimą, yra etiški „gavėjai“. Ar tai etiška, kad žmogus nusprendžia šių robotų likimą? Šiandien, žmogaus etika ir moralė negali būti vertinama kaip tobula ir pavyzdinė ateities civilizacijai (Bostrom et al., 2014). Viena iš priežasčių yra ta, kad žmogus negali išspręsti visų pripažintų etinių problemų. Kita – tai, kad žmogus negali atpažinti visų etinių problemų.

Galutinis ir labiausiai siekiamas mašinų etikos tikslas – sukurti mašiną, kuri pati laikytųsi idealaus etikos principo ar principų rinkinio. Teoriškai tai lengva, bet praktiškai sunku suformuluoti DI sistemų etinius principus. Pavyzdžiui, jei programuojame robotus taip, kad jie visada nekenktų, pirmiausia privalu įsitikinti, kad robotai supranta, kas yra žala. Dėl to kyla dar viena problema – kaip ir kas turi būti etiniai žalos standartai? Šioje vietoje yra reikalingas pasaulinis ir visuotinis etikos lygis bei supratimas. Išmokyti DI etikos, reiškia, kad būtina sumažinti informacijos asimetriją tarp DI programuotojų ir etikos standartų kūrėjų (Floridi et al., 2019; Morley et al., 2019).

Etinių ir moralinių iššūkių susijusių su DI, supratimas ir sprendimas vis dar yra labai ribotas. Tai nėra paprasta „teisinga ar neteisinga“, „gera ar bloga“, „dorybė ar yda“ problema. Tai ne problema,

kurią gali išspręsti maža žmonių grupė. Etiniai ir moraliniai su DI susiję klausimai yra kritiniai reikia aptarti dabar. Todėl labai svarbus visuomenės dėmesys, taip pat šalių kuriamos strategijos ir etikos gairės. Bandant suformuoti pažangių technologijų standartus, taip pat geriau suprasime žmogaus etiką ir moralę.

Didėjant „didžiųjų duomenų“ ir skaičiavimo metodų sklaidai, technologinė raida sparčiai plinta ir vis intelektualesni algoritmai tapo puikiu šaltiniu naujovėms ir verslo modeliams.

Verbeek (2011) savo veikalė „Moralizuoti technologijas: daiktų moralės supratimas ir kūrimas“ siekia išplėsti etikos sritį, kad ji geriau atitiktų technologinį amžių ir tai darydamas atskleidžia neatskiriamą žmonijos ir technologijų prigimtį. Remiantis filosofo indėliu, technologijos gali būti laikomos „moraliniais tarpininkais“, kurios yra žmogaus pasaulio suvokimo ir sąveikos atspindys, atskleidžiantis galimą elgesį. Kadangi kiekviena technologija daro poveikį mūsų suvokimo ir sąveikos su pasauliu būdai ir netgi mąstymo būdai, jokia technologija nėra moraliai neutrali - ji tarpininkauja mūsų gyvenimui. Masolo et al. (2020) techninius artefaktus paaiškino, kaip žmogaus sukurtus daiktus (objektus), kurie turi funkciją ir naudojimo planą. Jie susideda iš produktų, gautų atliekant technologinius veiksmus, nurodant požiūrį, kurio kasdien laikomasi sprendžiant praktines problemas, įskaitant tas, kurios susijusios su mūsų norais ir poreikiais. Techniniai artefaktai apima būtinybę laikytis naudojimo taisyklių ir sukurti parametrus, susijusius su individų ir socialinių institucijų vaidmenimis (Vermaas et al., 2011).

Sudėtingos sistemos kelia diskusiją apie techninių artefaktų ir socio-techninių sistemų atsakomybę ir etiką. Tokiems klausimams kaip kūrėjų atsakomybė ir moralės egzistavimas ne žmogaus veiksmuose – daugiausia dėmesio skiriant technologiniams objektams – reikia atsakymo ir viešų diskusijų.

Magrani (2019) teigia, kad kai intelektualūs dalykai, tokie kaip robotai, tampa autonomiškesni ir prisiima daugiau atsakomybės, jie turi būti užprogramuoti turėdami moralinius sprendimų priėmimo įgūdžius mūsų pačių saugumui užtikrinti.

Patvirtindamas šią tezę ir nagrinėdamas mašinų moralę, Verbeekas (2011), supranta, kad: mašinos dažniau veikia atviroje socialinėje aplinkoje, todėl tampa vis svarbiau sukurti jautrią funkcinės moralės tipą, etiškai svarbioms situacijoms numatyti ir taikyti.

Puikus pavyzdys yra „Microsoft“ robotas „Tay“, kuris padeda iliustruoti žmogaus elemento poveikį visuomenei. 2016. „Microsoft“ sukūrė DI programą. „Tay“ yra apdovanotas giliu mokymosi gebėjimu. Remdamasis internetine sąveika su kitais žmonėmis ir tokiu pagrindu kurdamas autentiškas išraiškas, robotas formavo savo pasaulėžiūrą. Tačiau patirtis pasirodė pražūtinga, įmonė turėjo išjungti robotą nepaėjęs nei 24 valandas, dėl nerimą keliančių rezultatų: naudotojų paskatintas

robotas pradėjo mechaniškai kartoti rasistinio pobūdžio, seksistinio, diskriminacinio ir pornografinio turinio posakius. Programų, galinčių mokytis ir prisitaikyti prie funkcijų, susijusių su žmonėmis, įgyvendinimas kelia naujų etinių ir teisinio reguliavimo problemų, nes tai padidina galimybę pasiekti visiškai kitokių rezultatų nei buvo tikėtasi (Neff et al., 2016; Wolf et al., 2017).

Pirmiausia, norit naudoti DI sistemas, kurios sąveikauja per socialinius tinklus ar žiniasklaidą, reikia apmąstyti etinius reikalavimus, kurie užtikrintų tinkamą tokio tipo technologijų plėtojimą. Šie mechanizmai veikia kaip visuomenės veiksniai ir galiausiai daro poveikį juos supančiai aplinkai, nors jie ir nėra žmogiškieji elementai. Todėl kalbama ne tik apie naujų technologijų „naudojimą“ ir „taisymą“, bet daugiausia apie tinkamą jų raidos etinę orientaciją (Miller et al., 2017).

„Microsoft“ teigė, kad netinkamas „Tay“ veikimas atsirado dėl vartotojų, kurie pasinaudojo programos pažeidžiamumu. Tačiau tai kūrėjų neatleidžia nuo atsakomybės apsvarstyti galimų žalingų pasekmių atsiradimą naudojant programinę įrangą (Miller et al., 2017; Schwartz, 2019).

Siekiant paversti DI sistemas vis labiau pritaikomomis ir veikiančiomis panašiai kaip žmogus, atsiranda didesnė galimybė sistemų nenusipėjamo elgesio atsiradimui. Taigi technologiją reikia numatyti ne tik kaip įrankį, atliekantį iš anksto nustatytą funkciją, bet ir sukurti tinkamą veikimo būdą. Verta pabrėžti, kad algoritmai gali prisitaikyti, kad atsirastų naujų algoritmų ir naujų būdų atlikti savo užduotis (Domingos, 2015), kad rezultato pasiekimo būdą būtų sunku paaiškinti net programuotojams, sukūrusiems algoritmą (Doneda et al., 2016). Be to, kuo labiau pritaikomos DI programos, tuo labiau nenusipėjami yra jų veiksmai, keliantys naują riziką. Todėl programų kūrėjams būtina gerai suprasti etinę ir teisinę atsakomybę, susijusią su šia veikla.

Kompiuterinių mašinų asociacijos etikos kodekse (Miller et al., 2017) nurodoma, kad DI sistemų kūrėjai (programuotojai), neatsižvelgdami į ankstesnį teisinį reguliavimą, turėtų parengti išsamų kompiuterinių sistemų ir jų poveikio vertinimą, įskaitant galimos rizikos ir žalos analizę. Taip pat reikalinga stebėseną, kad būtų galima patikrinti veiksmus, kurių buvo imtasi prieš sistemos sukūrimą, ypač ankstyvais įgyvendinimo etapais. Pavyzdžiui, „Tay“ atveju kūrėjai turėjo intensyviai stebėti roboto elgesį pirmąsias 24 valandas nuo paleidimo. Turėtų būti sudaryta galimybė užkirsti kelią galimai žalai ir iš anksto stebinti, o ne ištaisyti klaidas, ypač kai jų neįmanoma numatyti.

Norėdami apriboti neigiamų pasekmių galimybes, programinės įrangos kūrėjai turi atpažinti, apsibrėžti ir apriboti potencialiai pavojingas ir nenusipėjamas programos sąveikas su visuomene, kol ji nebus pakankamai išbandyta kontroliuojamoje aplinkoje. Po šio etapo vartotojai turėtų būti informuojami apie nenusipėjamą programos pažeidžiamumą ir galimas netikėtos elgsenos pasekmes (Miller et al., 2017).

Gebėjimas kaupti patirtį ir mokytis iš masinio duomenų apdorojimo kartu su gebėjimu veikti savarankiškai ir savarankiškai rinktis, gali būti laikomas teisinės atsakomybės prielaida. Kadangi, šiandien, DI nėra pripažįstamas teisės subjektu, jis negali būti laikomas atskirai atsakingu už galimą jo daromą žalą. Remiantis JT konvencijos dėl elektroninių ryšių naudojimo tarptautinėse sutartyse 12 straipsnį, asmuo (fizinis ar subjektas), kurio vardu buvo sukurta programa, galiausiai turi būti atsakingas už bet kokius veiksmus, kuriuos sukėlė mašina. Šis samprotavimas grindžiamas nuostata, kad įrankis neturi savo valios (Čerka ir kt., 2015).

Kita vertus, jei padaroma žala dėl DI artefakto, kita atsakomybės rūšis yra tokia, kuri daro analogiją su atsakomybe, priskiriama tėvams dėl savo vaikų veiksmų ar atsakomybe dėl gyvūno, savininkams žalos atveju. Šiuo požiūriu atsakomybė už šio artefakto veiksmus gali tekti ne tik jo gamintojui ar programuotojams, bet ir vartotojams, kurie buvo atsakingi už jų „mokymą“ (Čerka ir kt., 2015). Kita galimybė yra modelis, kuriame pagrindinis dėmesys skiriamas programuotojų ar vartotojų galimybei numatyti šių žalų atsiradimo galimybę. Pagal šį modelį programuotojas ar vartotojas gali būti laikomas atsakingu, jei elgėsi apgaulingai arba buvo aplaidus, atsižvelgdamas į rezultatą, kurį būtų galima nuspėti (Hallevy, 2010).

Iki šiol pasaulio teismai nėra aiškiai suformulavę atsakomybės, susijusios su intelektinės nuosavybės kūrimu, apibrėžimo, kurios būtų imtasi dėl padarytų klaidų. Nustatant griežtą atsakomybę, pažeidėjui sukuriama didelė paskata elgtis kruopščiai, siekiant sumažinti galimą žalą. Jei DI veiksmas padaro žalos dėl aplaidumo, apgaulės ar gamybos defekto taisyklės dažniausiai nurodytų jo kūrėjų atsakomybę. Tačiau dažnai nėra lengva žinoma kaip programos sureaguos į tam tikras komandas ar užduotis. Naudojant DI programas, kurios remiasi mašininiu mokymusi ir ypač giliojo mokymosi mechanizmais, kai pats programinės įrangos pobūdis apima ketinimą sukurti nenuspėjamą veiksmą, kuris bus nustatytas tik apdorojant visą informaciją, kurią gavo programa. Galiojantys įstatymai nėra pakankami, kad būtų užtikrintas teisingas reguliavimas DI kontekste.

Įvairių sričių mokslininkai yra susirūpinę ir svarsto, kad suteikus mašinoms autonomini „mąstymo“ gebėjimą, kartu bus suteikta galimybė elgtis priešingai, nei nurodyta taisyklėse. Norint to išvengti, reikia ypatingą dėmesį skirti DI technologijų kūrėjų, mokslininkų poveikio sritims kuriant ir plėtojant techninius artefaktus (Vladeck, 2014). Kūrėjai yra atsakingi už tai, kad apgalvotų darbo schemą bei vertybes, kuriomis remiantis bus kuriamos sistemos, jos funkcija bei naudojimosi vadovas (instrukcija). Atkreipti dėmesį į tai, kas nepateks į projektavimo ir naudojimo vadovą taip pat kokios nenuspėjamos problemos, nepriklausančios nuo kūrėjo gali kilti. Mokslininkams ir akademinėms bendruomenėms būtina skatinti plačias diskusijas apie etines gaires, kurios turėtų būti intelektualių mašinų tinkamo kūrimo vadovas.

DI yra vis didesnis prioritetas vyriausybinių institucijų politikos darbotvarkėse tiek nacionaliniu, tiek tarptautiniu lygiu. Daugelyje nacionalinių vyriausybių iniciatyvų iki šiol daugiausia dėmesio buvo skiriama DI naudojimui produktyvumui ir konkurencingumui skatinti. Taip pat vis daugiau susidomėjimo kelia tokie DI klausimai kaip: skaidrumas, žmogaus teisės ir etika. Daugelis šalių paskelbė nacionalines DI strategijas ir politines iniciatyvas, kuriomis dažniausiai siekiama užtikrinti lyderio pozicijas DI srityje. Strategijose ir iniciatyvose nustatomi tikslai ir uždaviniai, kuriems reikalingi suderinti visų suinteresuotųjų šalių veiksmai. Vyriausybės vaidmuo dažnai yra sutelkiantis ir tarpininkaujantis. Be to, kai kurios šalys sukūrė arba priskyre atsakomybę konkrečiam viešajam subjektui DI ir duomenų etikos klausimais (OECD, 2020).

Šalims kuriant DI strategijas, labai svarbu, kad jos nurodytų galimą žalą ir galimybes, susijusias su žmogaus teisėmis ir pasirengtų ginti bei skatinti žmogaus teises šios technologijos kontekste (Stanford, 2020). DI samprotavimuose turėtų būti atsižvelgiama į visuomenės vertybes, moralinius ir etinius sumetimus; pasverti atitinkamus skirtingų suinteresuotųjų šalių vertybių prioritetus įvairiuose daugiakultūriuose kontekstuose; paaiškinti jos samprotavimus bei garantuoti skaidrumą (Dignum, 2018). Žmogaus teisėmis pagrįstas požiūris į DI yra geriausias būdas vyriausybėms apsaugoti piliečius nuo galimos žalos atsiradimo, taip pat užtikrinti didžiausią galimą naudą (Stanford, 2020).

Todėl manau, kad ES šalių tarptautinės ir nacionalinės iniciatyvos, gairės bei strategijos visame pasaulyje gali būti puiki perspektyva ir siekis teisingo ir etiško DI kūrimo, vystymo ir naudojimo visuomenėje (1 lentelė).

1 lentelė. Etinės iniciatyvos ir žala

Iniciatyva	Vieta	Pagrindiniai sprendžiami klausimai	Publikacijos	Finansavimo šaltiniai
DI etikos institutas¹	Vokietija	Į žmogų orientuota inžinerija ir dėmesys kultūriniam ir socialiniam DI pažangos įtvirtinimui, apimant filosofiją, etiką, sociologiją ir politikos mokslus		Pradinis finansavimo šaltinis „Facebook“ (2019 m. 7,5 mln. dolerių per penkis metus)
Etiško DI ir mašininio mokymosi institutas²	Jungtinė Karalystė	Institutas siekia, kad DI būtų vystomas pagal atsakingos mašinos mokymosi principus: kontrolės, žalos atlyginimo, šališkumo, aiškumo, skaidrumo, atkuriamumo, išlaidų, privatumo, pasitikėjimo ir saugumo		nežinomas
Etinio DI švietimo institutas³	Jungtinė Karalystė	Galimos grėsmės jauniems žmonėms, švietimas dėl DI technologijų augimo ir etinio vystymo vadovaujantis „EdTech“ (edukacinėmis technologijomis)		nežinomas
Gyvenimo ateities institutas⁴	Jungtinės Valstijos	Užtikrinti, kad DI vystymasis būtų naudingas žmonijai, skiriamas dėmesys saugumui ir egzistencinei rizikai (autonominiai ginklai), žmogaus kontrolė DI srityje ir galimi pavojai	„Asilomar DI“ principai „Asilomar AI Principles“	Privatus sektorius: Elonas Muskas („SpaceX“ ir „Tesla“), Jaanas Talinas („Skype“), Mattas Wage'as (finansininkas), Nisanas Stiennonas (programinės įrangos inžinierius), Sam Harris, George'as Godula (technikos verslininkas) ir Jokūbas Trefethenas (Harvardas)
Skaičiavimo mašinų asociacija⁵	Jungtinės Valstijos	Skaidrumas, tinkamumas naudoti, saugumas, prieinamumas, atskaitomybė bei skaitmeninis kompiuterių ir tinklų įtraukimas tyrimų srityse (kūrimas ir įgyvendinimas)	Teiginiai: algoritmų skaidrumas ir atskaitomybė (2017 m. sausio mėn.), Skaičiavimas ir tinklo saugumas (2017 m. gegužės mėn.), prieinamumas, patogumas ir skaitmeninis	nežinomas

			įtraukimas (2017 m. rugsėjo mėn.)	
JSAI⁶	Japonija	Užtikrinti, kad DI tyrimai ir taikomoji veikla išliktų naudinga žmonių visuomenei, o tyrimai būtų atliekami etiškai ir moraliai	„Etikos gairės“ „Ethical Guidelines“	nežinomas
AI4ALL⁷	Jungtinės Valstijos	Įvairovė ir įtraukimas – DI socialiniai naudai ir žmonijos labui		Google
Ateities visuomenė⁸	Jungtinės Valstijos	DI poveikis ir valdymas iš esmės naudingas moksliniams tyrimams ir valdymui, teisei ir švietimo koordinavimui	„Principų projektas DI valdymas“ „Draft Principles for the Governance of AI“ (paskelbta 2017 m., vėliau 2019 m.)	nežinomas
DI dabarties institutas⁹	Jungtinės Valstijos	DI socialinės pasekmės, ypač teisės, ekonomikos, šališkumo, saugumo, infrastruktūros srityse		Įvairios organizacijos: „Microsoft“, „Google“, „Ford“, „DeepMind“
IEEE¹⁰	Jungtinės Valstijos	Gairės DI ir intelektualioms sistemoms, kad būtų išlaikomas orientuotas į žmogų ir tarnaujantis žmonijos vertybėms bei principams požiūris. Pirmenybę teikti etikai: žmogaus teisių, gerovės, atskaitomybės, skaidrumas ir supratimas apie netinkamą naudojimą užtikrinimui	„Etiškas dizainas“ „Ethically Aligned Design“ (2019 m. kovo mėn.)	

¹ The Institute for Ethics in Artificial Intelligence. Prieiga per internetą: <<https://ieai.mcts.tum.de/>>.

² The Institute for Ethical Artificial Intelligence and Machine Learning. Prieiga per internetą: <<https://ethical.institute/>>.

³ The Institute for Ethical Artificial Intelligence in Education. Prieiga per internetą: <<https://www.buckingham.ac.uk/research-the-institute-for-ethical-ai-in-education/>>.

⁴ The Future of Life Institute. Prieiga per internetą: <<https://futureoflife.org/>>.

⁵ The Association for Computing Machinery. Prieiga per internetą: <<https://www.acm.org/>>.

⁶ The Japanese Society for Artificial Intelligence. Prieiga per internetą: <<http://ai-elsi.org/archives/514>>.

⁷ AI4ALL. Prieiga per internetą: <https://ai-4-all.org/>.

⁸ The Future Society. Prieiga per internetą: <<http://thefuturesociety.org/>>.

⁹ The AI Now Institute Prieiga per internetą: <<https://ainowinstitute.org/>>.

¹⁰ IEEE. Prieiga per internetą: <<https://www.ieee.org/>>.

Partnerystė DI srityje ¹¹	Jungtinės Valstijos	DI technologijos: sauga, sąžiningumas, atskaitomybė, skaidrumas, darbas ir ekonomika, bendradarbiavimas, socialinis visuomenės poveikis		Partnerystę užmezgė DI tyrėjų grupė atstovaujanti šešioms didžiausioms pasaulio technologijų kompanijoms: „Apple“, „Amazon“, „DeepMind“, „Google“, „Facebook“, „IBM“, „Microsoft“
Atsakingos robotikos fondas ¹²	Olandija	Atsakinga robotika (projektavimas, kūrimas, naudojimas, reguliavimas ir įgyvendinimas). Technologines naujoves lydintys klausimai ir galimas poveikis		nežinomas
AI4People ¹³	Belgija	DI socialinis poveikis ir pagrindiniai principai, politika ir praktika, kuria remiantis galima sukurti „gerą DI visuomenę“	„Etinė gėrio sistema DI draugija“ „Ethical Framework for a Good AI Society“	„Atomium“ — Europos mokslo institutas, žiniasklaida ir demokratija. Skirtas finansavimo projekto mokslinis komitetas
DI etikos ir valdymo iniciatyva ¹⁴	Jungtinės Valstijos	Siekia užtikrinti, kad automatikos ir mašinų technologijų mokymasis būtų tiriamas, plėtojamas ir diegiamas taip, kad būtų užtikrintos vertybės		Harvardo Berkmano Kleino centras ir „MIT Media Lab“, Majamio fondas, „Luminate“, „Red Hoffman“ ir Williamo bei Flora Hewletto fondai
Saidot: įgalinti atsakingas DI ekosistemas ¹⁵	Suomija	Padėti įmonėms, vyriausybėms ir organizacijoms vystyti atsakingas DI sistemas, kad būtų užtikrinta skaidri, atsakinga, patikima DI taryba. Įgalinti organizacijas plėtoti į žmogų orientuotą DI, daugiausia dėmesio skiriant pasitikėjimui ir atskaitomybei DI ekosistemose (Saidot, 2019)		

¹¹ The Partnership on AI. Prieiga per internetą: < <https://www.partnershiponai.org/> >.

¹² The Foundation for Responsible Robotics. Prieiga per internetą: < <https://responsiblerobotics.org/> >.

¹³ AI4People. Prieiga per internetą: < <https://ai4people.eu/> >.

¹⁴ The Ethics and Governance of Artificial Intelligence Initiative. Prieiga per internetą: < <https://aiethicsinitiative.org/> >.

¹⁵ Saidot: Enabling responsible AI ecosystems. Prieiga per internetą: < <https://www.saidot.ai/platform> >.

EuRobotics¹⁶	Europa	Išsaugoti ir išplėsti Europos pažangą robotikos srityje DI industrializacija ir ekonominis poveikis		EK
Duomenų etikos ir inovacijų centras¹⁷	Jungtinė Karalystė	Nustatyti ir pašalinti reguliavimo spragas, maksimali DI nauda visuomenei		Jungtinės Karalystės vyriausybė
SIGAI¹⁸	Jungtinės Valstijos	DI principų taikymo ir palaikymo skatinimas proceso metu. Visuomenės švietimas DI klausimais per įvairius leidinius ar forumus		Mašininio skaičiavimo asociacija „The Association for Computing Machinery“
Kiti svarbūs tarptautiniai pokyčiai				
Monrealio konvencija¹⁹	Kanada	Socialiai atsakingas DI vystymas, sujungiant 400 dalyvių iš įvairių sektorių nustatyti, etinius ir moralinius iššūkius. Pagrindinės vertybės: gerovė, autonomija, teisingumas, privatumas, žinios, demokratija ir atskaitomybė		Monrealio universitetas su „Fonds de“ parama
UNI globalioji sąjunga²⁰	Šveicarija	Skaidrumas darbo vietose taikant DI. Darbuotojų interesų saugojimas, išlaikant kontrolę	„10 geriausių etinio intelekto principų“ „Top 10 Principles for Ethical AI“	nežinomas
EURON²¹	Europa (koordinatorių Švedija)	Mokslinių tyrimų koordinavimas, švietimas ir mokymas. Pramoniniai ir tarptautiniai ryšiai robotikos srityje	„Robo-etikos planas“ „Roboethics Roadmap“	EK (2000-2004)
EUROP²²	Europa	Suburti Europos robotikos ir DI bendruomenę. Pramonės susitelkimas į konkurencingumą ir naujoves		EK

¹⁶ EuRobotics. Prieiga per internetą: <<https://www.eu-robotics.net/>>.

¹⁷ The Centre for Data Ethics and Innovation. Prieiga per internetą: < <https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation>>.

¹⁸ SIGAI. Prieiga per internetą: <<https://www.acm.org/special-interest-groups/sigs/sigai>>.

¹⁹ The Montréal Declaration. Prieiga per internetą: < <https://recherche.umontreal.ca/english/strategic-initiatives/montreal-declaration-for-a-responsible-ai/>>.

²⁰ The UNI Global Union. Prieiga per internetą: < <http://www.thefutureworldofwork.org/docs/10-principles-for-ethical-artificial-intelligence/>>.

²¹ The European Robotics Research Network. Prieiga per internetą: <<https://cordis.europa.eu/project/id/730994>>.

²² The European Robotics Platform. Prieiga per internetą: <<https://op.europa.eu/en/publication-detail/-/publication/5ee0fc69-4066-4f30-a51d-efb3fb7452db>>.

Visos 1 lentelėje paminėtos iniciatyvos sutinka, kad DI turėtų būti tiriamas, kuriamas, projektuojamas, diegiamas, stebimas ir naudojamas etiškai, tačiau kiekviena iš jų turi skirtingas prioritetas. Iš iniciatyvų iškyla keletas pagrindinių klausimų, kuriuos galima suskirstyti į šias kategorijas (EPRS, 2020):

1. Žmogaus teisės ir gerovė. Ar DI yra naudingiausias žmonijai ir žmogaus gerovei?
2. Emocinė žala. Kaip DI gali pakeisti žmogaus emocinę ar psichologinę sveikatą?
3. Atskaitomybė ir atsakomybė. Kas yra atsakingas už DI ir jo veiksmus?
4. Saugumas, privatumas, prieinamumas ir skaidrumas. Kaip subalansuoti prieinamumą ir skaidrumą su privatumu ir saugumu?
5. Saugumas ir pasitikėjimas. Kas bus, jei visuomenėje DI bus laikomas nepatikimu arba kels grėsmę saugumui?
6. Socialinė žala ir socialinis teisingumas. Kaip bus užtikrinta, kad DI būtų patikimas, nešališkas, nediskriminuotų ir atitiktų visuomenės moralę ir etiką?
7. Finansinė žala. Kaip bus išvengta neigiamo DI poveikio ekonomikai (nedarbas, sumažėjęs darbų pasirinkimas)?
8. Teisėtumas ir teisingumas. Kaip bus užtikrinta, kad DI ir jo surinkti duomenys būtų naudojami, apdorojami ir valdomi teisėtai ir tinkamai? To reguliavimas? Ar DI turėtų būti suteikta asmenybė?
9. DI kontrolė ir etiškas naudojimas - arba netinkamas naudojimas. Kaip DI gali būti naudojamas neetiškai - ir kaip nuo to apsisaugoti? Kaip užtikrinti, kad DI bus visiškai kontroliuojamas žmogaus, net kai jis vystosi ir „mokosi“ ?
10. Žala aplinkai ir tvarumas. Kaip apsisaugoti nuo galimos žalos aplinkai, susijusios su DI kūrimu ir naudojimu? Tvari gamyba.
11. Informuotas naudojimas. Ką reikia daryti, kad visuomenė būtų informuota bei supažindinta su tinkamu DI naudojimu?
12. Egzistencinė rizika. Kaip išvengti DI ginklavimosi „varžybų“, prevenciškai sušvelninti ir sureguliuoti galimą žalą ir užtikrinti, kad pažangus mašininis mokymasis būtų progresyvus ir valdomas?

Siekiant išvengti neigiamų padarinių būtina numatyti, kokios prevencinės priemonės galėtų sumažinti galimą DI žalą iki minimumo bei numatyti netinkamos DI veiklos kontrolės priemones. Tyrimai rodo, kad efektyviausias poveikis netinkamam elgesiui – ankstesnių priežasčių prevencija, kai veikiamas ne subjektas, o tiesioginės priežastys, kurios lemia atitinkamą subjekto poelgį (Harry, 2014). Vadinasi, labai svarbu išsiaiškinti, kokie veiksniai nulemia subjekto nusikalstamą veikimą ir priimti atitinkamus sprendimus, kurie padėtų išvengti neigiamo poveikio. Siekiant sukurti kuo tobulesnę DI, labai daug informacijos ir įvairių mokslinių straipsnių galima rasti apie saugaus DI sukūrimą, tačiau informacijos apie piktavališko DI kūrimo procesą nėra. Pistono et al. (2019) mano,

kad pastarosios informacijos trūkumas kenkia tinkamo DI kūrimui ir vystymui, kadangi nėra žinoma, ko programavimo procese reikėtų vengti, norint sukurti žmogui ir aplinkai nepavojingą DI. Vienintelis netinkamas programavimo ar kūrybos momentas gali privesti prie didelės žalos sukūrimo, todėl svarbu iš anksto užkirsti kelią galimiems pavojams.

Vienas iš prevencijos būdų tai DI kūrimo priežiūros taryba, kurios pagrindinis tikslas būtų užtikrinti, kad Europos institucijos ir įstaigos gerbtų teisę į privatumą ir duomenų apsaugą, kuomet yra tvarkomi asmens duomenys, taip pat kokiomis vertybėmis ir tikslais yra grįstos DI technologijų vystymas. Ši taryba būtų sudaryta iš DI technologijos žinovų grupės, kuri vertintų kiekvieną DI kūrybos projektą ir pasiūlymą, spręstų apie jo priėmimą ar atmetimą, prižiūrėtų kūrybos procesą ir reikalautų DI programuotojų bei kūrėjų atskaitingumo. Tokia atsakinga priežiūra padėtų kontroliuoti DI gamybos spartumą ir išvengti piktavališko ar netinkamo DI sukūrimo. Kadangi DI yra kuriamas pasitelkiant uždarus šaltinius, patentuotą programinę ir techninę įrangą, bet kokia žmonių grupė, kuri turi tikslą sukurti piktavališką DI, turi visas tinkamas sąlygas tokio DI kūrybos procesui, nes patentų teisė, autorių teisių įstatymai ar net pramonės teisė užtikrina pastarajam apsaugą. Tokia tarnyba galėtų užkirsti kelią įstatymo spragų naudojimuisi įvesdama prievolę teikti tarybai visą informaciją apie autonominės sistemos kūrimą. Taip būtų sumažinta tikimybė, kad sukurtas DI taps neetišku, nekontroliuojamu ar žalingu (Thomas, 2019).

Panašų pasiūlymą teikia ir kiti mokslininkai, norėdami įsteigti Tarptautinę DI organizaciją (Erdelyi et al., 2018), kurios tikslas ir paskirtis būtų tapati priežiūros tarybos tikslams ir paskirčiai. Dar kiti siūlo paskelbti DI vystymosi aktą, kuriame būtų nurodyta, kuo vadovaujantis DI turėti būti kuriamas, kam DI kūrėjas / programuotojas yra atskaitingas, nurodant konkrečias organizacijas (Scherer, 2015). Šio akto dėka DI kūrimas ir veikimas būtų iš dalies teisiškai reglamentuotas, įvesta daugiau aiškumo ir konkretumo. Europos Parlamentas 2017 m. vasario mėn. ėmėsi iniciatyvos dėl DI kontrolės ir paskelbė rezoliuciją, kurios pagrindu rekomenduoja Sąjungos šalims įvesti taisykles dėl atsakomybės dėl DI kuriamos žalos, o konkrečiai siūlo šalims kartu sukurti Europos robotų ir DI agentūrą, kurios tikslas būtų DI ir robotų techninė, etinė ir moralinė kontrolė (Siau et. al, 2018). Taigi tam, kad išvengti DI nusikalstamo veikimo yra labai svarbu sukurti prevencinę sistemą, kurios pagrindu būtų prižiūrimas DI kūrimas, veikla, nustatytos taisyklės, kuriomis privalėtų vadovautis visi, norintys dalyvauti veikloje, susijusioje su DI (Taddeo et al., 2015). Tai, kad teisės sistemos sukurti įstatymai padeda netinkamų ketinimų turintiems asmenims diegti programiškai pavojingą informacinės technologijos sistemą, rodo, jog įstatymų leidėjai ir teisininkai nėra susipažinę su DI veikimo principais bei neturi supratimo apie tokios autonominės sistemos programavimo ypatybes.

Norint išvengti DI nusikalstamo ir netinkamo veikimo, siūloma teisės sistemos kūrėjams bendradarbiauti su informacinių technologijų specialistais, skatinti DI tyrimus ir plėsti žinias šioje

sirtyje (Brundage et al., 2018). Tai, pirmiausia, įpareigoja įstatymo leidėją užtikrinti, kad esami teisės aktai netrukdytų DI ekspertams plėsti žinių lauko apie DI technologijos saugaus ir piktavališko programavimo niuansus. Antra, abipusis bendradarbiavimas užtikrins, kad teisinės sistemos kūrėjai bus pilnai informuoti apie DI galimybes, privalumus ir trūkumus, o tai palengvins tinkamų ir teisingų teisės aktų projektų paruošimą ir priėmimą. Kadangi DI pats iš savęs yra dvigubo naudojimo technologija (jis gali būti autonominis, pats daryti sprendimus ir veikti arba juo galima pasinaudoti kaip įrankiu konkrečiam tikslui pasiekti), ekspertai ir inžinieriai turi pareigą visomis jiemis galimomis ir prieinamomis priemonėmis skatinti tik teigiamą DI technologijos naudojimą. Tai siūloma įgyvendinti kuriant viešą diskusiją apie DI, įtraukiant visas suinteresuotas šalis ir ekspertus, pavyzdžiui, visuomenę, nacionalinio saugumo specialistus, neapsaugoto DI ir kibernetinės saugos žinovus, verslo atstovus, kurie remiasi DI naudojimu savo produkcijos kūrimo bei etikos specialistus. Ši vieša diskusija padėtų gilinti žinias autonominės sistemos egzistavimo srityje ir išvengti ne tik netinkamo DI naudojimo, bet ir galimai sumažinti potencialią DI kuriamą žalą, kadangi plėstūsi supratimas apie DI programavimo principus, galimas technologines problemas autonominėje sistemoje ir kaip jas maksimaliai sumažinti.

Apibendrinant, galima spręsti, kad neetiškas ir netinkamas DI kūrimas ir jo nusikalstamas veikimas gali būti kontroliuojamas, jeigu teisininkai ir įstatymų leidėjai bendradarbiaus su DI technologijos ekspertais ir mokslininkais, plės žinias ir suvokimą, apie DI veiklą, skatins tyrimus ir palaikys DI kūrimo priežiūros tarybos ar panašios organizacijos veiklos vykdymą. Įvairiomis iniciatyvomis siekiama nustatyti ir suformuoti etišką DI sistemų kūrimą ir pritaikymą, kurios sukuria didžiausią naudą bei pirmenybę teikia visuomenei ir aplinkai, sumažindamos galimą neigiamą DI poveikį, tai yra, daugiausia dėmesio skiriant, kad DI būtų skaidrus ir atsakingas (IEEE, 2017).

IŠVADOS

1. Nėra vieningo DI sistemų apibrėžimo, tačiau iš įvairių mokslinių šaltinių darytina išvada, kad DI yra suvokiamas kaip sistema pasitelkianti mašinas atlikti užduotis, kurių atlikimui paprastai būtinas žmogaus intelektas: gebėjimas suprasti, bei mokytis iš aplinkinių faktorių, pavyzdžiui regėjimo suvokimas, kalbos atpažinimas ir vertimas, sprendimų priėmimas, gebėjimas apdoroti istorinius ir momentinius duomenis, kad būtų galima parengti prognozes, pasiūlymus, rekomendacijas ir kitus sprendimus.
2. Išanalizavus teisės aktus bei mokslinių šaltinių analizę, DI sukurto objekto / kūrinio galimybes autorinių teisių apsaugai, pastebėta, kad DI sukurtas objektas nei Lietuvoje, nei užsienio šalyse autorinėmis teisėmis nėra apsaugoti, tačiau teik užsienio šalyse, tiek Lietuvoje šis klausimas jau pradėtas spręsti kaip būtinas ir reikalingas. Pagal galiojančius teisės aktus autorių teisių subjektas yra tiesiogiai susijęs su kūrinio sukūrimu ir pagrindinis aspektas yra kūrinio sukūrimo faktas, todėl DI sukurdamas autonomiškai kūrinį tampa autorių teisių subjektu. Pagal galiojančius teisės aktus, kūrinys pripažįstamas kaip autorių teisių objektas jei yra originalus kūrybinės veiklos rezultatas ir turi objektyvios išraiškos formą, DI sukurtas objektas / kūrinys atitinka keliamus reikalavimus.
3. Vykstant nuolatiniam tobulėjimui technologijų srityje, dėl kurių robotai šiuo metu jau geba savarankiškai priimti sprendimus, kyla poreikis sukurti tinkamai pritaikytą taisyklių sistemą, kurioje naudojimas DI, suvaldytų tokius rizikos faktorius, kaip tam tikras elgesio nenuspėjamumo lygmuo bei sąveikavimas su aplinka. Etikos principų įtraukimas į DI technologijų plėtros procesą yra ne tik būtinas saugiklis, užkertantis kelią žmonijos susinaikinimui, bet ir būdas geriau suprasti, kaip atsakingas DI technologijų naudojimas leidžia gauti maksimalią naudą valstybei, visuomenei ir kiekvienam žmogui. Svarbiausias iššūkis, pasitikėjimo vertas etiškas / moralus DI.
4. Apie etines mašinų ir automatikos diegimo pasekmes ir principus jau buvo kalbėta apie 50–60 metus (Samuel, 1959; Wiener, 1988), vis didėjantis DI naudojimas daugelyje sričių kelia naujų svarbių klausimų apie jo tinkamumą (Yu et al., 2018). DI etikos sritis yra dar tik pradinėje stadijoje, kurią reikia konceptualizuoti, kaip būtų galima pasiekti DI raidą, apimančius etinius aspektus. Todėl kuriant DI sistemas bei rengiant strategijas ir dokumentus svarbu laikytis skaidrumo, atskaitomybės, sąžiningumo ir kitų vertybinių principų, kurie užtikrintų etišką, patikimą ir į žmogų orientuotą DI.

LITERATŪRA

1. Adler S. 2019. A Quick Guide to Artificial Intelligence (AI) [interaktyvus]. Intelligent automation network, [žiūrėta 2020 m. rugsėjo 9 d.]. Prieiga per internetą: <<https://www.intelligentautomation.network/decision-ai/news/a-basic-guide-to-ai>>.
2. Adoriasoft. 2018. Current state of artificial intelligence and its future possibilities [interaktyvus]. Adoriasoft, [žiūrėta 2020 m. rugsėjo 29 d.]. Prieiga per internetą: <<https://adoriasoft.com/blog/current-state-artificial-intelligence-future-possibilities/>>.
3. Agency for Digital Italy. 2019. Artificial Intelligence task force [interaktyvus]. The Agency for Digital Italy, [žiūrėta 2020 m. spalio 7 d.]. Prieiga per internetą: <<https://ia.italia.it/en/>>.
4. Ahuja A. S. 2019. The impact of artificial intelligence in medicine on the future role of the physician [interaktyvus]. PMC, [žiūrėta 2020 m. spalio 5 d.]. Prieiga per internetą: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6779111/>>.
5. AI For Humanity. 2018. AI for humanity: French Strategy for Artificial Intelligence [interaktyvus]. AI For Humanity, [žiūrėta 2020 m. spalio 10 d.]. Prieiga per internetą: <<https://www.aiforhumanity.fr/en/>>.
6. AI HLEG. 2019b. Policy and investment recommendations for trustworthy AI [interaktyvus]. European Commission, [žiūrėta 2020 m. spalio 29 d.]. Prieiga per internetą: <<https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>>.
7. AI Topics. 2020. A brief history of AI [interaktyvus]. AITopics, [žiūrėta 2020 m. spalio 5 d.]. Prieiga per internetą: <<https://aitopics.org/misc/brief-history>>.
8. Amilevičius D. 2017. Dirbtinis intelektas ir besiformuojančių technologijų etika. Naujasis židinys-Aidai, Nr. 4, 19-24 p.
9. Ananny M., Crawford K. 2018. Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability [interaktyvus]. SAGE journals, [žiūrėta 2020 m. rugsėjo 27 d.]. Prieiga per internetą: <<https://journals.sagepub.com/doi/full/10.1177/1461444816676645>>.
10. Anderson J., Rainie L. 2018. Artificial Intelligence and the Future of Humans [interaktyvus]. Pew research.center, [žiūrėta 2020 m. gruodžio 3 d.]. Prieiga per internetą: <<https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/>>.
11. Andrew N. 2016. What artificial intelligence can and can't do right now [interaktyvus]. Harvard business review, [žiūrėta 2020 m. rugsėjo 10 d.]. Prieiga per internetą:

- <<https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now>>.
12. Austrian Council on Robotics and Artificial Intelligence. 2018. Die Zukunft Österreichs mit Robotik und Künstlicher Intelligenz positiv gestalten [interaktyvus]. Österreichischer Rat für Robotik und Künstliche Intelligenz, [žiūrėta 2020 d. spalio 7 d.]. Prieiga per internetą: <https://www.acrai.at/wpcontent/uploads/2019/04/ACRAI_whitebook_online_2018-1.pdf>.
 13. Balkin J. 2015. The path of robotics law [interaktyvus]. California law review, [žiūrėta 2020 m. rugjūčio 25 d.]. Prieiga per internetą: <https://www.californialawreview.org/path_robotics_law/>.
 14. Bartkus J., Stundys T. 2018. Dirbtinio intelekto teisinė atsakomybė: ko galima tikėtis?. Vilniaus universitetas: teisės mokslo pavasaris, p. 8.
 15. Beato G. 2015. Roombas in the Big House? [interaktyvus]. Reason, [žiūrėta 2020 m. spalio 10 d.], Prieiga per internetą: <<https://reason.com/2015/03/03/roombas-in-the-big-house/>>.
 16. Berg. 2020. Pancreatic cancer [interaktyvus]. Berg, [žiūrėta 2020 m. lapkričio 27 d.]. Prieiga per internetą: <<https://www.berghealth.com/research/research-development/pancreatic-cancer/>>.
 17. Bossmann J. 2016. Top 9 ethical issues in artificial intelligence [interaktyvus]. World economicforum , [žiūrėta 2020 m. rugsėjo 29 d.]. Prieiga per internetą: <<https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence>>.
 18. Bostrom N., Yudkowsky E. 2014. The ethics of artificial intelligence [interaktyvus]. MIRI, [žiūrėta 2020 m. rugsėjo 20 d.]. Prieiga per internetą: <<http://intelligence.org/files/EthicsofAI.pdf>>.
 19. Botha M. 2019. The limits of artificial intelligence [interaktyvus]. Towards data science, [žiūrėta 2020 m. rugsėjo 30d.]. Prieiga per internetą: <<https://towardsdatascience.com/the-limits-of-artificial-intelligence-fdcc78bf263b>>.
 20. Brundage M., Avin S, Clark J., Toner H., Eckersley P, Garfinkel B., Dafoe A., Scharre P, Zeitzoff T., Filar B., Anderson H., Roff H., Allen G., Steinhardt J., Flynn C., Ó hÉigeartaigh S., Beard S., Belfield H., Farquhar S., Lyle C., Crootof R., Evans O., Page M., Bryson J., Yampolskiy R., Amodei D. 2018. The malicious use of artificial intelligence: forecasting, prevention, and mitigation [interaktyvus]. Cornell University, [žiūrėta 2020 spalio 15 d.]. Prieiga per internetą: <<https://arxiv.org/abs/1802.07228v1>>.
 21. Calo R. 2018. Artificial intelligence policy: a primer and roadmap [interaktyvus]. University of California, [žiūrėta 2020 m. rugsėjo 25 d.]. Prieiga per internetą: <https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Calo.pdf>.
 22. Cellan-Jones R. 2014. Stephen Hawking warns artificial intelligence could end mankind [interaktyvus]. BBC news, [žiūrėta 2020 m. birželio 16 d.]. Prieiga per internetą:

- <<http://www.bbc.com/news/technology-30290540>>.
23. Chalmers D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. USA: Oxford University Press.
 24. Charney R. 2015. Can androids plead automatism? A review of when robots kill: artificial intelligence under the criminal law by Gabriel Hallevy. *Toronto Fac. L. Rev.* 69, [žiūrėta 2020 m. balandžio 4 d.]. Prieiga per internetą: <<https://heinonline.org/HOL/LandingPage?handle=hein.journals/utflr73&div=8&id=&page=>>.
 25. Childs M. 2011. John McCarthy: Computer scientist known as the father of AI [interaktyvus]. *Independent*, [žiūrėta 2020 m. gruodžio 6 d.]. Prieiga per internetą: <<https://www.independent.co.uk/news/obituaries/john-mccarthy-computer-scientist-known-as-the-father-of-ai-6255307.html>>.
 26. Chollet F. 2017. *Deep learning with Python* [interaktyvus]. Manning publications.
 27. Crawford K., Calo R. 2016. There is a blind spot in ai research. *Nature*, vol 538, p. 311–313.
 28. Čaplinskas A. 2015. Informacinė visuomenė, dirbtinis intelektas ir teisė [interaktyvus]. MRU, [žiūrėta 2020 m. rugsėjo 29 d.]. Prieiga per internetą: <<https://repository.mruni.eu/handle/007/14284>>.
 29. Čerka, P., Grigienė, J., Sirbikytė, G. 2015. Liability for damages caused by artificial intelligence [interaktyvus]. *Computer Law & Security Review*, [žiūrėta 2020 m. gruodžio 1 d.]. Prieiga per internetą: <<https://www.sciencedirect.com/science/article/abs/pii/S026736491500062X>>.
 30. Das Sejuti. 2020. Rome Call For AI Ethics: A humanising pledge signed by the tech giants & the catholic church [interaktyvus]. *India magazine*, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <<https://analyticsindiamag.com/rome-call-for-ai-ethics-a-humanising-pledge-signed-by-the-tech-giants-the-catholic-church/>>.
 31. DHL. 2018. Artificial intelligence in logistics [interaktyvus]. DHL Trend Research, [žiūrėta 2020 m. lapkričio 27 d.]. Prieiga per internetą: <<https://www.dhl.com/content/dam/dhl/global/core/documents/pdf/glo-core-trend-report-artificial-intelligence.pdf>>.
 32. Die Bundesregierung. 2018. Strategie Künstliche Intelligenz der Bundesregierung [interaktyvus]. Die Bundesregierung, [žiūrėta 2020 m. spalio 7 d.]. Prieiga per internetą: <<https://www.bundesregierung.de/resource/blob/997532/1550276/3f7d3c41c6e05695741273e78b8039f2/2018-11-15-ki-strategie-data.pdf>>.
 33. Digital Poland Foundation. 2019. Map of the Polish AI [interaktyvus]. DigitalPoland, [žiūrėta 2020 m. spalio 10 d.]. Prieiga per internetą: <<http://www.startupbridge.eu/wp-content/uploads/2019/02/map-of-the-polish-AI-2019.pdf>>.
 34. Dignum V. 2018. Ethics in artificial intelligence: introduction to the special issue [interaktyvus].

- Ethics and Information Technology, [žiūrėta 2020 m. rugsėjo 25 d.]. Prieiga per internetą: <<https://link.springer.com/article/10.1007/s10676-018-9450-z>>.
35. Domingos, P. 2015. The Master Algorithm: how the quest for the ultimate learning machine will remake our world. New York: Basic Books.
36. Doneda D., Almeida, V. 2016. What is algorithm governance? IEEE internet computing [interaktyvus]. ResearchGate, [žiūrėta 2020 m. spalio 26 d.]. Prieiga per internetą: <https://www.researchgate.net/publication/305801954_What_Is_Algorithm_Governance>.
37. Dubrovskiy D. 2013. Problema “soznanie i mozg”: informatsionnyiy podhod. Znanie Ponimanie Umenie , Nr. 4, p. 92–98.
38. Edwards L, Veale M. 2018. Enslaving the algorithm: from a „right to an explanation“ to a „right to better decisions“? [interaktyvus]. Cornell University, [žiūrėta 2020 m. rugsėjo 10 d.]. Prieiga per internetą: <<https://arxiv.org/abs/1803.07540>>.
39. EESRK. 2017. Dirbtinis intelektas. Jo poveikis bendrajai (skaitmeninei) rinkai, gamybai, vartojimui, užimtumui ir visuomenei [interaktyvus]. Europos ekonomikos ir socialinių reikalų komitetas, [žiūrėta 2020 m. spalio 10 d.] Prieiga per internetą: <<https://eur-lex.europa.eu/legal-content/LT/ALL/?uri=CELEX%3A52016IE5369>>.
40. EK. 2018. Dirbtinis intelektas Europai [interaktyvus]. EUR-Lex, [žiūrėta 2020 m. lapkričio 6 d.]. Prieiga per internetą: <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AC%3A2019%3A110%3AFULL>>.
41. Ekonomikos ir inovacijų ministerija. 2019. Lietuvos dirbtinio intelekto strategija. Ateities vizija [interaktyvus]. Kurk Lietuvai, [žiūrėta 2020 m. rugsėjo 30 d.]. Prieiga per internetą: <http://kurk.lt/wp-content/uploads/2019/04/DI_strategija_LT_koreguota.pdf>.
42. EPRS. 2020. The ethics of artificial intelligence: issues and initiatives [interaktyvus]. European Parliamentary Research Service, [žiūrėta 2020 m. gruodžio 15 d.] Prieiga per internetą: <[https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU\(2020\)63445_2_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)63445_2_EN.pdf)>.
43. Erdelyi O., Goldsmith J. 2018. Regulating artificial intelligence: proposal for a global solution [interaktyvus]. SSRN, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3263992>.
44. European Commission. 2012. Public Attitudes towards Robots [interaktyvus]. European Commission, [žiūrėta 2020 m. spalio 10 d.]. Prieiga per internetą: <<https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/SPECIAL/surveyKy/1044/p/3>>.
45. European Commission. 2017. Attitudes towards the impact of digitisation and automation on daily life [interaktyvus]. European Commission, [žiūrėta 2020 m. spalio 10 d.] Prieiga per

interneta:

<<https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/SPECIAL/surveyKy/2160>>.

46. European Commission. 2018 Directorate-General for Research and Innovation, Unit [interaktyvus]. European Group, [žiūrėta 2020 m. rugsėjo 25 d.]. Prieiga per internetą: <https://knowledge4policy.ec.europa.eu/organisation/dg-rtd-dg-research-innovation_en>.
47. European Commission. 2020. Artificial intelligence [interaktyvus]. European Commission, [žiūrėta 2020 m. lapkričio 20 d.]. Prieiga per internetą: <<https://ec.europa.eu/digital-single-market/en/artificial-intelligence>>.
48. European Union. 2020. AI Watch. Artificial Intelligence in medicine and healthcare: applications, availability and societal impact [interaktyvus]. JRC science for policy report, [žiūrėta, 2020 m. lapkričio 12 d.]. Prieiga per internetą: <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120214/jrc120214_ai_in_medicine_and_healthcare_report-aiwatch_v50.pdf>.
49. Europos Komisija. 2018. Dirbtinis intelektas Europai [ineraktyvus]. Briuselis, [žiūrėta 2020 m. gegužės 25d.]. Prieiga per internetą: <<https://ec.europa.eu/transparency/regdoc/rep/1/2018/LT/COM-2018-237-F1-LT-MAIN-PART-1.PDF>>.
50. Europos Parlamentas. 2017. Rezoliucija su rekomendacijomis Komisijai dėl robotikai taikomų civilinės teisės nuostatų [interaktyvus]. Europos Parlamentas, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_LT.html>.
51. Europos Sąjunga 2016. Europos Sąjungos pagrindinių teisių chartija [interaktyvus]. Europos Sąjungos oficialusis leidinys, [žiūrėta 2020 m. gegužės 25 d.]. Prieiga per internetą: <Nuoroda:<https://eur-lex.europa.eu/legal-content/LT/TXT/PDF/?uri=CELEX:12016P/TXT&from=GA>>.
52. Europos Sąjunga. 2020. Duomenų apsauga pagal BDAR [interaktyvus]. Your Europe, [žiūrėta 2020 m, gegužės 25 d.] Prieiga per internetą: <https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_lt.htm>.
53. Europos saugos ir sveikatos darbe agentūra 2021. Europos darbuotojų saugą ir sveikatą reglamentuojančios direktyvos [interaktyvus]. Europos saugos ir sveikatos darbe agentūra, [žiūrėta 2020 m. gegužės 25 d.]. Prieiga per internetą: <<https://osha.europa.eu/lt/safety-and-health-legislation/european-directives>>.

54. Floridi L., Cowls J. 2019. A unified framework of five principles for AI in society. Harvard Data Science Review [interaktyvus]. HDSR, [žiūrėta 2020 m. rugsėjo 27 d.]. Prieiga per internetą: <<https://hdsr.mitpress.mit.edu/pub/10jsh9d1/release/7>>.
55. Floris G. 2020. A pluralist account of the basis of moral status [interaktyvus]. Springer link, [žiūrėta 2020 m. gruodžio 27 d.]. Prieiga per internetą: <<https://link.springer.com/article/10.1007/s11098-020-01513-z>>.
56. Frankenfield J., Scott G. 2020. Artificial Intelligence (AI) [interaktyvus]. Investopedia, [žiūrėta 2020 m. spalio 19 d.]. Prieiga per internetą: <<https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>>.
57. Garnham A. 2017. Artificial intelligence: an introduction. Routledge : Psychology revivals.
58. Gaubienė N. 2019. Lietuvos dirbtinio intelekto strategija: ar teisingai suprantamas dirbtinis intelektas? [interaktyvus]. TeisėPro, [žiūrėta 2020 m. gegužės 15 d.]. Prieiga per internetą: <http://www.teise.pro/index.php/2019/08/26/n-gaubiene-liuvos-dirbtinio-intelektto-strategija-ar-teisingai-suprantamas-dirbtinis-intelektas/#_ftnref1>.
59. Genzelis B. 1986. Esė apie mąstytojus. Vilnius: Mintis.
60. Giubilini A., Savulescu J. 2018. The artificial moral advisor. the ‘ideal observer’ meets artificial intelligence. Philosophy & Technology, Nr. 31(2), p. 169–188.
61. GOS. 2019. National approach to artificial intelligence [interaktyvus]. Government Offices of Sweden, [žiūrėta 2020 m. spalio 8 d.]. Prieiga per internetą: <<https://www.government.se/4a7451/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf>>.
62. Hallevy G. 2010. The Criminal liability of artificial intelligence entities-from science fiction to legal social control. Akron Intellectual Property Journal, vol. 4, p 171-201.
63. Hallevy G. 2011. Unmanned vehicles: subordination to criminal law under the modern concept of criminal liability [interanktyvus]. Journal of Law, Information and Science, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <<http://www.austlii.edu.au/au/journals/JILawInfoSci/2012/12.html>>.
64. Hallevy G. 2013. When robots kill: artificial intelligence under criminal law. Boston: Northeastern University Press.
65. Harry S. 2014. Machine learning and law [interaktyvus]. Washington Law Review, [žiūrėta 2020 m. gegužės 22 d.]. Prieiga per internetą: <<https://scholar.law.colorado.edu/articles/81/>>.
66. IBM. 2019. IBM Watson Health in oncology [interaktyvus]. IBM, [žiūrėta 2020 m. lapkričio 27 d.]. Prieiga per internetą: <<https://www.ibm.com/downloads/cas/0ZRYPWL9>>.
67. IEEE. 2017. Ethically Aligned Design: First Edition. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems [interaktyvus]. The Institute of Electrical and

- Electronics Engineers, [žiūrėta 2020 m. gruodžio 1 d.] Prieiga per internetą: <<https://standards.ieee.org/news/2019/ieee-ead1e.html>>.
68. Yeung K. 2019. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework [interaktyvus]. Council of Europe, [žiūrėta 2020 m. lapkričio 13 d.]. Prieiga per internetą: <<https://rm.coe.int/responsability-and-ai-en/168097d9c5>>.
69. Igorevna O. I., Kudryashe F. V. 2017. The creative ability of artificial intelligence. Routledge, creativity studies, vol. 10(2), p. 135–144.
70. Yu H., Shen Z., Miao C., Leung C., Lesser V.R., Yang Q. 2018. Building ethics into artificial intelligence [interaktyvus]. Cornell University, [žiūrėta 2020 m. rugsėjo 10 d.]. Prieiga per internetą: <<https://arxiv.org/abs/1812.02953>>.
71. Jasanoff. S. 2016. The ethics of invention: technology and the human future [interaktyvus]. Harvard Kennedy school, [žiūrėta 2020 m. rugsėjo 25 d.]. Prieiga per internetą: <<https://www.hks.harvard.edu/publications/ethics-invention-technology-and-human-future>>.
72. JCR.2018. The impact of artificial intelligence on learning, teaching, and education [interaktyvus]. European Union, [žiūrėta 2020 m. rugsėjo d.]. Prieiga per internetą: <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113226/jrc113226_jrcb4_the_impact_of_artificial_intelligence_on_learning_final_2.pdf>.
73. Johnson D., Verdicchio M.. 2017. Reframing AI discourse. minds and machines, vol. 27, Nr. 4, p. 575–590.
74. Kallem S. 2012. Artificial intelligence algorithms [interaktyvus]. Journal of computer engineering, [žiūrėta 2020 m. gegužės 22 d.]. Prieiga per internetą: <<http://www.iosrjournals.org/iosr-jce/papers/Vol6-Issue3/A0630108.pdf>>.
75. Kurakin A, Goodfellow, I. Bengio S. 2016. Adversarial examples in the physical world [interaktyvus]. Cornell University, [žiūrėta 2020 m. lapkričio 6 d.]. Prieiga per internetą: <<https://arxiv.org/abs/1607.02533>>.
76. Kurk Lietuvai, Ekonomikos ir inovacijų ministerija. 2019. Lietuvos dirbtinio intelekto strategija [interaktyvus]. EIMIN, [žiūrėta 2020 m. spalio 30 d.]. Prieiga per internetą: <[https://eimin.lrv.lt/uploads/eimin/documents/files/DI_strategija_LT\(1\).pdf](https://eimin.lrv.lt/uploads/eimin/documents/files/DI_strategija_LT(1).pdf)>
77. Lavenda D. 2017. Vannevar Bush Points the Way to a New Era of Computing [interaktyvus]. CMS wire, [žiūrėta 2020 m. gruodžio 6 d.]. Prieiga per internetą: <<https://www.cmswire.com/digital-workplace/vannevar-bush-points-the-way-to-a-new-era-of-computing/>>.
78. LeCun Y., Bengio Y., Hinton G. 2015. Deep learning [interaktyvus]. Nature, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <<https://www.nature.com/articles/nature14539>>.

79. Lietuvos Respublikos terminų bankas. 2020. Intelektas [interaktyvus]. LRTB, [žiūrėta 2020 m. kovo 20 d.]. Prieiga per internetą: <<http://terminai.vlkk.lt/paieska?search=intelektas>>
80. Locke J. 1988. Locke: Two treatises of government student edition. England: Cambridge University press.
81. Lupkowski P. 2019. Turing's 1948 „Paper Chess Machine“ Test as a Prototype of the Turing Test [interaktyvus]. ResearchGate, [žiūrėta 2020 m. lapkričio 10 d.]. Prieiga per internetą: <https://www.researchgate.net/publication/333994449_Turing's_1948_'Paper_Chess_Machine'_Test_as_a_Prototype_of_the_Turing_Test>.
82. Magrani E. 2019. New perspectives on ethics and the laws of artificial intelligence [interaktyvus]. Interner policy review, [žiūrėta 2020 m. lapkričio 16 d.]. Prieiga per internetą: <<https://policyreview.info/articles/analysis/new-perspectives-ethics-and-laws-artificial-intelligence>>.
83. Maknickienė N. 2015. Paramos sistema investuotojui valiutų rinkoje: daktaro disertacija. Vilnius.
84. Maršalkaitė. 2018. Dirbtinis intelektas, mašininis mokymasis ir gilieji tinklai: kas slepiasi už šių savokų? [interaktyvus]. Techo, [žiūrėta 2020 m. gegužės 15 d.]. Prieiga per internetą: <<http://techo.lt/dirbtinis-intelektas-masininis-mokymasis-ir-gilieji-tinklai-kas-slepiasi-uz-siu-savoku/>>.
85. Masolo C., Sanfilippo E. 2020. Technical artefact theories: a comparative study and a new empirical approach [interaktyvus]. Review of philosophy and psychology, [žiūrėta 2020 m. gegužės 20 d.]. Prieiga per internetą: <<https://link.springer.com/article/10.1007/s13164-020-00475-9>>.
86. MEAEF. 2017. Finland's Age of Artificial Intelligence [interaktyvus]. Ministry of Economic Affairs and Employment of Finland, [žiūrėta 2020 m. spalio 30 d.]. Prieiga per internetą: <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkoju_lkaisu.pdf>.
87. MEAEF. 2018a. Artificial intelligence programme [interaktyvus]. Ministry of Economic Affairs and Employment of Finland, [žiūrėta 2020 m. spalio 30 d.]. Prieiga per internetą: <<https://tem.fi/en/artificial-intelligence-programme>>.
88. MEAEF. 2018b. Work in the age of artificial intelligence. Ministry of Economic Affairs and Employment of Finland, [žiūrėta 2020 m. spalio 30 d.]. Prieiga per internetą: <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf>.
89. Medeiros J. 2016. The startup fighting cancer with AI [interaktyvus]. Wired, [žiūrėta 2020 m. lapkričio 13 d.]. Prieiga per internetą: <<https://www.wired.co.uk/article/ai-cancer-drugs-berg-pharma-startup>>.

90. Miller K.W., Wolf M.J. ir Grodzinsky S. 2017. Why we should have seen that coming: comments on microsoft's tay „experiment,” and wider implications. *The ORBIT Journal*, vol. 1, No 2, p. 1-12.
91. Mongirdas M. 2019. Dirbtinio intelekto DI strategija ir įgyvendinimas [interaktyvus]. Kurkl.lt, [žiūrėta 2020 m. gegužės 15 d.]. Prieiga per internetą: <<http://kurkl.lt/projektai/lietuvos-dirbtinio-intelektto-strategija-ir-igyvendinimas/>>.
92. Mongirdas. 2019a. Lietuvos dirbtinio intelekto strategija: ateities vizija [interaktyvus]. MITA, [žiūrėta 2020 m. gruodžio 6 d.]. Prieiga per internetą: <<https://mita.lrv.lt/uploads/mita/documents/files/Markas%20Mongirdas.pdf>>.
93. Moody T. C. 1994. Conversations with zombies. *Journal of Consciousness Studies*, Nr. 1(2), p. 196–200.
94. Morley J., Floridi L., Kinsey L., Elhalal A. 2019. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics* [interaktyvus], Nr. 26 [žiūrėta 2020 m. rugsėjo 25 d.]. Prieiga per internetą: <<https://link.springer.com/article/10.1007/s11948-019-00165-5>>.
95. Muller C., Bostrom N. 2016. Future progress in artificial intelligence: A survey of expert opinion. *Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science)*, vol 376, p. 555-572.
96. Neff G., Nagy P. 2016. Automation, algorithms, and politics| talking to bots: symbiotic agency and the case of Tay [interaktyvus], Nr. 10(2016), [žiūrėta 2020 m. spalio 9 d.]. 4915–4931 p. Prieiga per internetą: <<https://ijoc.org/index.php/ijoc/article/view/6277>>.
97. OECD. 2020. AI policies and initiatives [interaktyvus]. OECDiLibrary, [žiūrėta 2020 m. rugsėjo 20 d.]. Prieiga per internetą: <<https://www.oecd-ilibrary.org/sites/cf3f3be0-en/index.html?itemId=/content/component/cf3f3be0-en#boxsection-d1e13778>>.
98. PAFL. 2020. Rome Call for AI Ethics [interaktyvus]. Pontifical Academy for Life, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <<http://www.academyforlife.va/content/pav/en/events/intelligenza-artificiale.html>>.
99. Paglia V 2020. Rome Call for AI Ethics [interaktyvus]. Rome call, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <<http://www.academyforlife.va/content/pav/en/events/intelligenza-artificiale.html>>.
100. Piano S. 2020. Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications* [interaktyvus], Nr. 7 (9) [žiūrėta 2020 m. rugsėjo 25 d.]. Prieiga per internetą: <<https://www.nature.com/articles/s41599-020-0501-9>>.

101. Pistono F, Yampolskiy.V. 2016.Unethical research: how to create a malevolent artificial intelligence [interaktyvus].]. Cornell University, [žiūrėta 2020 spalio 15 d.]. Prieiga per internetą: <<https://arxiv.org/abs/1605.02817>>.
102. Ramonienė M., Brazauskienė J., Burneikaitė N., Daugmaudytė J., Kontutytė E., Pribušauskaitė J. 2012. Lingvodidaktikos terminų žodynas. Vilnius : Vilniaus universitetas : Vilniaus universiteto leidykla.
103. Reed C. 2018. How should we regulate artificial intelligence? [interaktyvus]. The Royal Society, [žiūrėta 2020 m. gruodžio 6 d.]. Prieiga per internetą: <<https://royalsocietypublishing.org/doi/10.1098/rsta.2017.0360>>.
104. Rome Call for AI Ethics. 2020. Rome Call for AI Ethics [interaktyvus]. Academyforlife, [žiūrėta 2020 m. gegužės 25 d.]. Prieiga per internetą: <http://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/CALL%2028%20febraio/AI%20Rome%20Call%20x%20firma_DEF_DEF_con%20firme_.pdf>.
105. Rome call. 2020. Rome Call for AI Ethics [interaktyvus]. Rome call, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <<https://romecall.org/2020/03/02/rome-call-for-ai-ethics/>>.
106. Russel S., Norvig P. 2010. Artificial intelligence: A modern approach [interaktyvus]. AIMA, [žiūrėta 2020 m. gruodžio 6 d.]. Prieiga per internetą: <<http://aima.cs.berkeley.edu/index.html>>.
107. Samuel A. L. 1959. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, vol. 3, Nr. 2, p. 210 – 229.
108. Scherer U. 2015. Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. Harvard Journal of Law & Technology, vol. 29, No. 2, p. 48.
109. Schwartz O. 2019 In 2016, Microsoft’s racist chatbot revealed the dangers of online conversation [interaktyvus 2020 m. spalio 10 d.]. IEEE Spectrum, [žiūrėta 2020 m. gruodžio 1 d.]. Prieiga per internetą: <<https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation>>.
110. Searle, J. 1980. Minds, brains, and programs [interaktyvus]. Behavioral and Brain Sciences, žiūrėta 2020 m. gruodžio 6 d.]. Prieiga per internetą: <<https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A>>.
111. Serpa Y. 2020. The Rome Call for AI Ethics [interaktyvus]. Medium, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <<https://medium.com/towards-artificial-intelligence/the-rome-call-for-ai-ethics-2c1a00ad528>>
112. Shroff R. 2020. When Blockchain Meets Artificial Intelligence [interaktyvus]. The

- Startup, [žiūrėta 2020 m. lapkričio 24 d.]. Prieiga per internetą: <<https://medium.com/swlh/when-blockchain-meets-artificial-intelligence-e448968d0482>>.
113. Siau K., Wang W. 2018. Building trust in artificial intelligence, machine learning, and robotics [interaktyvus]. ResearchGate, [žiūrėta 2020 m. rugsėjo 15 d.]. Prieiga per internetą: <https://www.researchgate.net/publication/324006061_Building_Trust_in_Artificial_Intelligence_Machine_Learning_and_Robotics>.
- Stahl B., Timmermans J., Mittelstadt B. 2016. The ethics of computing: a survey of the computing-oriented literature. ACM Journals, ACM Computing Surveys, vol. 48, Nr. 4, p. 1–38.
114. Stanford. 2020. National artificial intelligence strategies and human rights: a review [interaktyvus]. Stanford University, [žiūrėta 2020 m. gruodžio 2 d.]. Prieiga per internetą: <<https://cyber.fsi.stanford.edu/gdpi/content/national-artificial-intelligence-strategies-and-human-rights-review>>.
115. Stanford. 2020a. Nietzsche's Moral and Political Philosophy [interaktyvus]. Stanford Encyclopedia of Philosophy, [žiūrėta 2020 m. rugsėjo 25 d.]. Prieiga per internetą: <<https://plato.stanford.edu/entries/nietzsche-moral-political/>>.
116. Sternberg R. J. 2020. Human intelligence psychology [interaktyvus]. Britannica, [žiūrėta 2020 m. lapkričio 12 d.]. Prieiga per internetą: <<https://www.britannica.com/science/human-intelligence-psychology>>.
117. Stilgoe J., Winfield A. 2018. Self-driving car companies should not be allowed to investigate their own crashes [interaktyvus]. The Guardian, [žiūrėta 2020 m. spalio 1 d.]. Prieiga per internetą: <<https://www.theguardian.com/science/politicalscience/2018/apr/13/self-driving-car-companies-should-not-be-allowed-to-investigate-their-owncrashes>>.
118. Swedish AI Council. 2019. Swedish AI Council, [žiūrėta 2020 m. spalio 7 d.]. Prieiga per internetą: <<https://swedishaicouncil.com>>.
119. Štilis D., Kiškis M., Limba T., Rotomskis I., Agafonov K., Gulevičiūtė G., Panka K. 2016. Interneto ir technologijų teisė. Kaunas: MRU.
120. Taddeo M., Floridi L. 2018: Regulate artificial intelligence to avert cyber arms race [interaktyvus]. Nature, [žiūrėta 2020 m. spalio 11 d.]. Prieiga per internetą: <<https://www.nature.com/articles/d41586-018-04602-6>>.
121. Task Force on Artificial Intelligence of the Agency for Digital Italy. 2018. White paper on artificial intelligence at the service of citizens [interaktyvus]. Agenzia per l'Italia Digitale, [žiūrėta 2020 m. spalio 8 d.]. Prieiga per internetą: <<https://libro-bianco-ia.readthedocs.io/en/latest/>>.
122. TAT Institute. 2019. What Alan Turing means to us [interaktyvus]. The Alan Turing

- Institute, [žiūrėta 2020 m. rugsėjo 10 d]. Prieiga per internetą: <<https://www.turing.ac.uk/blog/what-alan-turing-means-us>>.
123. The Danish Government. 2018. Strategy for Denmark's Digital Growth [interaktyvus]. Strategy for Denmark's digital growth, [žiūrėta 2020 m. spalio 16 d.]. Prieiga per internetą: <https://eng.em.dk/media/10566/digital-growth-strategy-report_uk_web-2.pdf>.
124. The Danish Government. 2019. National Strategy for Artificial Intelligence [interaktyvus]. Strategy for Denmark's digital growth, [žiūrėta 2020 m. spalio 16 d.]. Prieiga per internetą: <https://eng.em.dk/media/13081/305755-gb-version_4k.pdf>.
125. The Oxford English Dictionary. 2020. Definition of artificial intelligence noun [interaktyvus]. Oxford Learner's Dictionary, [žiūrėta 2020 m. gruodžio 6 d.]. Prieiga per internetą: <<https://www.oxfordlearnersdictionaries.com/definition/english/artificial-intelligence?q=artificial+intelligence>>.
126. Thomas M. 2019. 6 Dangerous risks of artificial intelligence [interaktyvus]. Built in, [žiūrėta 2020 m. spalio 7 d.]. Prieiga per internetą: <<https://builtin.com/artificial-intelligence/risks-of-artificial-intelligence>>.
127. UNESCO Publishing, Netexplo. Human decisions thoughts on AI [interaktyvus]. UNESDOC Digital Library, [žiūrėta 2020 m. spalio 4 d.]. Prieiga per internetą: <<https://unesdoc.unesco.org/ark:/48223/pf0000261563>>.
128. UNICEF. World economic forum. 2018. Children and AI where are the opportunities and risks? [interaktyvus]. UNICEF, [žiūrėta 2020 m. birželio 10 d.]. Prieiga per internetą: <https://www.unicef.org/innovation/sites/unicef.org/innovation/files/2018-11/Children%20and%20AI_Short%20Version%20%283%29.pdf>.
129. Verbeek, P. 2011. Moralizing technology: understanding and designing the morality of things [interaktyvus]. The University of Chicago Press, [žiūrėta 2020 m. spalio 20 d.]. Prieiga per internetą: <<https://press.uchicago.edu/ucp/books/book/chicago/M/bo11309162.html>>.
130. Vermaas, P.; Kroes, P., Poel, I., Franssen, M., Houkes, W. 2011. A philosophy of technology: from technical artefacts to sociotechnical systems [interanktyvus]. Morgan & Claypool, [žiūrėta 2020 m. birželio 7 d.]. Prieiga per internetą: <<https://ieeexplore.ieee.org/document/6812712>>.
131. Villani, C. 2018. For a meaningful artificial intelligence: towards a French and European strategy [interaktyvus]. Aiforhumanity, [žiūrėta 2020 m. spalio 7 d.]. Prieiga per internetą: <https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf>.
132. Vincent C 2020. Ethics of Artificial Intelligence and Robotics [interaktyvus]. Stanford Encyclopedia of Philosophy, [žiūrėta 2020 m. rugsėjo 25 d.]. Prieiga per internetą: <<https://plato.stanford.edu/entries/ethics-ai/>>.

133. Vincent J. 2019. AI systems should be accountable, explainable, and unbiased, says EU [interaktyvus]. The Verge, [žiūrėta 2020 m. rugspėjo 30 d.]. Prieiga per internetą: <<https://www.theverge.com/platform/amp/2019/4/8/18300149/eu-artificial-intelligence-ai-ethical-guidelines-recommendations>>.
134. Vladeck, D. C. (2014). Machines without principals: liability rules and artificial intelligence [interaktyvus]. Washington Law Review, [žiūrėta 2020 m. spalio 20d.]. Prieiga per internetą: <<https://digitalcommons.law.uw.edu/wlr/vol89/iss1/6/>>.
135. Wakefield K. 2020. Artificial Intelligence in business [interaktyvus]. SAS, [žiūrėta 2020 m. gegužės 12 d.]. Prieiga per internetą: <https://www.sas.com/en_gb/insights/articles/analytics/applications-of-artificial-intelligence.html>.
136. Wiener N. 1988. The human use of human beings: cybernetics and society. New York: Da Capo Press.
137. Wolf M., Miller K., Grodzinsky F. 2017. Why we should have seen that coming: comments on Microsoft's „Tay“ experiment, and wider implications [interaktyvus]. Sciencedirect, [žiūrėta 2020 m. gruodžio 1 d.]. Prieiga per internetą: <<https://www.sciencedirect.com/science/article/pii/S2515856220300493?via%3Dihub>>.