

Vilniaus Universitetas

TARPTAUTINIŲ SANTYKIŲ IR POLITIKOS MOKSLŲ INSTITUTAS

Tarptautinių santykių ir diplomatijos magistro programa

KAROLINA RIMKUTĖ

II kurso studentė

Politinė kova dėl dirbtinio intelekto kaip etinė problema

MAGISTRO DARBAS

Darbo vadovas: prof. dr. Tomas Janeliūnas

Vilnius, 2022 sausio 10 d.

Magistro darbo vadovo išvados dėl darbo gynimo:

.....
.....
.....
.....

(data) (v., pavardė) (parašas)

Magistro darbas įteiktas gynimo komisijai:.....

(data) (Gynimo komisijos sekretoriaus/ės parašas)

Magistro darbo recenzentas:

(v., pavardė)

Magistro darbų gynimo komisijos įvertinimas:

.....

Komisijos pirmininkas/ė:

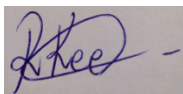
Komisijos nariai:

PATVIRTINIMAS APIE ATLIKTO DARBO SAVARANKIŠKUMĄ

Patvirtinu, kad įteikiamas darbas (*Politinė kova dėl dirbtinio intelekto kaip etinė problema*) yra:

- . Atliktas mano pačios ir nėra pateiktas kitam kursui šiame ar ankstesniuose semestruose;
- . Nebuvo naudotas kitame Institute/Universitete Lietuvoje ir užsienyje;
- . Nenaudoja šaltinių, kurie nėra nurodyti darbe, ir pateikia visą panaudotos literatūros sąrašą.

Karolina Rimkutė



Rimkutė K. Politinė kova dėl dirbtinio intelekto kaip etinė: Tarptautinių santykių ir diplomatijos specialybės, magistro darbas / VU Tarptautinių santykių ir politikos mokslų institutas; darbo vadovas T. Janeliūnas. – V., 2022. – 72 p.

Reikšminiai žodžiai: dirbtinis intelektas, politika, etika, tarptautiniai santykiai, galia

Šiame darbe nagrinėjama dirbtinio intelekto ir etikos santykio problema tarptautiniuose santykiuose. Dirbtinis intelektas yra technologija, ženkliai keičianti mūsų kasdienybę – jos pagalba tikimasi išspręsti aibę problemų, su kuriomis susiduria šiuolaikinė visuomenė. Tuo pačiu metu suvokiama, kad DI – tai technologija, kelianti grėsmę pamatinėms demokratinėms vertybėms. Būtent todėl ES aktyviai bando kontroliuoti DI panaudojimą. Darbo ašis sukasi apie tai, kad nors pagrindinės Vakarų organizacijos (ES, NATO) skiria nemažą žmogiškąjį ir finansinį kapitalą, siekdamas reguliuoti DI naudojimą, tuo pačiu metu nepakankamai kalbama apie problemas susijusias su DI sukūrimu. Darbe atskleidžiamos esminės su DI sukūrimu susijusios etinės problemos: diskriminaciniai klasifikavimo algoritmai, duomenų gavimo problema, DI sąsajos su ekologine krize. Teigiama, kad etikos ignoravimas DI sukūrimo proceso metu yra tiesiogiai susijęs su požiūriu į DI, kuris leidžia politiniame kontekste DI matyti kaip universalų galios įrankį. Toks DI supratimas padaro jį vienu iš prioritetinių tikslų pagrindinių pasaulinių galios centrų konkurencinėje kovoje. Etinis šio proceso reguliavimas nėra parankus nei vienam iš galios centrų, todėl dažnai vietoje realių problemų sprendimų susiduriame tik su tariamai „etiška“ politikų ir verslo pasaulio atstovų retorika.

TURINYS

Įvadas.....	6
1. Kas yra DI?.....	12
1.1. Formalus apibrėžimas.....	12
1.2. DI kaip kūnas ir galios įrankis.....	15
2. DI kaip etinė problema galios centrų kovos kontekste.....	18
3. Etinis reguliavimas DI panaudojimo stadijoje.....	23
3.1. Europos Sąjungos požiūris į DI: DI strategija.....	23
3.2. NATO požiūris į DI: DI strategija.....	25
4. Etinio reguliavimo stoka DI sukūrimo stadijoje.....	28
5. Pagrindinės su DI sukūrimu susijusios etinės problemos.....	30
5.1. Atsakomybės problema.....	30
5.2. Duomenų tvarkymo problema.....	34
5.3. Ekologijos problema.....	37
6. Etinio DI kūrimo proceso reguliavimo problema galios centrų konkurencijos kontekste.....	39
Išvados.....	43
Literatūros sąrašas.....	45
1 priedas.....	53
2 priedas.....	57
3 priedas.....	63
4 priedas.....	68
Summary.....	72

ĮVADAS

Tyrimo aktualumas. Mašinos be vairuotojų važinėja miesto gatvėmis, robotai pristato maistą išalkusiems gyventojams, kompiuteriai prognozuoja ligas, rašo populiarius straipsnius, atlieka konsultantų vaidmenį bei „dirba“ teismuose vos ne pilnu etatu – tai ne Aizeko Azimovo romanas, o visai netolima ateitis. Tačiau į Google naršyklę įvedus „dirbtinis intelektas“ ir paspaudus keletą nuorodų, pasidarys aišku, kad tai tik pradžia. Mokslininkai žada, kad dirbtinis intelektas (toliau – DI) pakeis gydytojus, prognozuos ateitį, darys rimtus atradimus ir panašiai. Didžiosios valstybės ir transnacionalinės korporacijos taip pat nestovi nuošalyje – milijardai dolerių investuojami į šią sritį (Analytics Insight, 2021). Pavyzdžiui, 2020 metų duomenimis, privačios investicijos į DI JAV sudarė 23,6 mlrd. dolerių, Kinija buvo antroje vietoje, jos investicijos siekė 9,9 mlrd. JAV dolerių (Liu 2021). Investicijos į DI siekia milijardus ir toliau auga. Tačiau tokie pinigai investuojami ne tik dėl to, kad tikimasi, jog DI galu gale atsipirks. Jau nekyla abejonių, kad lenktynės dėl DI yra nacionalinio saugumo klausimas (National Security Commission on Artificial Intelligence, 2021). Taigi konkurencija nebėra tik ekonominė ar technologinė, bet ir geopolitinė. Tą rodo ne tik skaičiai, bet ir didžiųjų valstybių pareiškimai, kaip kad 2017 m. Rusijos prezidento Vladimiro Putino komentaras: „Kas pirmaus dirbtinio intelekto srityje, tas taps pasaulio valdovu“ (TACC 2017). „Mąstanti mašina“ ne tik palengvina eilinio piliečio ar bankininko gyvenimą – ne kuo blogiau ji tvarkosi ir su karine technika, ir su kibernetiniu saugumu. Tad iš vienos pusės lenktynės darosi vis karštesnės, iš kitos – kelia rimtą susirūpinimą dėl nederamo DI panaudojimo pasekmių.

JAV ir ES bando padaryti taip, kad DI naudojimas neprieštarautų pagrindinėms demokratinėms vertybėms. Bandoma apsaugoti asmens laisvę, sprendžiamas atsakomybės klausimas, ieškoma būdų kaip apsaugoti visuomenę nuo „išankstinio robotų nusistatymo“ ir panašiai. Vienaip ar kitaip visi šie klausimai priklauso etikos sričiai. Tad galima sakyti, kad aktyviai dirbama ties etiniu DI panaudojimo klausimu. Tačiau skiriant tiek dėmesio rezultatui, kur kas mažiau kalbama apie kelią, kurį reikia nueiti norint jį pasiekti. O kelias visai nėra lengvas. Griežtos globalios konkurencijos sąlygos verčia varžovus kiek atlaidžiau žiūrėti į kaštus, reikalingus norint pasiekti geidžiamą tikslą. Visa tai primena ginklavimosi varžybas, vykusias tarp JAV ir Sovietų Sąjungos. Nepaisant to, kad nuolat buvo pasirašomos sutartys, ribojančios ginklų panaudojimą, paties ginklų kūrimo proceso vis nepavykdavo tinkamai sureguliuoti (Council on Foreign Relations 2021). Tad nenuostabu, kad kuriant bei išbandant naujas raketas buvo daroma žala ir paprastiems žmonėms, ir gamtai. Yra pagrindo manyti, kad ne mažiau žalos kelia DI sukūrimo procesas.

Visų pirma, reikia atkreipti dėmesį į tai, jog DI nėra technologija, egzistuojanti nieko nekainuojančio kodo pavidalu. DI turi kūną – galingus kompiuterius, kurie sunaudoja nemažai

elektros energijos bei prisideda prie vis spartėjančio klimato kaitos proceso. Antra, DI nėra *a priori* „protingas“ – jį reikia apmokyti, o tam reikia duomenų. Tačiau kas bus, jei Misteris X nenorės, kad jo duomenys būtų naudojami šiam tikslui? Kaip šiuo atveju turi būti sprendžiamas privatumo klausimas? Ir pagaliau, klausimas, kuris svarbus tiek DI sukūrimo, tiek naudojimo procese: nelygybės klausimas. Kuriant DI naudojami tokie skambūs žodžiai kaip „inovacijos“, „šviesi ateitis“, „nauji senų problemų sprendimai“ ir panašiai. Tačiau visa tai liečia tik labai nedidelę pasaulio žmonių dalį – intelektualinį, politinį bei ekonominį elitą. Tuo tarpu kitų žmonių likimas nebeatrodo taip šviesiai, įskaitant ir paprastus buhalterius ar vairuotojus, kurie, galiausiai, gali netekti darbo bei būti išstumti iš darbo rinkos.

Visa tai yra DI sukūrimo kaina. Atrodytų, kad reikėtų būti labai atsargiems einant šia linkme, tačiau, kaip jau buvo sakoma, DI yra ne eilinė technologija tarp kitų, – tai yra reikšmingas galios įrankis, todėl jo sukūrimo kaina nueina į antrą planą. Magistrinis darbas yra skirtas būtent šiai temai.

Šiame darbe bus bandoma atskleisti, kuria prasme galima kalbėti apie galios centrų siekį išnaudoti DI kaip galios įrankį, leidžiantį didinti savo galią – politinę, ekonominę, karinę, socialinę, geopolitinę, taip pat parodant, kaip šiame kontekste yra pažeidžiamos etinės normos. Darbe atsispindi, kad šiuo metu pagrindinis dėmesys yra kreipiamas į DI reguliavimo etapą, kai tuo tarpu nepakankamai skiriama dėmesio su DI sukūrimu susijusioms etinėms problemoms. Šis sukūrimo ir reguliavimo etapų išskyrimas reikšmingas tuo, kad leidžia atskirai aptarti galios centrų interesą ignoruoti pirminę DI vystymosi stadiją, sutelkiant dėmesį į reguliavimą, nors tai turės pasekmių toliau vystant šią technologiją ir diegiant globalias etines taisykles. Šiuo metu galima išskirti Kiniją, ES ir JAV kaip aktyviausius DI atžvilgiu globalias ambicijas turinčius veikėjus. Kalbėti apie globalų DI režimą dar anksti, nes net transatlantinių partnerių vizijos skiriasi. Todėl savo darbe naudoju besiformuojančio požiūrio į DI Vakaruose sampratą, kurią siūlo ES ir NATO, ne tik dėl to, kad jos labiausiai pažengusios, taikant etikos principus DI vystyme, bet ir dėl užčiuopiamų ženklų, kad griežtesnio reguliavimo valdymą ateityje galėtų taikyti ir JAV. Darbe pateikiamos šių svarbiausių Vakarų organizacijų paskelbti pagrindiniai DI strateginiai dokumentai. Jų lyginimas tarpusavyje ir bendrame geopolitiniame kontekste, kur pasireiškia ir nevalstybiniai veikėjai (kaip kad technologijų kompanijos), suponuoja įtampas, kylančias dėl DI matymo kaip galios įrankio ir siekio, kovų dėl diskurso kontekste, paversti savo valdymo modelį visuotinai dominuojančia galia.

Tyrimo prielaida. Darbe remsiuosi politikos kaip galios centrų kovos samprata. Iš vienos pusės, prielaida leidžia paaiškinti DI sukūrimo procesą kaip nulemtą galios centrų interesų. Iš kitos pusės, tai leidžia išryškinti etinio reguliavimo klausimą kaip vieną iš fundamentalių visuomenės laisvės apsaugos sąlygų.

Tyrimo problema. Kaip DI kaip galios įrankio supratimas koreliuoja su etiniu jo sukūrimo proceso reguliavimu?

Tyrimo tikslas. Atskleisti pagrindines etines dilemas DI sukūrimo stadijoje bei jų sąsają su galios centrų tikslais.

Tyrimo tikslui pasiekti keliami šie uždaviniai:

- Pirma, apibrėžti DI sąvoką.
- Antra, apibrėžti teorines prieigas prie DI sukūrimo etinio reguliavimo problemos rėmus.
- Trečia, analizuoti esamą situaciją DI etinio panaudojimo srityje.
- Ketvirta, artikuliuoti etinio reguliavimo DI sukūrimo proceso metu stoką kaip problemą.
- Penkta, ištirti kaštus DI sukūrimo etape.
- Šešta, atskleisti etinio reguliavimo DI sukūrimo proceso metu stokos sąsajas su DI kaip galios įrankio samprata.

Darbo struktūra.

Pirmame skyriuje apibrėžiama DI sąvoka. Šioje dalyje remiamasi Jack Copelando straipsniu „Artificial intelligence“ (Copeland 2021) bei Kate Crawford knyga *Atlas of AI* (Crawford 2021), kuriuose išskiriami techniniai ir politiniai apibrėžimo aspektai. Antrame skyriuje apibrėžiamos galios ir etikos sąvokos bei bandoma pateikti teorinį modelį galintį paaiškinti esamą situaciją DI sukūrimo srityje. Pasitelkiama tokiais autoriais kaip Joseph Nye, Michel'is Foucault'o, Judith Butler, Zygmuntas Baumanas, Friedrichas Nietzsche. Trečiame skyriuje analizuojama esama situacija DI etinio panaudojimo srityje. Centriniai šios analizės dalies tekstai – Europos Sąjungos (toliau – ES) ir NATO DI strategijos. Ketvirtame skyriuje bandoma atskleisti etinio reguliavimo stoką DI sukūrimo proceso metu kaip problemą. Penktame skyriuje kalbama apie DI sukūrimo kaštus, atskirai nagrinėjamos „moralinės“ ir „materialinės“ su atsakomybe susijusios problemos. Šeštame skyriuje surinktas faktinis materialas susiejamas su teorinėmis darbo prielaidomis.

Tyrimo tezė. DI sukūrimo proceso etinio reguliavimo stoka koreliuoja su galios centrų siekiu didinti savo įtaką naudojant DI kaip įrankį.

Tyrimo metodas. Darbas parašytas taikant diskurso analizės metodą. Papildomai, remiantis NATO DI strategijoje išskiriamomis interesų grupėmis, įskaitant civilinį privatų sektorių, akademinę bendruomenę, kaip turinčiomis reikšmingą vaidmenį kuriant DI (NATO 2021), darbo tikslui atskleisti surinkti ir darbe naudoti pusiau struktūruoti interviu. Jie buvo organizuoti, siekiant surinkti tyrimui reikalingos informacijos. Ši medžiaga atliko meta komentaro vaidmenį, padėjusį išryškinti kontraversijas ir atskleisti naujas klausimo matymo perspektyvas. Klausimai rengti atspindint darbo uždavinius ir tikslą, o jų formuluotės derintos su ekonomikos atašė Lietuvos Nuolatinėje atstovybėje Briuselyje Mindaugu Simonaičiu, rengiant pilotinį tyrimą, bei Briuselyje veikiančios konsultacijų įmonės Cream Consulting duomenų mokslininku (ang. data scientist) Robin Munier. Transkribuotus pokalbius originalo kalba galima rasti prieduose.

Atliktų tyrimų apžvalga. DI yra technologijos šaka, būdinga dirbtinai intelektualioms sistemoms. Nagrinėta literatūra parodė, kad yra tipiška reikšti susirūpinimą, reaguojant į naujų technologijų gimimą. Niekam nekelia abejonų, kad tokios technologijos kaip branduolinis ginklas kelia etines ir politines diskusijas, siekiama politiškai kontroliuoti tokių technologijų vystymosi kryptį, kad tai galiausiai nesukeltų grėsmės žmonijos išlikimui. Atitinkamai, vis kuriamos technologijos kelia naujus iššūkius esamoms nusistovėjusioms normoms, todėl yra siekiama technologijų vystymosi ir visuomenės santykių pokytį apibrėžti atitinkamu reglamentavimu.

Profesoriaus Vincento C. Müllerio 2020 m. publikacija Stanfordo filosofijos enciklopedijoje „Ethics of Artificial Intelligence and Robotics” (Müller 2021) aktualizuoja DI kaip skaitmeninę technologiją, kuri turės ženkliai įtaką žmonijos vystymuisi artimiausioje ateityje. Tolesniam darbui svarbu, kad autorius padeda apsibrėžti, kokios DI problemos yra laikomos etinėmis problemomis. Pasak autoriaus, medijos daugiausia dėmesio skiria rizikoms ir saugumui, poveikio prognozavimui, formuojant įspūdį, kad mes žinome, kas yra etiška ir kaip to pasiekti, belieka susitarti dėl techninių aspektų. Toks visuomenės nuomonės formavimas daro tiesioginę įtaką politikai ir pramonei, kuri prisitaikydama prie naujų tendencijų save ima pozicijuoti kaip „etišką“ (klimato kaitos kontekste, tai naujoji „žalia“), taip plaudama pačią etikos sampratą. Skirtį, remiantis Mülleriu, brėžia tai, kad DI etikos problema suponuoja, jog nėra iš karto žinoma, ką daryti yra teisinga. Šia prasme, problema yra ne DI taikymas pramonėje (darbų praradimai; DI mašinų taikymas karo lauke), o klausimas, ar leistina DI naudoti šiose srityse tam tikromis aplinkybėmis.

Į tai, kad nepakanka ieškoti techninių sprendimų, sprendžiant su DI susijusias etines problemas, taip pat gilinasi Kate Crawford, buvusi „Microsoft Research New York City“ vyriausioji mokslo tyrėja, šių metų pradžioje išleidusi plačiai DI sferoje aptarinėjamą knygą *Atlas of AI* (Crawford 2021). Norėdama parodyti, kad DI nėra neutralios, objektyvios techninės sistemos, taip pat kad tai nėra protingi agentai, kurie priima geriausią sprendimą visose situacijose, siūlo DI analizuoti plačiau: per resursus, darbo jėgą, duomenis. Kadangi DI remiasi šiais dalykais, ji daro išvadą, kad šia prasme jis nėra dirbtinis. Tačiau ir ne intelektualus – nes tai ne smegenys, ne žmogus, kurį galėtum išmokyti. Nagrinėdama ryšį tarp aplinkosaugos, darbo jėgos ir duomenų apsaugos, tokias problemas kaip DI sukūrimo kaštai, ji parodo, kad pastarieji yra sąmoningai nutylimi, akcentuojant tik techninę DI pusę. Todėl tyrėja siūlo į DI žiūrėti kontekstualiai, įvertinant visą grandinę – nuo materialijų išteklių išgavimo, ten dažnai nelegaliai veikiančių darbininkų, vėliau – menkai apmokamos darbo jėgos, pajungtos kurti įspūdį, kad sistemos autonomiškos ir intelektualesnės nei yra šiai dienai, iki naudos gavėjų, technologijų kompanijų, kurios nenori prisiimti atsakomybės už industrijai naudojamus kaštus, žalingos įtakos gamtai ir bendrą hierarchinės sistemos reproduktivumo poveikį visuomenei. Dėl šių susiklosčiusių aplinkybių autorė gilinasi į tai, kam naudinga esama sistema ir kas yra grandinės apačioje.

Nagrinėjant, kaip technologijos keičia santykį tarp galios polių ir kaip tai paveiks visuomenę, aiškėja, kad DI etikos tyrinėtojus daugiau mažiau galima suskirstyti į dvi grupes. Pirmieji, manydami, kad įmanoma sukurti DI prilygstantį žmogaus protui, ar net peržengiantį jo galimybes, kalba apie *galimas* etines problemas ateityje. Antrieji, labiau yra susirūpinę dabartine situacija dėl taikomojo DI (ang. applied AI).

Pirmajai priklausantys mokslininkai, įskaitant Stepheną Hawkingą ir Maxą Tegmarką (Tegmark 2016) mano, kad superintelligentiškos mašinos yra įmanomos. O jų sukūrimo pasekmės gali būti „arba geriausia, arba blogiausia, kas kada nors nutiko žmonijai“ (Hern 2016). Nors DI siejamas su neužtikrintumu ir nežinomybe, vis tiek matome gana platų sutarimą, jog DI radikalčiai palies kiekvieną žmonijos gyvenimo aspektą. Filosofas Nickas Bostromas knygoje *Superintelligence: Paths, Dangers, Strategies* (Bostrom 2014) taip pat sprendžia šį klausimą, leisdamasis į diskusiją apie transhumanizmą. Pagrindinis jo knygos argumentas: pirmasis sukurtas superintelektas turės lemiamą pirmaujančiojo pranašumą, o pasaulyje, kuriame nėra kitos prilygstančios sistemos, jis bus labai galingas. Atitinkamai, tokia sistema formuos pasaulį pagal savo „nuostatas“ ir sugebės įveikti bet kokią žmonių pasipriešinimą. Autorius prognozuoja, kad bus labai sunku – ar net neįmanoma – sukurti superintelektą, kurio nuostatos būtų draugiškos žmonėms arba būtų valdomos. Mūsų galimybė išsigelbėti tuo tarpu yra matoma per „netiesioginį normatyvumą“ ir „nuosekliai ekstrapoliuotą valią“, kai pasinaudosime pačios dirbtinės sistemos intelektu, kad pasiektume naudingų rezultatų, kurių patys negalime matyti ar dėl kurių negalime susitarti iš anksto. Tyrėjas pabrėžia, kad iššūkis yra „išlaikyti savo žmogiškumą: išlaikyti savo mentalinį ir emocinį stabilumą“. Jis rekomenduoja moksliniams tyrimams vadovauti ir juos valdyti laikantis griežtos etikos principų. Vis tik, jo teigimu, mums prireiks išmaniųjų technologijų, kad galėtume susidoroti su iššūkiais, kiliančiais artimiausiu metu. Kitaip tariant, prieinama prie tam tikros rizikos pusiausvyros.

Antroji tyrėjų grupė labiau linkusi kalbėti apie DI dabartį ir kaip tai, kas daroma dabar, galėtų paveikti mūsų ateitį. Šie tyrėjai skiria nemažą dėmesį šalutinių poveikių ir būsimų masinio algoritmų diegimo rizikoms vertinti. Analitikai, tokie kaip Aneeshas Aneeshas iš Stanfordo universiteto numato, kad algoritmai perims viešąją ir privačią veiklą naujoje „algokratinio valdymo“ eroje, kuri pakeis „biurokratinę hierarchiją“ (Aneesh 2002). Kiti, kaip kad Harvardo profesorė Shoshana Zuboff, kalba apie „stebėjimo kapitalizmo“ iškilimą, kuris diktuoja ekonominį elgesį „informacinėje civilizacijoje“ (Zuboff 2019). Gilinantis į DI sukūrimo žalą galima rasti keletą vertingų studijų, kur tyrėjai, suvokdami algoritmų pažeidimus, įskaitant galimybę juos išnaudoti siaurų interesų taikymui, leidžiasi į diskusiją apie tai, kaip kurti visuomenei naudingą DI. Pavyzdžiui, Lee Rainie ir Janna Anderson studija „Code-Dependent: Pros and Cons of the Algorithm Age“ (Rainie, Anderson 2017) ar kt. Į sukūrimo stadiją gilinasi, todėl darbu aktuali, prancūzų tyrėjų Lê Nguyễn Hoang ir El Mahdi El Mhamdi knyga *Le fabuleux chantier : Rendre l'intelligence artificielle robustement bénéfique*

(Hoang, Mhamdi 2019), kur šiuolaikiškai, taikant holistinio metodo kryptį, kalbama apie DI naudą (bendra algoritmo prasme). Jų modelis tarpdiscipliniškai įtraukia psichologijos, filosofijos, sociologijos ir kompiuterių mokslo aspektus. Autoriai laikosi prielaidos, kad algoritmai mus supa visur kasdieniame gyvenime ir daro įtaką milijardų žmonių sprendimams, todėl svarbu diskutuoti apie socialinį jų poveikį ir tokią algoritmų koncepciją, pagal kurią jų „elgesys“ atitiktų vertybes, kurias norėtume įtraukti į DI metodus.

Vis tik analizuojant, kaip atsiskleidžia DI etinės problemos Vakarų organizacijų – NATO ir ES – diskusijose, darbe gilinamasi į ES ir NATO DI strategijas. Iš vienos pusės, šis lyginimas leidžia kalbėti apie Vakaruose besiformuojantį požiūrį į DI. Kita vertus, jis atskleidžia, kad nors abu dokumentai postuluoja bendras vertybes, įskaitant žmogaus teises, orientaciją į žmogų, jie iš principo neįtraukia kaštų dalies. Tai leidžia kalbėti apie šių Vakarų organizacijų etinio reguliavimo stoką DI sukūrimo stadijoje ir su tuo susijusias problemas.

Lietuvoje yra nemažai autorių vienaip ar kitaip analizuojančių technologijų keliamus etinius iššūkius. Būtų galima išskirti šiuos tyrėjus, prisidėjusius prie supratimo apie DI keliamus iššūkius ir galimybes: Doc. Dr. Nijolė Maknickienė (Maknickienė, Urbonavičiūtė 2020), Prof. Dr. Artūras Serackis (Serackis, Jonkus 2014). Nepaisant to, kad apie DI yra rašoma ir kalbama, vis dėlto tyrimų skirtų etinio DI sukūrimo stadijos reguliavimo problemai rasti nepavyko. DI kaip tarptautinių santykių fenomenas plačiai aptarinėjamas ir konferencijų formuose. Pavyzdžiui, 2021 m. spalį Klaipėdos universitete vyko tarptautinė nuotolinė konferencija „Visuomenės gerovė, dirbtinis intelektas ir etiniai iššūkiai“, kur savo pranešimus skaitė Prof. Dr. Algimantas Kirkutis, Prof. Dalia Marija Stančienė; lapkritį Vilniuje vyko konferencija AI Boost 2021, kur be užsienio ekspertų savo pranešimus skaitė prof. habil. dr. Antanas Čenys, Dr. Jurgita Kapočiūtė-Dzikienė, Dr. Tadas Žižiūnas. Čia galima paminėti kolegas Justiną Lingevičių (Lingevičius 2021) ir Justiną Stanevičių (Stanevičius 2019), kurie nagrinėja DI klausimą saugumo ir tarptautinio bendradarbiavimo aspektu.

1. KAS YRA DI?

1.1. FORMALUS APIBRĖŽIMAS

Jau buvo minėta, kad DI turi būti suprastas ne vien kaip kompiuterinė programa, bet ir kaip galios įrankis. Pirmas DI apibrėžimas susijęs su filosofine bei psichologine intelekto *per se* samprata bei jo techninėmis sukūrimo sąlygomis. Antras, su politiniais DI sukūrimo motyvais. Būtent šių dviejų koncepcijų aptarimui skirtas šis skyrius.

Kalbėdami apie DI susiduriame su painiava, kurią kelia žodžių algoritmas, DI ir mašininis mokymas (toliau – MM) netinkamas vartojimas. Nors pagrįsde darbas referuoja į DI, aiškumo dėlei prasminga, visų pirma, pateikti jų visų apibrėžimus. Remiantis Aileen Scott, algoritmas yra automatinio nurodymo forma. Jis gali būti paprastų pavienių „jei-tada“ teiginių seka, pavyzdžiui, paspaudus šį mygtuką, atliekamas tas veiksmas, arba kartais tai gali būti sudėtingesnės matematinės lygtys (Scott 2021). Jack Copeland, savo straipsnyje apie DI, įėjusio į Encyclopædia Britannica (Copeland 2021), pateikia tokį apibrėžimą: „Dirbtinis intelektas – skaitmeninio kompiuterio ar kompiuteriu valdomo roboto gebėjimas atlikti užduotis, paprastai susijusias su protingomis būtybėmis. Techniškai tariant, DI – tai yra algoritmų rinkinys, galintis susidoroti su nenumatytais aplinkybėmis (t.y. gali modifikuoti savo algoritmus ir kurti naujus, reaguodamas į išmoktas įvestis ir duomenis, o ne pasikliauti vien įvestimis). Jis skiriasi nuo MM tuo, kad gali būti tiekiamas nestruktūrizuotais duomenimis ir toliau veikti. Juk struktūrizuoti duomenys – tai informacija, kuri lengvai gali būti suskirstyta į lenteles. Dažniausiai tai informacija, pasirodanti skaičių ir raidžių pavidalu. Ji lengvai surandama ir grupuojama („Excel“ tipo failai yra puikus šios informacijos pavyzdys). Nestruktūrizuoti duomenys – tai informacija, kuri nelengvai pasiduoda paieškos mechanizmams, ji beveik negali būti suskirstyta į lenteles. Priešingai nei struktūrizuoti duomenys, šią informaciją nelengva panaudoti tiesiogiai, prieš tai ji turi būti interpretuota. Tokios informacijos formatų tipų yra daugybė – tai gali būti paveiksliukai, audio įrašai, filmai, socialinių tinklų įrašai ir t.t. (Taylor 2021). Tai, kad DI sugeba apdoroti nestruktūrizuotą informaciją, suteikia jam didžiulį pranašumą prieš programas, dirbančias tik su struktūrizuotais duomenimis.

Kas liečia MM – tai yra algoritmų rinkinys, leidžiantis kompiuteriams mokytis iš ankstesnių rezultatų ir gauti informacijos atnaujinimą be žmogaus įsikišimo. Tačiau norint atlikti užduotį, kaip minėta, jam tiekiamas didžiulis struktūrizuotų duomenų kiekis (Scott 2021). Tad DI ir MM yra labai glaudžiai susiję, MM apskritai yra laikomas DI pogrupiu (ang. subset). „Protingas“ kompiuteris naudoja DI, kad galvotų kaip žmogus ir pats atliktų užduotis. MM yra tai, kaip kompiuterinė sistema

lavina savo intelektą (Azure 2022). Tad šiuo atveju kalbame apie tam tikrą DI „patirties“ formavimąsi.

Padarius skirtį tarp sąvokų, galima kalbėti apie tai, kokias savybes tenkina DI termino taikymas. Į DI referuojama projektuojant sistemas, aprūpintas žmonėms būdingais intelektiniais procesais, tokiais kaip gebėjimas samprotauti, atrasti prasmę, apibendrinti arba mokytis iš praeities patirties“ (Copeland 2021). Norint geriau suprasti, kas yra DI, reikia atidžiau pažiūrėti, kaip yra suprantamas mokymosi procesas.

a) Gebėjimas mokytis

Bandydamas eksplikuoti mokymosi sąvoką Copelandas pateikia tokį pavyzdį. Vapsvos (kaip kad *Sphex ichneumoneus*) turi instinktą prieš įnešdamos maistą į savo urvą patikrinti, ar jis saugus, ar jo neužėmę įsibrovėliai. Tai yra logiškas veiksmas. Jei į šį procesą įsikiša žmogus, t.y. jei esant vapsvai viduje padėtume maistą per kelis centimetrus nuo įėjimo į jos urvą, vabzdys, užuodęs ir išlindęs laukan, vis tiek atliktų instinktyvų patikros veiksmą, nepaisant to, kad pats ką tik ten buvo ir patikrino, jog vieta saugi ir grėsmės nėra. Šį veiksmą vapsva kartotų tiek kartų, kiek būtų iš naujo dedamas maistas, keičiant jo padėtį. Taigi būtent intelektas leistų vapsvai racionalizuoti savo elgesį ir atsisakyti nereikalingo veiksmo, atitinkančio aplinkybes.

Kalbant apie mokymąsi, Jason Brownlee išskiria tris pagrindinius mokymosi iš problemų tipus: prižiūrimą, neprižiūrimą ir pastiprintą. Mokantis prižiūrint, dažniausiai naudojamas bandymų ir klaidų metodas, kai išbandome kompiuterius, kad gautume žinomą atsakymą, pirmiausia įvedame reikšmes ir palyginame jas su teisingu atsakymu. Taip kompiuteriai įgyja patirties, kad išvengtų ankstesnių klaidų ir pakoreguotų atsakymą kitoms reikšmėms, kol pateiks tikslų atsakymą. Mokymuisi be priežiūros taikomas tas pats metodas, išskyrus tai, kad kompiuteriams nepateikiamas teisingas atsakymas – jie turi rasti struktūrą, išsiaiškinti paslėptus modelius, kad gautų teisingą atsakymą. Sustiprintame mokymėsi kompiuteris turėtų išsiaiškinti, kaip veikti sąveikaujant su aplinka, kad būtų pasiektas tam tikras tikslas, pavyzdžiui, vairuoti automobilį ar žaisti šachmatais. Be šių trijų pagrindinių mokymosi metodų, yra pusiau prižiūrimas mokymasis, kuris sujungia prižiūrimą ir neprižiūrimą mokymąsi taip, kad kompiuteriui būtų duodamas nepilnas mokymo signalas. Perėję šias mokymosi fazes, kompiuteris įgyja galimybę „apibendrinti“, kur gali sugalvoti sprendimus, su kuriais anksčiau nebuvo susidūrę. Programa, taikydama ankstesnę patirtį, mokosi ir taiko žinias analogiškomis naujoms situacijoms. Tai gali būti taikoma ir kalbos mokymėsi, kada prirašius „add ed“ taisyklę, programa galėtų formuoti būtojo laiko anglų kalbos veiksmažodžius, remdamasi ankstesne patirtimi (Brownlee 2019).

b) Gebėjimas samprotauti

Tuo tarpu intelekto savybė samprotauti suprantama kaip proto gebėjimas daryti situaciją atitinkančias išvadas. Williamas J. Littlefieldas II išskiria tris samprotavimų tipus. Du iš jų visi išmokome mokykloje: dedukcinį ir indukcinį samprotavimą. Kompiuteriai gali puikiai atlikti abu šiuos veiksmus. Pirmuoju atveju prielaidų teisingumas garantuoja išvadų teisingumą. Indukciniu samprotavimu, remiantis surinktais duomenimis, kuriami modeliai, apibūdinantys situaciją ir leidžiantys nuspėti ateitį, tačiau gavus vis naujų duomenų, modelis peržiūrimas ir tobulinamas, kad kuo geriau atitiktų situaciją. Tačiau įvedama ir abdukcija, kai, remiantis turimomis žiniomis, faktais ar teorijomis, sukuriama nauja hipotezė. „Skirtingai nuo indukcijos ar dedukcijos, kai mes pradėdame nuo atvejų, kad padarytume išvadas apie taisyklę, arba atvirkščiai, su abdukcija, mes sukuriame hipotezę, kad paaiškintume atvejo ir taisyklės ryšį, [tai] <...> edukuotas spėjimas“ (Littlefield II, 2019). Kompiuteriui labai sunku gerai atlikti tokią užduotį. Littlefieldas sako, kad jie įstrigtų begalinėje kilpoje: „Iš dalies abdukcija yra sudėtinga dėl to, kad turime daryti išvadą apie kai kurias tikėtinas hipotezes iš tikrai begalinio paaiškinimų rinkinio...“ (Littlefield II, 2019). Kaip galima būtų išspręsti šią problemą iki šiol nėra aišku. Juk žmogus rinkdamasis tarp kelių hipotezių dažnai vadovaujasi intuicija, tuo tarpu kompiuteriui kol kas trūksta šios savybės.

c) Gebėjimas spręsti problemas

Problemos sprendimas apibrėžiamas kaip gebėjimas taikant įvairius veiksmus sistemingai ieškoti sprendimo, kad būtų pasiektas iš anksto užsibrėžtas tikslas. Taikomi specialiosios ir bendrosios paskirties metodai. Pirmasis, kai yra konkreti problema ir išnaudojami specifiniai situaciją atitinkantys ypatumai. Tuo tarpu bendrosios paskirties metodo pavyzdžiu galėtų būti priemonių pabaigos analizė, kitaip tariant, laipsniškas skirtumo tarp esamos būsenos ir galutinio tikslo mažinimas. Programa parenka priemonę iš sąrašo, kad pasiektų tikslą. Tai gali pasitarnauti tiek stalo žaidimuose, tiek matematinių įrodymų kūrime, tiek manipuliuojant virtualiais objektais kompiuterio sukurtame pasaulyje (Kapoor, Yadav, Adlakha 2015). Kitaip sakant DI gali metodiškai ieškoti problemos sprendimo būdo tol, kol maksimaliai neatitiks rezultato ir tikslo sutapimo kriterijaus.

d) Suvokimas

Suvokimas, kaip jį apibrėžia Malcolmas Tatumas, yra kita DI savybė, vartojama norint nustatyti kompiuterinės sistemos gebėjimą interpretuoti duomenis panašiai kaip žmonės naudoja savo pojūčius, kad susisiektų su juos supančiu pasauliu. Suvokimo tikslas yra aprūpinti kompiuterinę

sistemą reikalinga programine įranga, kad būtų galima atpažinti vaizdus, garsus ir net prisilietimą tokiu būdu, kuris pagerintų žmonių operatorių ir mašinų sąveiką (Tatum 2012). Programos jau pajėgios ne tik identifikuoti individus, bet turime ir autonomiškai valdomas ir atvirame kelyje judančias transporto priemones bei robotus, gebančius surinkti paskirus objektus ar sukonstruoti objektus iš atsitiktinių detalių.

e) Kalba

Kalba taip pat priskiriama intelektui. Jai būdinga turėti prasmę pagal susitarimą. Kitaip tariant, ji neapsiriboja tik tariamu žodžiu, – yra ir natūrali kalbos reikšmė. Čia turime omenyje ir skirtingą reikšmę skirtingose šalyse turinčius kelio ženklus ar rankų gestus. Arba, pavyzdžiui, skirtumą tarp teiginio, kuriuo pranešame apie artėjantį lietų ir fizikinio aiškinamojo teiginio per slėgio kritimą. Todėl nors nėra visuotinai priimto sutarimo, remiantis viena iš teorijų galėtume teigti, kad kalba yra istorinė, – suprasti kalbą, tai kartu ir jausti kontekstą, įsitraukti į kalbinius žaidimus kalbinėje bendruomenėje (Copeland 2021). Problema yra ta, kad sakiniai nėra vien žodžių kombinacija, maža to, kiekvienas sakinytis gali būti suinterpretuotas labai skirtingai. O jeigu prisiminsime, kokį vaidmenį mūsų bendravime atlieka intonacijos ir pauzės, galima bus lengviau suprasti, su kokiais sunkumais susiduria mokslininkai bandantys išmokyti DI kalbėti.

f) Apibendrinimas

Copeland teigimu, jeigu būtų įmanoma sukurti kompiuterį, atitinkantį visas šias charakteristikas, kalbėtume apie „stiprųjį DI“ (Copeland 2021). Tačiau stiprus DI, suprantamas tiesiog kaip mąstančios mašinos, daugelio ekspertų nuomone nepasiteisino, nes net ir artimiausioje ateityje neaišku, ar turėsime skruzdės intelektiniams gebėjimams prilygstančią mašiną.

Tad tenka kalbėti apie kitus DI tipus – taikomąjį DI (ang. applied AI) ir pažinimo modeliavimą (ang. cognitive simulation). Su taikomuoju DI susiduriame kalbėdami apie ekspertinę medicinos diagnostiką ar prekybos vertybiniais popieriais sistemą, todėl tai jau duoda apčiuopiamą naudą. Tuo tarpu pažinimo modeliavimai taikomi siekiant suprasti žmogaus proto veikimą, pavyzdžiui, atpažįstant veidus ar atsimenant dalykus, todėl atradimai šioje srityje ženkliai prisideda prie neuromokslų ir kognityvinės psichologijos.

1.2. DI KAIP KŪNAS IR GALIOS ĮRANKIS

Taip galima būtų apibrėžti DI techniškai. Tačiau toks supratimas neįtraukia socialinės bei politinės DI dimensijų. Ar galima, kalbant apie DI, jas ignoruoti? Taip. Tačiau eiti šiuo keliu yra pavojinga – šiuo atveju abstrahuotis nuo politikos reiškia nutylėti apie vieną iš esminių su DI susijusių klausimų. Ką tu norima pasakyti? Jeigu kažkas rašytų knygą apie šachmatus, jis galėtų skirti net visą skyrių tam, kad Šaltojo karo metu vieno ar kito šachmatininko pergalė buvo traktuojama vos ne kaip visos kapitalistinės ar socialistinės stovyklos pergalė. Tačiau jeigu šito skyriaus nebūtų knyga apie šachmatus nieko neprarastų, nes jie niekada nebuvo realiu galios įrankiu. Tuo tarpu DI yra galios įrankis, o ten, kur yra galia, apie politiką kalbėti būtina.

Jau minėta Kate Crawford yra viena iš autorių, einančių šiuo keliu. Ji pastebi, kad DI nėra abstraktus debesis (ang. cloud), o veikiau įkūnytas ir materialus, pagamintas iš gamtos išteklių, kuro, žmonių darbo jėgos, infrastruktūros, logistikos. Tai reiškia, kad jis nėra autonomiškas, nes reikalauja intensyviu apskaičiavimu pagrįsto mokymo, didelių duomenų rinkinių ir iš anksto nustatytų taisyklių. Maža to – jis pilnai priklausomas nuo visuomenėje veikiančių politinių ir socialinių struktūrų. Tai reiškia, kad jis nėra prieinamas bet kam. Būtent reikalingo jam vystyti kapitalo poreikis lemia jo tarnystę dominuojančiam interesui. Šia prasme galime DI matyti kaip galios registrą (Crawford 2021, 8). Tokiu būdu DI yra instrumentalizuojamas galios siekimui. Kate Crawford teigia, kad DI – tai sistemos, sukeliančios gilius ir ilgalaikius geomorfinius pokyčius mūsų planetoje, didinančios darbo jėgos nelygybės formas.

Bandant atskleisti požiūrį į šią problemą svarbu nubrėžti paraleles tarp DI sukūrimo bei kasybos. Lewisas Mumfordas primena, kad kasyba pasitarnaudavo galios stiprinimui dviem būdais. Iš vienos pusės, gaunama žaliava ir pinigais, tiesiogiai reikalingais karui, iš kitos pusės – skatindama ginklų gamybos industrializaciją, o tokiu būdu finansininkai galėjo pasipelnyti iš abiejų procesų (Mbembe 2017). Tad nepaisant to, kad šachtose dirbdavo ir tebedirba tūkstančiai žmonių, pasipelnydavo ir pasipelno iš jų darbo tik vienetai. Taip, paprasti darbininkai gauna algą, tačiau dažnai žala, kurią jie padaro savo sveikatai, yra kur kas didesnė negu finansinė nauda. Maža to, kasybos daro didžiulę žalą ekologijai – ne vien dėl to, kad žemėje gręžiama skylė, bet ir dėl to, kad kasybos skatina kitų pramonės sričių vystymąsi, o tai savo ruožtu dar labiau užteršia gamtą (pavyzdžiui, vien gautų žaliavų transportavimas versdavo statyti geležinkelius ir didinti traukinių, varomų anglimi, skaičių) (Crawford 2021, 33-34).

DI taip pat dirba pagal vad. kasybos (ang. mining) principą – jo stublinančių rezultatų pagrindas yra duomenų kasyba. Tačiau šių duomenų „kasyba“, t.y. jų gavimas bei analizavimas, taip pat kaip ir aukso kasyba, turi aibę pusių apie kurias dažnai užmirštama. Visų pirma, DI tiesiogine to žodžio prasme prasideda nuo šachtų – juk jose gaunami resursai reikalingi serverių, kompiuterių, laidų ir t.t. gamybai. Tačiau tai tik pradžia. Kada žaliavos yra išgautos, iš jų reikia pagaminti visus būtinus DI elementus. Po to prasideda darbas su duomenimis – jų gavimas ir analizavimas taip pat

nevyksta tuščioje erdvėje ir reikalauja nemažai pastangų. Šių „paslėptų“ pusių sąrašą galima būtų tęsti ir toliau (ką, pavyzdžiui ir daro Lewis Mumford). Tačiau jau šioje vietoje išvada yra ganėtinai aiški – DI yra vienas iš megaprojektų, kurie reikalauja kur kas daugiau žaliavų, pinigų ir „žmogiškųjų išteklių“ negu tai atrodo iš pirmo žvilgsnio.

Ar galima tikėtis, kad kada nors visos minėtos išlaidos „atsipirks“? Tai, kad DI suteikia pranašumą politinėje bei ekonominėje kovoje, padaro jį galios įrankiu. Tai reiškia, kad vienetai, kuriems jis yra prieinamas taps dar galingesni, o dauguma, kuriai jis nėra prieinamas, bus dar labiau atskirti nuo galios.

Dėl to technologijos tampa centriniu tyrinėjimo objektu ne tik kai kalbame apie ekonominių augimą, gerovės kūrimą ar skurto mažinimą, bet ir sprendžiant platesnes visuomenės problemas. Technologijų prigimtį tyrinėjantys Philipas Faulkneris, Clive Lawsonas ir Jochen Runde savo straipsnyje „Theorising technology“ (Faulkner, Lawson, Runde 2010) eksplikuoja Tony Smitho idėją, kad „technologija kapitalizme visų pirma yra priemonė pasiekti kapitalo tikslą – valorizacijos, o technologiniai pokyčiai labai dalinai ir nesaugiai skatina žmogaus tikslus“ (Smith 2010). Jų apžvalgoje atsispindi autoriaus mintis, kad kapitalizmas turi struktūrinį reikalavimą diegti naujas technologijas. Remiantis tuo, technologijos yra esminė nuolatinio naujų produktyvių galimybių paieškos proceso dalis, atsižvelgiant į tai, kad produktyvumo didinimas kitais kanalais, pavyzdžiui, darbo proceso organizavimu ir įgūdžių įgijimu, turi ženklių trūkumų. Smithas teigia, kad naujai klijuojamos etiketės technologijoms kaip endogeninėms ar „žinių ekonomikos“ atsiradimas viso labo atkartoja Markso idėjas apie technologijų, kaip priemonės nuolatinei produktyvumo pažangai, svarbą. Taip pat, kad įvairioms technologijoms (pvz., lygiaverčių tinklų kūrimui (ang. Peer-to-peer-networking)) nebus leista pasiekti savo potencialo žmogaus klestėjimui, nes tai prieštarautų kapitalo klestėjimui. Jis taip pat teigia, kad įvairių pastarojo meto ekonominių problemų, ypač finansinių krizių, šaknys yra techninių pokyčių pobūdis, perteklinis kaupimas, ir kad nėra akivaizdžių priežasčių manyti, kad tokių krizių galima išvengti koku nors tvariu būdu kapitalistinėje gamybos sistemoje (Faulkner, Lawson, Runde 2010). Šis darbas skirtas būtent šiai problematikai, todėl DI kaip galios įrankis bus tolesnio tyrimo dėmesio centre.

2. DI KAIP ETINĖ PROBLEMA GALIOS CENTRŲ KOVOS KONTEKSTE

Kalbant apie nepakankamą etinį reguliavimą DI sukūrimo proceso etape galima apsiriboti faktų konstatavimu bei aprašyti esamą situaciją DI reguliavimo srityje. Tačiau naivu būtų manyti, kad tokia deskriptyvi prieiga nereikalauja jokio teorinio pagrindimo. Juk, galu gale, reikia iš anksto žinoti, kas yra etika, kada jai nusižengiama, ir kodėl apskritai apie tai verta kalbėti. Todėl nepaisant to, kad šiame darbe pagrindinis dėmesys skiriamas faktų surinkimui, būtent teorinis darbo pagrindas vaidina svarbų vaidmenį suteikiant darbui vientisumo. Šiame darbe į DI bandoma žiūrėti pro vieną iš pagrindinių politikos konceptų prizmę, t.y, pro galios (ang. power) sąvoką. Galia šiame darbe apibrėžiama kaip noras primesti savo valią kitiems tam, kad galia taptų dar galingesnė. Etika suprantama priešingai – apeliuojant į Kantą etiniu elgesiu laikomas toks veiksmas, kurio pagrindinė sąlyga – subjekto autonomiškumas. O pats etinis elgesys yra tas, kuris atitinka kategorinį imperatyvą (Putinaitė 1998, 132). Taip suprantant etiką pasidaro aišku, kad galios, siekiančios tik savo ekspansijos kitų sąskaita, veiksmai negali būti traktuojami kaip etiški. Atsispiriant nuo šių teorinių prielaidų bandoma parodyti, kad DI kaip universalios galios įrankio supratimas tiesiogiai koreliuoja su nepakankamu etiniu reguliavimu DI kūrimo proceso metu.

Norint pagrįsti aukščiau įvardintą koncepciją reikia pradėti nuo galios sąvokos apibrėžimo. Talcottas Parsons savo straipsnyje „On the Concept of Political Power” (Parsons 1963, 232-262) pastebi, kad nepaisant to, kad galios apibrėžimai priklauso nuo konteksto ir aibės kitų faktorių, visi jie centruojasi aplink idėją, kad galia yra sugebėjimas efektyviai realizuoti savo tikslus (ang. to get things done), ypač tuo metu, kai tikslų realizavimas susiduria su tam tikra opozicija. Tas, kuriam pavyksta įveikti opoziciją ir priversti kitus vykdyti savo valią, tampa galios centru. Nuo šiol šis galios centras turi tik vieną tikslą – stiprinti savo paties galią. *Būtent tai vadiname politika. O kelių didžiųjų pasaulinių galios centrų kovą – tarptautiniais santykiais.* Klausimas tik tas, kaip priversti kitus vykdyti savo valią, t.y. tapti galios centru.

Josephas Nye pasiūlė skirstyti galią į kietą (ang. hard) ir švelnią (ang. soft). „Kietą“ galią tyrėjas sieja su ginkluotu smurtu, karine intervencija, ekonominiu spaudimu. Kitaip sakant, „kietos“ galios pagrindas yra materialiniai resursai, kurių pagalba priverčiame kitą vykdyti mūsų valią. Tuo tarpu „švelnioji“ galia yra gebėjimas pasiekti tai, ko norime, remiantis savanorišku žmonių dalyvavimu, o ne per prievartą. „Švelnioji“ galia naudoja patrauklius politinius įvaizdžius bei kultūrinės vertybes, būtent jų pagalba galios centrai pasiekia savo tikslą (Nye 2004, 255-256). Šiuo atveju subjektas pats pareiškia norą judėti pageidautina galios centrui linkme. Tokiu būdu, patrauklių idėjų, įvaizdžių ir konceptų pagalba dėka galios centrai realizuoja savo tikslus.

Dažniausiai „švelnioji“ galia siejama su diskurso samprata. Būtent diskurso pagalba galia realizuoja savo tikslus nenaudojant prievartos. Diskursas formuoja tam tikras mąstymo matricas,

prioritetus ir vertybes, padaro kultūrinę erdvę simboliškai reikšminga. Vienas iš pirmųjų tyrėjų, susiejusių diskurso ir valdžios sąvokas, buvo Michel'is Foucault'o. „Diskursai yra taktiniai elementai arba blokai galios santykių lauke; viduje vienos ir tos pačios strategijos gali būti skirtingi ir net prieštaraujantys vienas kitam diskursai“ (Foucault 1978, 101-102). Diskursas gali transliuoti valdžią ir ją kurti. Jis vienu metu yra ir galios įrankis, ir jos efektas (Foucault 1978, 101). Judith Butler savo knygoje *The Psychic Life of Power: Theories in Subjection* vystydama Foucault'o idėjas pastebi, kad „Galia prasiskverbia į mus ir mus susilpnina, mes interiorizuojame arba priimame jos terminus <...> Subjekcija kaip tik ir reiškia šią fundamentalią priklausomybę nuo diskurso, kurio mes niekada nepasirenkame, bet kuris (paradoksaliu būdu) inicijuoja mūsų veiklą ir ją palaiko“ (Butler 1997, 2). Galia, tokiu būdu, realizuojasi diskursuose, kurie gimdo paklusnumo efektą.

Šiuo metu, vieni iš galingiausių Vakarų kultūros diskursų gali būti laikomi demokratinis, etinis, ir... vartotojiška kultūra. Tai nėra atsitiktinė terminų seka ir ryšį tarp jų galima būtų bandyti plačiau paaiškinti. Tačiau šiuo atveju svarbu atkreipti dėmesį kiek į kitą jų santykio aspektą, o būtent į tai, kad tarp šių dviejų diskursų vyksta nebyli kova. Jeigu demokratinės vertybės ir etinės normos nėra taip lengvai interiorizuojamos ir reikalauja atsakingų sprendimų, vartotojiška kultūra, anot Zygmunto Baumano, pasižymi savo ypatingu „lengvumu“. Šis pasaulis „panašus į stalą, kuriame stovi pernelyg daug skanių patiekalų – net patys rajausi negali paragauti visko“ (Bauman 2000, 63). Pagrindinės jėgos, organizuojančios šiuolaikinį gyvenimą, yra pagundos – jos verčia vartotojus vartoti vis daugiau ir daugiau. Būtent vartodamas individas patiria malonumą ir tuo pačiu metu įgauna identitetą (Bauman 2000, 83-84). Baumanas pastebi, kad tokioje visuomenėje įvyko galios instrumentų pokytis. Sename pasaulyje, kurį Baumanas vadina Fordo pasauliu, galia realizuodavo savo tikslus disciplinos ir griežtos kontrolės pagalba. Buvo manoma, kad kontrolės idealas yra Panoptikumai ar Vyresnysis Brolis. Tuo tarpu vartotojiškoje visuomenėje žiniasklaida nurodo individui judėjimo kryptį. Reginiai (ang. spectacles) pakeitė visus bandantį sukontroliuoti viršininką, tačiau neprarado jo galios disciplinuoti. „Paklusimas standartams <...> nuo šiol pasiekiamas pagundos ir masalo pagalba“ (Bauman 2000, 86). Tokiu būdu, žmogus, nepatirdamas jokio išorinio spaudimo, yra lengvai nukreipiamas reikiama linkme ir kontroliuojamas.

Guido Parietti siekdamas konceptualizuoti galią, taigi ir politiką, pateikia analizę, rodančią, kaip dauguma diskusijų apie valdžią – politologijoje ir filosofijoje, nuo Weberio iki Lukeso, įskaitant Dahlą ir Searle'ą, nėra konceptualios, o empirinės ir operatyvinės. Jo teigimu, iki šiol siūlomi galios apibrėžimai yra apskriti (ang. circular) ir atskirti nuo sąvokos prasmės natūralioje kalboje, todėl nustoja būti susiję su galia. Teigiama, kad „diskusijos sukasi apie faktines implikacijas ir valdžios prielaidas, tuo tarpu numanoma samprata labai nesikeičia. Kitaip tariant, dažniausiai vartojamus apibrėžimus galima redukuoti į vieną formą, kuri yra tautologinė: „turi galią, jei gali daryti tą ir aną“. Tai kyla iš prielaidos, kad galia yra kaip reiškinys ar objektas, kurį reikia stebėti empiriškai“ (Parietti

2018). Jo teigimu, reikia nagrinėti kategorinę galios sampratos formą, kai galia reiškia ne daiktą, o sąlygą, kuriai esant tam tikri dalykai gali būti daromi ir galvojami – atitinkantys galimybę, o ne būtinybę.

Taigi iš to, kas buvo pasakyta, galima padaryti tokias išvadas. Galia – tai sugebėjimas realizuoti savo tikslus kitų pagalba. Tai galima daryti naudojant „kietą“ arba „švelnią“ galią. Tačiau toks apibrėžimas ignoruoja vieną svarbų klausimą, būtent klausimą apie galios tikslus ar tikslą. Jeigu, pavyzdžiui, galios centras nori sutelkti žmones tam, kad būtų padarytas naujas kelias, kuris pasitarnaus visiems kolektyvo nariams, ar tai yra galios centro tikslas? Veikiau tai yra viso kolektyvo tikslas, kurį tam tikras individas, ar grupė individų reflektavo ir paskatino realizuoti. Galios centro tikslas, priešingai, gali būti tik vienas – jo paties galios stiprinimas. Šiuo atveju, kalbant apie galią referuojama į autorių, kuris padarė nemaža įtaką jau minėtiems Foucault'o ir Baumanui, t.y. kalbama apie Friedrich'ą Nietzsche. Kaip pastebi Martinas Heideggeris, Nietzsche kalbėjo apie galią (vok. Macht), kaip apie vieną iš save orientuotą dalyką. Vienintelis galios tikslas – ji pati, t.y. galios ekspansija (Хайдеггер 2018, 77-79). Tad galutinis galios apibrėžimas turėtų skambėti taip: galia – tai sugebėjimas didinti savo galią kitų pagalba.

Turint rankose tokį apibrėžimą Nietzsche ir etiką traktavo kaip dar vieną galios įrankį, padedantį jai stiprinti save. Tačiau tai nėra vienintelis etikos supratimas. Galima būtų sakyti, kad Kantas apibrėžė etiką priešingai. Nietzsche'io etika buvo būdas įtikinti kitus daryti tai, kas paranku man. Kantui, priešingai, etika neturi nieko bendro su noru įtikinti kitą. Ji nėra teorija ar diskursas, paaiškinantis, kas yra gerai ir kas blogai. Etika Kantui yra autentiškas egzistavimo būdas. Jo kategorinis imperatyvas kaip žinia skamba taip: „Elkis taip, kad tavo valios maksima visada galėtų kartu būti visuotinio įstatymų leidimo principas“ (Kantas 1980, 45). Šis imperatyvas nėra susijęs nei su visuotiniu gėriu, nei su visuotine nauda. Jis nėra taisyklė, apibrėžianti, kaip turi atrodyti etinis veiksmas, jis yra etinio veiksmo išraiška suformuluota *post-factum*. Šiuo atveju svarbu tai, kad etika yra suprantama kaip asmeniška etika. Ji yra giliai susijusi su individo *autonomija*. Kanto etika iš esmės yra individo laisvės ir autonomijos apologijos viršūnė. Tuo tarpu Nietzsche supranta etiką priešingai, t.y. kaip bandymą nuslopinti autonomiją. Laisvas yra tik tas, kuris stovi anapus gėrio ir blogio. Laisvas yra tas, kuris siekia stiprinti savo galią. Skirtumas tarp Kanto ir Nietzsche's šiuo atveju yra radikalus. Kantui etika yra autonomijos išraiška, maža to, individo laisvė yra savipakankama – jos realizavimas neimplikuoja kito laisvės atėmimo. Nietzsche's laisvė, priešingai, implikuoja pono ir vergo dialektiką: mano laisvė stiprinama *kito* laisvės atėmimo sąskaita.

Norint padaryti bendrą išvadą, paremtą šių dviejų mąstytojų įžvalgomis, galima būtų pasakyti, kad jeigu tam tikra teorija paaiškina, kaip reikia elgtis ir kodėl tai reikia daryti – ši teorija nėra etika, o tik dar vienas diskursas. Tuo tarpu jeigu kalbame apie atsakingą veiksmą, paremtą savipakankamos

individo autonomijos principu, – kalbame apie etiką. Trumpai tariant, apie etiką kalbame tada, kada veiksmas neslopina kitų laisvės.

Po bandymo konceptualizuoti galią ir etiką galima pereiti prie technologijų ir galios teoretizavimo. Kaip pažymi Justinas Lingevičius ir Dovilė Jakniūnaitė studijoje „Skaitmeninė geopolitinė konkurencija dirbtinio intelekto amžiuje: Jungtinių Amerikos Valstijų, Kinijos ir Europos Sąjungos vizijos“, nors bandymas konceptualizuoti DI pasireiškia net viena kitai prieštaraujančiomis priegomis, bendrai žiūrint galima sutarti, kad DI matomas kaip bendroji technologija (angl. general purpose technology), kurios vystymas turės įtakos ekonomikai, inovavijų kūrimui. Pasak autorių, yra prasminga galvoti apie DI kaip apie strateginį išteklių, kurio efektyvumui reikia valstybės institucijų ir rinkos veikėjų bendro darbo. Dėl to teigiama, kad DI keis didžiųjų valstybių galių pusiausvyrą (Lingevičius, Jakniūnaitė 2021, 77-78). Kitaip tariant, DI tampa vienu iš pagrindinių įrankių galios centrų kovos kontekste.

Padarius šias išvadas galima pabandyti atsakyti į klausimą: kokią vietą DI užima galios ir etikos opozicijos polyje? Kaip jau buvo rašoma, galia realizuoja save dvejais būdais, todėl kalbame apie „kietą“ ir „švelnią“ galias. Skaitant Baumaną gali pasirodyti, kad „kietosios“ galios epocha jau praėjo ir Vakarų pasaulis gali būti apibrėžtas kaip „švelniosios“ galios karalystė. Tačiau jau tuo metu kai Baumanas rašė savo knygą buvo daug kalbama apie tai, kad Didysis Brolis vis dėlto stebi mus. Dabar šis teiginys skamba juo aktualiau. Besivystant DI technologijoms stebėjimo ir kontrolės galia sugrįžta į mūsų pasaulį neregėtu mastu. Tad žymusis panoptikumus niekur nedingo, tačiau dabar jo renkama informacija gali būti naudojama dvejopai – iš vienos pusės, išorinei piliečių kontrolei, iš kitos – vidinei. Aišku, „kietoji“ galia Vakarų pasaulyje pagrinde kelia grėsmę tik nusikaltėliams, tačiau radikalus „kietos“ valdžios pavojus eiliniams piliečiams taip pat niekur nedingo, į ką, pavyzdžiui, nurodo kai kurie ypatingai aršūs *lockdown* 'ų kritikai. Bet kaip ten bebūtų, tarptautiniuose santykiuose „kietoji“ galia matoma kur kas ryškiau – karinės intervencijos, ekonominės sankcijos, slaptos spectarnybų operacijos yra dažnai pasitaikantys dalykai.

Šioje „kietosios“ ir „švelniosios“ galių koegzistavimo situacijoje DI vaidina universalios galios įrankio vaidmenį. Iš vienos pusės, jis naudojamas informacijos surinkimui ir reklamos personalizavimo tikslams. Turint omenyje, kad jau dabar DI rašo straipsnius, gali būti, kad jis kurs atskirą reklamą paskiram individui. Tokiu būdu DI gali tapti galingu diskursų kūrėju. Iš kitos pusės, DI galimybės karinės technikos, prognozavimo ir kibernetinio saugumo srityse taip pat suteikia ji kontroliuojantiems galios centrams didžiulį pranašumą.

Atsispiriant nuo šių teorinių prielaidų teigiama, kad nepakankamas etinis reguliavimas DI sukūrimo proceso metu, susijęs būtent su DI kaip universalios galios įrankio supratimu. Etnis požiūris į DI visuose jo vystymo etapuose reikštų norą naudoti jį paskirų individų autonomijos gynimo tikslams. Tuo tarpu, šiuo metu išgalėjęs DI supratimas liudija apie priešingas tendencijas.

DI suprantamas kaip kontrolės arba pelno didinimo įrankis, kitaip sakant, kaip galios stiprinimo įrankis. Ar galima šiuo atveju kalbėti apie etinį reguliavimą jau DI sukūrimo etape? Apie tai galima būtų kalbėti tik tuo atveju, jeigu šis procesas būtų pasiekiamas visuomenei, o ne vien keliems pasauliniams žaidėjams. Tačiau kol kas visuomenės įsitraukimas į DI sukūrimo procesą ir jo reguliavimą yra stebėtinai mažas. Tačiau ši situacija pilnai atitinka galios centrų interesus. Geresnis etinis reguliavimas DI sukūrimo proceso metu sulėtintų šį procesą ir leistų nesilaikantiems taisyklių aplenkti savo varžovus. Tačiau šiuo atveju būti atsilikusių reiškia būti pralaimėtoju. Kaip pastebi Vincentas C. Mülleris, DI veikia rinkoje pagal principą „nugalėtojas gauna viską“ (Müller 2020). Tokiu būdu DI tampa tikslu, kurio reikia pasiekti „bet kokia kaina“. Priešingu atveju, kiekvienas iš pagrindinių pasaulio galios centrų rizikuoja prarasti savo kaip galios centro statusą. Šios pozicijos atskleidimui ir yra skirtas šis darbas.

3. ETINIS REGULIAVIMAS DI PANAUDOJIMO STADIJOJE

Apibendrinat tai, kas pasakyta aukščiau, galima padaryti tokias išvadas: DI – tai kompiuterio sugebėjimas spręsti aibę intelektualių problemų taip, kaip tai darytų žmogus. Tačiau tai nėra vakuume egzistuojantis protas. DI yra vienas iš pagrindinių transnacionalinių korporacijų ir didelių politinių žaidėjų konkurencijos objektas, jis reikalauja didžiulių investicijų ir daro įtaką ekologiškai situacijai. Ar nėra taip, kad kova dėl DI peržengs visas moralines normas ir vyks tokiu pat laukiniu būdu, kaip kadaise vyko branduolinės varžybos tarp JAV ir Sovietų Sąjungos? Iš pirmo žvilgsnio atrodo, kad moralinis DI reguliavimas yra ir JAV, ir Kinijos, ir ES dėmesio centre. ES apskirtai gali būti laikoma susirūpinimo etine problematika etalonu. Tačiau ar galima teigti, kad ES programoje pakankamai akcentuojama etinė DI sukūrimo proceso pusė? Žemiau bus bandoma parodyti, kad ES rūpesčio centre etiniai DI naudojimo, bet ne sukūrimo principai.

Kaip pastebi John Tasioulas, Oksfordo universiteto AI etikos instituto direktorius: „Kalbant apie etikos klausimų svarbą DI, kyla klausimas, ar tai teisinga etika ir ar teisingi žmonės priima sprendimus? Daug žmonių priimančių sprendimus yra galingi ir niekam neatskaitingi ekonomikos veikėjai, kaip kad *Think tank*‘ai. Kitaip sakant, problema yra ta, kad sprendimai dėl DI jau yra priimami tam tikrų galios centrų. Ir šiuo atveju nėra „neutralių“ sprendimų, visi jie priklauso etikos sričiai. Jei ketinate naudoti AI veido atpažinimui, tai jau yra etinis pasirinkimas. Tai implikuoja etines vertybes. Jei nuspręsite naudoti autonominius ginklus, tai vėlgi yra etinis pasirinkimas. Taigi tai visada bus etika. Lieka tik klausimas, ar tai yra gera etika ar bloga etika, ir kas daro šiuos sprendimus“ (žr. priedą nr. 1). Būtent šis momentas privertė ES ir NATO žymiai atidžiau pažiūrėti į etinį DI reguliavimo klausimą.

3.1. EUROPOS SĄJUNGOS POŽIŪRIS Į DI: DI STRATEGIJA

ES yra parašiusi nemažai dokumentų, kurie liečia vieną ar kitą DI gyvenimo ciklo dalį, įskaitant „Dirbtinį intelektą Europai“ (DI strategiją) (European Commission 2018a), „Baltąją knygą DI“ (European Commission 2020), „2021 Koordinuotą planą DI“ (European Commission 2021a), „Pasiūlymu dėl Reglamento dėl DI akto“ (toliau – Pasiūlymas) (European Commission 2021b) ir kt. Dėmesys DI nestebina turint omenyje, kad dabartinė ES Komisijos vadovė Ursula von der Leyen savo kandidatūros kalboje tarp kitų dalykų akcentavo, kad tyrimai DI srityje yra tarp prioritetinių kadencijos tikslų (Balazs 2019). Tiksliai formuluotė skambėjo taip: „Per pirmąsias 100 savo darbo dienų pateiksiu teisės aktus dėl suderinto Europos požiūrio į dirbtinio intelekto pasekmes žmogiškiesiems ir etiniams aspektams. Taip pat reikėtų atsižvelgti į tai, kaip galime panaudoti didelius duomenis naujovėms, kurios kuria gerovę mūsų visuomenei ir verslui“ (Leyen 2019).

Kalboje taip pat buvo pabrėžta, kad Europa turi padidinti savo konkurencingumą lenktyne su JAV ir Kinija. Kadangi dabartinis DI vedamas perversmas prilyginamas industrializacijos epochai (Thomas 2020), kuomet žaibišku greičiu auga produktyvumas, kuriasi nauji, inovatyvūs verslai, o atsilaisvinę žmogiškieji išteklių bematant vėl tampa įdarbinami kitose pareigose, niekam nekyla abejonių, kad DI yra svarus ekonominio vystymosi faktorius.

Jau 2018 m. 25 ES valstybės pasirašė „Deklaraciją dėl bendradarbiavimo dirbtinio intelekto srityje“ (European Commission 2018b), po ko sekė minėta DI strategija. Tuo parodyta, kad ES prioritetas – patikimo (ang. *trustworthy*) DI sukūrimas, t.y. tokio DI, kurio vartotojai gali juo naudotis be jokios rizikos. Siekiant šio tikslo tais pačiais metais buvo paskirti 52 tyrėjai į vad. Aukšto lygio DI ekspertų darbo grupę. Šie ekspertai gavo užduotį sukurti gaires dėl etinių, teisinių ir socialinių aspektų, liečiančių DI. 2019 m. birželio 26 d. ši grupė pirmojoje europinėje aljanso asamblėjoje pateikė du dokumentus. Pirmas iš jų – „Patikimo dirbtinio intelekto etikos gairės“ (European Commission 2019a, 15). Šiame dokumente ekspertai pateikia savo poziciją ir principus dėl DI etinio patikimumo. Pagrindinės normos, kurias turėtų atitikti patikimas DI, yra šios: legalumas – visų įstatymų paisymas, duomenų surinkimo etape bei DI naudojimo atveju; etiškumas – etinių vertybių paisymas; patikimumas tiek techniniu, tiek ir socialiniu atžvilgiu. Dokumente taip pat pateikti septyni reikalavimai, kuriais remiantis būtų galima DI vadinti patikimu. Tai: 1 – žmogiškasis veiksnys ir žmogaus atliekama priežiūra, įskaitant pagrindines teises; 2 – techninis patvarumas ir saugumas, įskaitant atsparumą išpuoliams ir saugumą, atsarginį planą ir bendrą saugumą, tikslumą, patikimumą ir atkuriamumą; 3 – privatumas ir duomenų apsauga, įskaitant pagarbą privatumui, duomenų kokybę ir vientisumą, ir prieigą prie duomenų; 4 – skaidrumas, įskaitant atsekamumą, paaiškinamumą ir informacijos sklaidą; 5 – įvairovė, nediskriminavimas ir teisingumas, įskaitant nepagrįsto šališkumo vengimą, prieinamumą, universalumą ir suinteresuotųjų subjektų dalyvavimą; 6 – visuomenės ir aplinkos gerovė, įskaitant tvarumą, ekologiškumą, socialinį poveikį, visuomenę ir demokratiją; 7 – atskaitomybė, įskaitant galimybę atlikti auditą, neigiamo poveikio sumažinimą ir informavimą apie jį, kompromisus ir teisių gynimą. Tais pačiais metais pasirodė kitas dokumentas – „Patikimo dirbtinio intelekto politikos ir investicijų rekomendacijos“ (European Commission 2019b), kuriame pateiktos 33 rekomendacijos, skatinančios vystyti patikimą DI. Dar vėliau, 2020 m. Europos Komisija išleido Baltąją knygą – „Dirbtinis intelektas. Europos požiūris į kompetenciją ir pasitikėjimą“ (European Commission 2020). Šiame dokumente pateikiamos konkrečios priemonės, pasitarnaujančios DI vystymui, jo įdiegimui ir panaudojimui. Taip pat kalbama apie fundamentalių žmogaus teisių gynimo aspektus. Baltoji knyga laikytina pagrindiniu strateginiu dokumentu, kuriuo rėmėsi 2021 m. balandžio 21 d. Europos Komisijos pristatytas teisinis Pasiūlymu dėl Reglamento dėl DI akto (European Commission 2021b) – pirmasis pasaulyje dokumentas dėl teisinio DI reguliavimo. Kaip savo studijoje teigia Lingevičius ir Jakniūnaitė: „Pasiūlymas plačiau nediskutuoja etikos gairių,

panašu, kad jas siekiama užkoduoti būtent per galimų pažeidimų valdymą – Kinija matoma kaip kuriantis ir eksportuojantis valstybinės informacijos kontrolės ir sekimo modelis, todėl ES svarbu rasti demokratinėmis vertybėmis grįstą DI reguliavimą (Broadbent, Arrieta-Kenna 2021). Tad formuluojamos rizikos, viena vertus, atskleidžia su DI ateitimi susijusią nežinomybę, bet kartu gali būti matomos kaip atsakas ir saugikliai Kinijos taikomoms praktikoms. EK neabejotinai sąmoningai DI reguliavimą formuluoja kaip europietišku vertybių gynimą nuo DI vystytojų Kinijoje (MacCarthy, Propp 2021). Siekiant sukurti realią ir pagrįstą atsvarą Kinijos praktikoms, ES ne tik svarbu teigti pasirinktus principus, bet ir juos perkelti į praktinį suformuluotų rizikų valdymą“ (Lingevičius, Jakniūnaitė 2021, 86). Taigi nors Pasiūlymas rodo interesą pasiekti tarpusavio sutarimą viduje, DI matomas ir kaip pagrindinis galios įrankis globaliai eksportuoti savo principus. Maža to, Europos Komisija ženkliai padidino investicijas į DI vystymą, tikintis, kad šios investicijos bus 20 mlrd. eurų per metus (kalbame apie valstybines ir privačias investicijas) (European Commission 2021c).

Matome, kad 2018–2020 m. žymi intensyvi ES darba, kuriant strateginę ES DI viziją. Savotišką kulminaciją pasiekta su minėtu Pasiūlymu ir tarptautinio bendradarbiavimo formato su JAV (JAV-ES Prekybos ir technologijų taryba) įkūrimu. ES DI dokumentai akcentuoja etiškumo, orientacijos į žmogų ir patikimumo momentus, o pati Europos Komisija siekia ES paversti „pasaulio lydere saugaus DI srityje“, tikintis sukurti valstybei naudingas DI sistemas (European Commission 2021d). Todėl galima drąsiai teigti, kad etinio DI sukūrimas yra vienas iš ES prioritetų. Tačiau, šioje vietoje pravartu prisiminti psichoanalitiku patarimą, klausytis ne vien to, apie ką kalbama, bet ir to, apie ką nutylima. Kam gi skirta nepakankamai dėmesio minėtuose dokumentuose? Problemoms susijusioms su DI sukūrimo procesu. Apie tai, su kuo galima sieti šią tylą bus kalbama toliau. Tačiau pirmiau būtina pažiūrėti į tai, kaip etikos ir DI santykio klausimas sprendžiamas NATO strategijoje.

3.2.NATO POŽIŪRIS Į DI: DI STRATEGIJA

Kaip pastebi John Tasioulas, “net Vakarų pasaulyje esama nesutarimo dėl etinio DI reguliavimo. “ES požiūris į reguliavimą kitoks nei Amerikoje <...> JAV yra labai atsparios visapusiškam reguliavimui, kaip kad yra ES. Taigi vienas įdomiausių klausimų yra, ar turime siekti visa apimančio reguliavimo, kaip siūlo ES, ar, veikiau, turime siekti tikslingo reguliavimo konkrečiais klausimais, pavyzdžiui, veido atpažinimo ar socialinio balo sistemų nustatyme” (žr. priedą nr. 1) Profesoriaus pastebėjimu, tarpininkų tarp šių dviejų galios centrų galėtų tapti Jungtinė Karalystė. Tačiau nereikia pamiršti, kad JAV ir nemažą ES šalių dalį jungia būtent NATO valstybių blokas. Todėl aptarus ES teisinę bazę dėl DI ir etikos svarbu apžvelgti ir NATO DI strategijoje (NATO 2021) išskiriamus strateginius tikslus. NATO kaip tokius įvardija keturis: 1 – sudaryti pagrindą NATO ir

sąjungininkams rodyti pavyzdį ir skatinti atsakingą DI kūrimą ir naudojimą sąjungininkų gynybos ir saugumo tikslais; 2 – pagreitinti ir integruoti DI pritaikymą plėtojant ir teikiant pajėgumus, didinant sąveiką Aljanse, įskaitant pasiūlymus dėl DI naudojimo atveju, naujų struktūrų ir naujų programų; 3 – apsaugoti ir stebėti DI technologijas ir gebėjimą diegti naujoves, atsižvelgiant į saugumo politikos sumetimus, pavyzdžiui, atsakingo naudojimo principų įgyvendinimą; ir 4 – nustatyti ir apsisaugoti nuo grėsmių, kylančių dėl valstybinių ir nevalstybinių veikėjų piktavališko DI naudojimo. Taigi galima teigti, kad tiek ES, tiek politinio-karinio NATO aljanso išskiriami DI tikslai didele dalimi sutampa.

Kalbant apie atsakingo DI panaudojimo gynyboje principus, įsipareigojama užtikrinti, kad DI taikomosios programos, kurias jie kuria ir ketina įdiegti įvairiais jų gyvavimo ciklo etapais atitiktų šešis principus: 1 – teisėtumas: DI programos bus kuriamos ir naudojamos pagal nacionalinę ir tarptautinę teisę, įskaitant tarptautinę humanitarinę teisę ir žmogaus teises; 2 – atsakomybė ir atskaitomybė: DI taikomosios programos bus kuriamos ir naudojamos laikantis atitinkamo lygio sprendimo (ang. judgement) ir rūpestingumo; siekiant užtikrinti atskaitomybę, taikoma aiški (ang. clear) žmogaus atsakomybė; 3 – paaiškinamumas ir atsekamumas: DI programos bus suprantamos ir skaidrios, įskaitant peržiūros metodus, šaltinių ir procedūrų naudojimą. Tai apima tikrinimo, vertinimo ir patvirtinimo mechanizmus NATO ir (arba) nacionaliniu lygiu; 4 – patikimumas: DI programos turės aiškius, gerai apibrėžtus naudojimo atvejus. Tokių pajėgumų sauga ir patikimumas bus tikrinami ir užtikrinami per visą jų gyvavimo ciklą, įskaitant nustatytas NATO ir (arba) nacionalines sertifikavimo procedūras; 5 – valdymas: DI programos bus kuriamos ir naudojamos pagal numatytas funkcijas ir leis: tinkamai sąveikauti žmogui ir mašinai; leis pastebėti ir išvengti nenumatytų pasekmių; ir suteiks galimybę imtis veiksmų, pavyzdžiui, sistemų atirbojimas (ang. disengagement) arba išjungimas, kai sistemos demonstruoja nenumatytą elgesį; 6 – šališkumo mažinimas: bus imtasi iniciatyvių veiksmų, kad būtų sumažintas bet koks nenumatytas šališkumas, kuriant ir naudojant DI programas ir duomenų rinkinius. Kartu paėmus šie tikslai ir principai atskleidžia Vakaruose besiformuojantį požiūrį į DI.

Įdėmiau nagrinėjant strategiją išaiškėja, kad yra aiškiai suvokiamos su DI sukūrimu, vystymu ir taikymu susijusios grėsmės, taip pat grėsmės, kylančios, jei priešiškos valstybės ar nevalstybinių veikėjų jėgos panaudotų DI technologijas prieš juos pačius. Iš teksto galima spręsti, jog neabejojama, kad DI vaidmuo ateityje augs ir taps vis reikšmingesniu galios svertu. Galima išskaityti siekį kuo sparčiau ir efektyviau išvystyti sąveikumą, suderinamumą, sklandų apsikeitimą informacija tarp partnerių, operatyvų reagavimą į krizes, galų gale, vienodą požiūrį DI srityje. Susidaro įspūdis, kad to siekiama tam, kad būtų galima įgyti daugiau galios konkurencinėje kovoje. Tiesa, tai jog išskiriamas noras užtikrinti, kad vykdoma veikla atitiktų bendrus NATO principus ir vertybes, leistų teigti, kad nors DI matomas kaip galios įrankis, juo kaip tokiu numatoma naudotis tik esant būtinybei

ir atsižvelgus į kitų valstybių ar nevalstybių veikėjų veiksmus. T.y. Aljansas siekia didinti galią, tačiau DI vystomas tiek, kad kuo geriau tarnautų bendram Aljanso saugumui ir gynybai. Nors strategijoje nekalbama apie DI naudojimą prieš autoritarinius režimus, yra suvokiama, kad jei sritis nebus pakankamai greitai vystoma ir nebus konkurencinio pranašumo, grėsmės atveju žala Aljansui gali būti didžiulė ar net neatitaisoma. Tą galima spręsti iš to, kad išskiriama grėsmė kritinei infrastruktūrai, civiliniam pasirengimui, pažeidimų kibernetinėje erdvėje sukūrimui. Pabrėžtina, kad nors strategijoje kalbama apie DI kūrimą, nekalbama apie kūrimo kaštus, tik įvertinamos jau dabar esančios ir potencialiai galinčios iškilti rizikos, susijusios su DI naudojimu. Atkreiptinas dėmesys, kad kalbama apie naujus bendradarbiavimo formatus tarp Aljanso ir partnerių, apimant reguliarius aukšto lygio dialogus, įtraukiant technologijų įmones strateginiu politiniu lygmeniu. Taip pat kalbama apie būtinybę dirbti išvien, kad būtų įgytas pranašumas ir pavyktų išvengti silpnų grandžių aljanse, mažinti pažeidimus sistemose ir visuomenėse, didinti bendrą atsparumą dezinformacijai ir greito (bemaž automatinio) reagavimo į krizes mechanizmus. Įdomu, kad išskiriama dezinformacija, nukreipta prieš DI naudojimą kariniais tikslais, kuri galimai neigiamai veiktų visuomenės nuomonę ir nuteiktų ją prieš industrijos vystymą šioje srityje. Tai tik patvirtina, kad DI matomas kaip galios įrankis, suvokiant, kad jo naudojimui ir vystymui reikalingas visuomenės palaikymas.

Kaip jau minėta, šališkumo mažinimas yra vienas iš numatytyjų DI panaudojimo gynyboje principų. Teigiama, kad palaikydami saugų ir atsakingą DI naudojimą, NATO ir sąjungininkai sąmoningai įgyvendins šališkumo mažinimo pastangas. Kaip vienas iš ryškesnių probleminių taškų išskiriamas šališkumas dėl individualių bruožų, pavyzdžiui, lyties, etninės priklausomybės ar asmeninių savybių. Tai rodo, kad NATO nesvetimas duomenų bazių šališkumo problematikos supratimas.

Taigi nemaža dalimi NATO ir ES požiūris į etinį DI reguliavimą sutampa. Pabrėžiama būtinybė apsaugoti asmens laisvę ir pagrindines jo teises. Tuo pačiu metu išskirtinis dėmesys NATO DI strategijoje skiriamas gynybos klausimui. Pabrėžiama, kad DI gali būti naudojamas kaip galios įrankis, tačiau tik tam, kad būtų užtikrintas fizinis ir kibernetinis aljanso saugumas. Kitaip sakant, ir šiuo atveju bandoma paisyti etinių DI panaudojimo normų. Nepaisant to, NATO DI strategijoje nekalbama apie etines problemas, susijusias su DI sukūrimu – šis klausimas tiesiog nėra tematizuojamas. Prie šios problemos tematizavimo dabar ir būtina pereiti.

4. ETINIO REGULIAVIMO STOKA DI SUKŪRIMO STADIJOJE

Kaip pastebi Michał Boni, buvęs Europos Parlamento narys ir Martens centro vyresnysis mokslinis bendradarbis, DI pakankamai reguliuojamas iš etinės pusės. Jo teigimu, „reguliavimo pakanka. Dešimtbalėje sistemoje vertinčiau 8/9. Svarbu, kad kalbame apie etinius DI funkcionavimo aspektus, kaip prisiminate, dar 2019 m. buvo įsteigta aukšto lygio DI ekspertų grupė. Ji pristatė Vokietijos DI etikos grupės požiūrį. Manau, aišku, kad visuose ES dokumentuose stengiamasi parodyti, kokie svarbūs yra etiniai aspektai. Kita vertus, yra problema, kaip etinį supratimą ir etinius principus perkelti į praktinius sprendimus“ (žr. priedą nr. 2). Tokią skirtį tarp teorijos ir praktikos, įvaizdžio kūrimo ir realaus veikimo galima paaiškinti savų interesų gynimo logika. Kaip pastebi Vincentas C. Mülleris, „pastebima tendencija, kad verslas, kariuomenė ir kai kurios viešojo administravimo institucijos „tiesiog kalba“ ir atlieka „etikos higieną“ (ang. ethics washing), siekdamas išsaugoti gerą įvaizdį visuomenėse, tuo tarpu tęsia įprastą veiklą kaip ir anksčiau.“ (Müller 2020). Tai įgimta galios centrų savybė priešintis jiems primetamiems apribojimams. Tačiau vis dėlto apie etinius DI panaudojimo kriterijus bent kalbama. Tuo tarpu su DI sukūrimo procesu situacija yra dar prastesnė. Šiame skyriuje norėtusi eksplikuoti būtent etinio DI sukūrimo proceso reguliavimo stokos problemą.

Šioje vietoje pravartu atsižvelgti į Kate Crawford daromą paralelę tarp kasybos ir DI. Kaip teigia autorė, kasybos verslas pelningas tik tuo atveju, jei neatsižvelgiama į visas su šia veikla susijusias išlaidas – poveikį aplinkai, prarastas kalnakasių gyvybes ar sveikatą, psichologinį ir socialinį aspektus, priverčiant žmones palikti gyvenamą vietą. Apskritai, dar XVI a. Agricola teigė, jog visiems aišku, kad daroma žala iš kalnakasybos yra didesnė nei iš kalnakasybos gaunamų metalų vertė (Crawford 2021, 26-28). Kai 1800-ais buvo išgaunamas auksas ir sidabras iš Nevados ir Kalifornijos žemių miestui statyti, šiandien nebeslepama, kad to pasekoje žemė buvo sudarkyta, vandens keliai užteršti, aplinkinis miškas sunaikintas, o žmonės priversti palikti savo namus. Kadangi gavybos šachtos buvo toli nuo miesto, buvo paprasta negalvoti, kokią žalą vietai daro iš jos gaunami metalai ir gautu pelnu kuriamas miestas. Kaip įdomiai pristatant Gray Brechino knygos *Imperial San Francisco: Urban Power, Earthly Ruin* mintį išskyrė Los Angeles Times, „Aukštieji šiuolaikinio miesto centro pastatai <...> primena apverstą minų peizažą ir tarnauja beveik tuo pačiu tikslu tiems, kuriems jie priklauso“ (Kirsch 2000). Crawford savo ruožtu knygoje daro išvadą, kad ekstrakcijos tarnavimas nedaugelio interesantų grupei panašiai veikia šiandien, jau technologinės industrijos laikais išgaunant baltą ličio kristalą, dėl ko Silicio slėnis, kaip pasaulinis aukštųjų technologijų ir inovacijų centras, prigludžia penkių didžiausių technologijų kompanijų – Microsoft, Amazon, Facebook, Google, Apple – biurus. Šioje vietoje be pėdsakų gamtai į akis krinta ir didžiulė nelygybė gyvenamosios vietos (ang. housing) ir prabangos prekių koncentracijos atžvilgiu. Autorė pateikia

pavyzdį, kad pačiame mieste tarp tokių turtingų rajonų kaip SoMair ir kelių eismo juostų magistralėse nusidriekusių „palapinių miestelių“ (kaip Division Street) atstumas yra visai nedidelis, o vaizdas radikaliai priešingas (Crawford 2021, 28). Nesunku rasti patvirtinančių duomenų, kad po technologijų bumo San Franciske dabar vienas didžiausių gatvėse gyvenančių benamių rodiklių JAV (Ho 2019). To akivaizdoje Jungtinių Tautų (JT) specialusis pranešėjas padėtį pavadino „nepriimtinu“ žmogaus teisių pažeidimu, kai tūkstančiai benamių gyventojų nė neturi priėjimo prie vandens, sanitarinių ir sveikatos paslaugų, priešingai nei rekordinis netoliese gyvenančių milijardierių skaičius (Gee 2018). Šis pavyzdys gerai iliustruoja, kad didžiausią gavybos naudą pasidalina nedaugelis.

Nepaisant to, jog trūksta studijų apie tai, kokia DI sukūrimo kaina, vis dėl to aišku, kad jo sukūrimas labai priklausomas nuo energijos ir resursų išgavimo – retų gamtos mineralų, naftos, anglies – pigios darbo jėgos ir masinių duomenų surinkimo. Ši ekstraktyvi DI politika byloja problemą, kad tikrosios šios gavybos sąnaudos niekada nepadengia pati pramonė. Konkurencijos genamas poreikis kurti vis greitesnius ir efektyvesnius modelius lemia ir vis daugiau papildomos energijos reikalaujančius modelius, kurie palieka vis didesnę anglies dioksido žymę mūsų planetoje. Dėka menkai apmokamos darbo jėgos, kuri atlieka paprastas užduotis programų kūrime ir tobulinime, turime sistemas, kurios atrodo intelektualesnės nei yra iš tiesų (Crawford 2021, 68).

Remiantis tuo galime teigti, kad DI tik skatina neteisingumo augimą bei blogina ekologinę situaciją. Tuo pačiu metu, kaip matėme, ES ir NATO dokumentuose šie klausimai yra apeinami. Kitaip sakant, problemos susijusios su DI sukūrimu lieka šešėlyje. Būtent todėl ši tema reikalauja kur kas atidesnio visuomenės ir politikų dėmesio. Bandant atskleisti problemos mastą žemiau bus bandoma aptarti pagrindines su DI sukūrimu susijusias etines problemas.

5. PAGRINDINĖS SU DI SUKŪRIMU SUSIJUSIOS ETINĖS PROBLEMOS

5.1. ATSAKOMYBĖS PROBLEMA

Tyrinėtojas pradėjęs tyrinėti etines problemas, susijusias su DI sukūrimu, labai greitai pastebi, kad yra žmonių, neigiančių patį šio klausimo adekvatumą. Klausimas, ar mokslininkas yra atsakingas už savo darbo rezultatus jau tapo klasikiniu. Jam savo dėmesį skyrė ir Kurtas Vonnegutas, ir Albertas Schweitzeris, ir Andrejus Sacharovas ir daugybė kitų. DI atveju jis vėl pasidaro itin aktualus. Galima išskirti dvi su atsakomybe susijusias problemas – „moralines“ ir „materialines“.

a) Moralinės problemos

Yra nemaža pavyzdžių, rodančių, kaip ne retai daromos klaidingos generalizacijos, remiantis ribotais šaltiniais, duomenų bazėmis, kurios turi trūkumų. Vis tik neretai jos naudojamos kaip galutinis šaltinis ir DI šaltinių mokymui. Ar mokslininkai kalti dėl šių klaidingų generalizacijų? Ar tai nereiškia, kad etine problematika turi būti susirūpinta jau pačioje DI sukūrimo pradžioje? Atrodo, kad mokslininkas tik turi „įdėti“ jau esamus duomenis į mašiną, o ji „neutraliai“ atliks visą darbą ir dėl to jos sprendimai taip pat bus objektyvūs. Bet juk jau pradinis materialas nėra etiškai neutralus. O tai reiškia, kad mokslininkai, ignoruojantys šitą pradinio materialo savybę, taip pat prisiima atsakomybę už DI veiklos rezultatus.

Norėdami pagrįsti, kad technologijos nėra objektyvios ir neutralios, o patys DI kūrėjai yra nesąmoningai veikiami savo išankstinių nusistatymų reikia paliesti tokius klausimus, kaip kad: kaip veikia ir kokią įtaką turi klasifikavimo sistemos? Kaip pastebi Geoffrey C. Bowkeris ir Susan Leigh Star, klasifikavimo sistemos labai plačiai įsigalėjusios mūsų kasdienoje, o mes ne tik kad dažnai jų nepastebime, bet ir nežinome, kaip jos mus veikia (Bowker, Star 1999). Šie autoriai pasirodo debatuose, kurie dažnai vyksta prieš klasifikavimo sistemos standartizavimą. Norėdami iliustruoti tą mastą, autoriai pateikia Tarptautinės ligų klasifikacijos (TLK) pavyzdį. Jis vertingas tuo, kad iš pirmo žvilgsnio atrodo, kad ligų klasifikavimas yra kažkas objektyvaus. Tačiau šios klasifikacijos kūrimas užtruko daug metų ir vis dar yra diskusijų objektas. Pavyzdžiui, atogrąžų šalys mano, kad atogrąžų ligų yra labai mažai, palyginti su „turtingo pasaulio“ ligomis, tokiomis kaip vėžys ir širdies ligos. Japonijoje širdies priepuoliai laikomi žemo pasireiškimo, todėl mirties liudijime insultas dažnai nurodomas kaip mirties priežastis. Tai neleidžia susidaryti tikslaus vaizdo, juolab lyginti statistikos tarp šalių. Autoriai taip pat parodo, kaip klasifikacija veikia mūsų pačių

saviidentifikaciją. Jie pateikia ekstremalų atvejį: tuberkuliozės pacientus, kurie visiškai priklausomi nuo savo gydytojų, kad nustatytų savo būklę, ir juodaodžius Pietų Afrikoje, kuriems taikomas apartheidas, dėl ko jie priversti pakeisti gyvenamąją vietą, darbą ir šeimą. Šia prasme, kai susiduriame su situacija, kada nežinome, kuriai iš kategorijų save priskirti, pavyzdžiui, pildant anketą, ir pasirenkame aktualiausia/artimiausia, šis veiksmas galiausiai apibrėžia mūsų tapatybę. Bowker ir Star taip pat pateikė Slaugos intervencijos klasifikacijos (NIC) pavyzdį, kai slaugytojai buvo paprašyti apibūdinti, ką jie daro, kad būtų galima klasifikuoti, standartizuoti ir integruoti rezultatus į atsiskaitymo sistemą. Nors kai kurie džiaugėsi, kad dabar bus viskas matoma „sistemoje“, kiti baisėjosi, kad negalės daryti to, kas jų manymu yra teisinga, nes tai nepaklius į sistemos „rėmus“. Čia aiškiai juntama įtampa tarp sistemos ir vietoje priimamų sprendimų, kuriuos reikia daryti, kad sistema iš tikrųjų veiktų. Kaip teigia autoriai, jokia klasifikavimo sistema nėra tobula. Visada bus elementų, kurie netinkamai patenka į kategoriją. Tačiau svarbu atsiminti, kad tai nėra elemento atspindys – tai sistemos netinkamumo atspindys. Mūsų apibrėžimai paprasčiausiai dažnai būna nepakankamai išsamūs. Galima sakyti, kad klasifikavimo sistemos paprastai būna „juodosios dėžės“ Bruno Latour prasme. Jie linkę tapti nematomi, ir mes pamirštame, kiek pastangų įdėjome juos kuriant, pamirštame politinius bei etinius klausimus, kurie buvo susiję su jų kūrimu (Bowker, Star 1999). Tokiu būdu už klasifikacinių sistemų slypi jų kūrėjų intencijos, tačiau jos dažnai nežinomos net patiems kūrėjams, o jų rezultatas, priešingai, yra viešai prieinamas ir dažnai daro įtaką priimamiems sprendimams.

Nuo to atsispiriant galima daryti paraleles su šiuolaikinėmis DI sistemomis, kurios ir toliau skirsto pasaulį į fiksuotas kategorijas. Tai turint omenyje, naivu tikėtis, kad technologijų sukūrimo etape daromos klasifikacijos yra objektyvios. Tai signalizuoja neetišką elgesį. Žiūrint iš šios perspektyvos, DI negali būti laikomas objektyvia ar neutralia technologija. Tai įgyja dar daugiau reikšmės, priėmus, kad klasifikacijos politika – svarbus DI etapas. Net kai aptinkame, jog DI sistemos veikia diskriminaciniu pagrindu, kompanijos imasi tik techninio lygio problemos sprendimo, nemąstant, kokios socialinės ir politinės teorijos slypi ar yra sutvirtinamos tokia klasifikacijos praktika. Kitaip tariant, klasifikacija – galios aktas, pasireiškiantis, pavyzdžiui, klijuojant etiketes DI, sekant žmones pagal vaizdų atpažinimą. Kaip taikliai pastebėjo Bowkeris ir Star, klasifikacijos ištirpsta infrastruktūroje, įprotyje, todėl priimamos kaip savaime suprantamas dalykas. Nors jos naudojamos techninių sistemų formavimo stadijoje, galiausiai atlieka dinamišką rolę formuojant socialinį-materialųjį pasaulį (Bowker, Star 1999). Autorė Safiya Noble savo tyrimuose parodė, kaip plačiai DI sistemose įsigalėjusi diskriminacinė praktika – nuo veido, balso atpažinimo, iki algoritmų metamų reklamų socialinėse medijose. Net kai tokios praktikos yra atsekamos, kaip minėjau, įprasta žmonėms sakyti, kad šališkumas išspręstas, nukrypimas techniškai ištaisytas. Tačiau tokiu būdu išvengiama visuomenės diskusijos apie tai, kodėl tokios šališkos formos ir diskriminacija apskritai ten atsirado (Noble 2018). Pavyzdys galėtų būti –

2014 m. Amazon atliktas eksperimentas (Lee, Resnick, Barton 2019), kurio metu buvo naudojami algoritmai įdarbinimo procese. Kuriant sistemas buvo neatsižvelgta, kad šios mažmeninės 60 procentų darbuotojų visame pasaulyje sudaro vyrai, kurie, be kita ko, užima 74 procentus vadovaujančių pareigų įmonėje. Žmonės buvo vertinami penkiabalėje sistemoje (taip pat kaip ir produktai). Gana greitai buvo aptiktas lyčių šališkumas. Sistemos inžinieriai panaudojo bazėje turimus pastarųjų 10-ies metų darbuotojų CV (kurie daugiausia buvo baltųjų vyrų) ir pagal šiuos duomenis parengė statistinį modelį su 50 tūkstančių terminų. Programa buvo išmokyta atpažinti žodžių šablonus gyvenimo aprašymuose, o ne reikalingus darbo pozicijai gebėjimus, ir šie duomenys buvo lyginami su bendrovės inžinerijos skyriumi, kuriame vėlgi daugiausia dirba vyrai, siekiant nustatyti kandidato tinkamumą. Dėl to DI „baudė“ bet kokį gyvenimo aprašymą, kurio tekste buvo žodis „moteris“, įskaitant tai, kad mažino balą toms, kurios lankė moterų koledžus. Net ir atlikus techninius taisymus šališkumas išliko, rezultatai atspindėjo lyčių šališkumą.

Šioje vietoje įdomu pastebėti, kad verslo atstovai dažnai net nemato problemų, susijusių su duomenų klasifikavimu, todėl tiesiog nuogaštuoja dėl to, kad duomenimis pasitikima per mažai ir tai atsiliepia verslo efektyvumui, pelno siekiui. Kaip teigia Andrew McAfee ir Erikas Brynjolfssonas Harvard Business Review: „manome, kad šiandien visame verslo pasaulyje žmonės per daug pasikliauja savo patirtimi ir intuicija ir per mažai pasikliauja duomenimis. Savo tyrimams sukūrėme 5 balų sudėtinę skalę, kuri užfiksavo bendrą įmonės duomenų valdymą. Iš viso 32 % mūsų respondentų savo įmones įvertino 3 arba žemiau šios skalės“ (McAfee, Brynjolfsson 2012).

Tačiau žiūrint į „žinių kūrimo“ problemą iš mokslinės pusės reikia pripažinti, kad viskas neapsiriboja vien „duomenimis“ – tai yra rimta problema net paties mokslininkams. Juk duomenys be algoritmo tėra tik žalia medžiaga, kurią dar reikia apdoroti. DI pradeda tai daryti pagal iš anksto nustatytus šablonus, tačiau kaip pastebi Vincentas C. Mülleris, tai, kad programa nuolat mokosi, reiškia ir tai, kad keičiasi jos duomenų apdirbimo šablonai. Tad net ją sukūrę programistai galu gale nežino, kokie šablonai buvo naudojami kiekvienu konkrečiu atveju (Müller 2020). Tai reiškia, kad žmogus, negavęs darbo dėl DI sprendimo, niekada nesužinos, dėl kokios priežasties tai įvyko. Garsus JAV politikas Henry Kissingeris atkreipė dėmesį į esminę demokratijos sprendimų priėmimo problemą, jei pasikliausime sistema, kuri tariamai yra pranašesnė už žmones, bet pati negali paaiškinti savo sprendimų. Kaip gerai apibendrina, tai reikš, kad „sukūrėme potencialiai dominuojančią technologiją, kai tuo tarpu ieškojome gyvenimo gairių (ang. guiding philosophy)“ (Kissinger 2018). Tad jau DI kūrimo procese reikia ieškoti būdų, kaip padaryti DI priimamus sprendimus skaidriais bei patikrinamais. Priešingu atveju, kaip pastebi Cave (Cave 2019), mes turėsime pasaulį, aprašytą Kafkos romanuose.

Tad jau vien faktas, kad klasifikacijos problema įvardinama „technine“, liudija, kad mokslininkai nėra linkę apmąstyti etinės pradinių duomenų surinkimo problemos. Tokios pozicijos

rezultatai pasimato ganėtinai greitai – DI pradeda veikti šališkai, kaip veidrodis, atspindėdamas pradines mokslininkų nuostatas. Tad nepaisant to, kad ne mokslininkas, o DI priima sprendimą, už jo sprendimą yra atsakingas būtent DI kūrėjas.

b) Materialinės problemos

Kitas atsakomybės tipas gali būti pavadintas „materialiniu“. Jis susijęs su žaliavų, reikalingų DI sukūrimui, kilmės problema. Skaičiuojama septyniolika retų žemės mineralų, reikalingų kurti mažesnius ir lengvesnius technologinius įrenginius. Jie randami įvairiuose prietaisų komponentuose (Magnuson 2021). Šiame kontekste taip pat pažymėtina, kad kai kalbame apie atsakomybės problemą, tai ji yra ne tik santykiyje su gamta, bet ir su žmonija, nes šių mineralų išgavimas paremtas vietiniu, geopolitiniu smurtu.

Šios globalų tinklą apimančios smurto formos yra bandomos spręsti valstybiniu lygiu, tačiau nepakankamai tikslingai. Kaip pastebi Tarptautinis darnaus vystymosi institutas, JAV 2010 m. yra išleidusi Dodd-Frank aktą (USA. Commodity Futures Trading Commission, 2010), reguliuojantį šių retų metalų išgavimą, tačiau dokumente tik bendrais bruožais užsimenama apie visą grandinę dėl jų išgavimo daromos žalos. Bendrai aktas siekė reformuoti finansų sektorių dar prieš 2008 m. finansų krizę. Jame aptinkame nuorodą į taip vadinamus „konfliktinius mineralus“, kitaip tariant, natūralius resursus, išgaunamus iš konflikto zonų, o iš jų pardavimo gautą pelną naudojant toliau remti konfliktus. Vienas tokių atvejų, kurio imasi nagrinėti institutas yra Kongo Demokratinėje Respublikoje kelis dešimtmečius besitęsiantis konfliktas. Apie situacijos progresą galima kalbėti tik formaliu aspektu, kadangi iš ten mineralų išgavimo veiklą vykdančių kompanijų jau reikalaujama informacijos apie mineralų kilmę, taip pat klausiant, ar juos pardavus nebuvo finansuojama ginkluota grupuotė regione (Church, Crawford 2018). Nors šiame tebesitęsiančiame konflikte miršta tūkstančiai, o milijonai žmonių yra priversti išsikelti, kompanijų atsakomybės problema nėra pakankamai sprendžiama (Burke 2017). Konfliktas naudingas analizei ir tuo aspektu, kad dabartinės kasybos sąlygos iliustruoja nekontroliuojamą šiandieninės vergovės formą (Sabo-Walsh 2017). O grandinės viršuje – ne viena didžiulė technologijų įmonė.

Nesunku surasti pavyzdžių, kaip kompanijos veikė šių problemų akivaizdoje. Pavyzdžiui, „Intel“ daugiau nei ketverius metus vykdė tyrimą apie savo tiekimo grandinę. Padaryta išvada, kad pati tiekimo grandinė yra itin sudėtinga – daugiau nei 16 tūkstančių tiekėjų daugiau nei šimte šalių teikia tiesiogines medžiagas įmonės gamybos procesams, įrankius ir mašinas savo gamykloms, taip pat logistikos ir pakavimo paslaugas. Nepaisant pastangų, „Intel“ ir „Apple“ buvo kritikuojamos už tai, kad siekdamos nustatyti naudingųjų iškasenų statusą „be konfliktų“, tikrino tik lydyklas (*smelters*), bet ne pačias kasyklas. Kitaip tariant, technologijų milžinai į studiją įtraukė tik lydymo

gamyklas už Kongo ribų, maža to, pačius auditus dažnai patikėdavo vietos gyventojams. Dėl to abejojama, kiek pasikliauti technologijų pramonės turimais sertifikatais, liudijančiais, kad išgavimas vyko atsietai nuo konfliktų. Problema yra ta, kad tiekėjai yra susieti su tūkstančiais gamintojų, įgyjančių apdorotas medžiagas iš didžiulio skaičiaus lydyklų. Šios įmonės savo ruožtu perka medžiagas iš dar didesnio skaičiaus prekybininkų, kurie užsiima tiek legalia, tiek nelegalia kasybos veikla. Finale šie mineralai, kurių kilmę taip sunku nustatyti, patenka į vieno ar kito kompiuterio ar telefono dalį (Crawford 2021, 35). Kito kompiuterių gamintojo „Dell“ teigimu, metalų ir mineralų tiekimo grandinių sudėtingumas kelia beveik neįveikiamus iššūkius gaminant elektronikos komponentus tam, kad juos būtų galima drąsiai įvardyti kaip niekaip nesusijusius su konfliktais. Elementai keliauja per tokį didelį skaičių subjektų grandinėje, kad neįmanoma nustatyti jų kilmės (DELL 2018). Tuo remdamiesi pastarieji patys legitimizuoja galimą išnaudotojišką praktiką, nes tai tiesiogiai susiję su gaunamu pelnu. Tiekimo grandinės nežinojimas-ignoravimas glūdi pačiame kapitalizme – nuo to, kaip įmonės apsisaugo per trečiųjų šalių rangovus ir tiekėjus, iki prekių žymėjimo ir reklamavimo vartotojams. Nors dauguma naudingųjų iškasenų nėra išgaunamos iš tiesioginių karo zonų, tai nereiškia, kad žmonės nėra išnaudojami ar kad dėl to nenukenčia gamta.

Materialinė kompanijų, kuriančių DI, atsakomybė iš vienos pusės yra rimtas uždavinys pačioms kompanijoms. Tačiau iš kitos pusės galima prisiminti kaip nuolankiai žiūrėjo daugelis šalių į Bokassos ar Kadafio žiaurumus tol, kol jie aprūpindavo Vakarų pasaulį naudingais resursais. DI reikalingų medžiagų gavimo sąlygos, kaip buvo parodyta, taip pat dažnai nėra „Intel“ ar „Apple“ rūpesčio centre. O tai, savo ruožtu, skatina smurto ir *status quo* palaikymą trečiosiose pasaulio šalyse.

5.2. DUOMENŲ TVARKYMO PROBLEMA

Šiandien iš viešai prieinamų šaltinių ekspansiškai renkami visų tipų duomenys. Juos technologijų kompanijos naudoja modelių kūrimui. Ir nors patys duomenų rinkiniai yra vieši ir suprantami kaip viešoji gėrybė, sukurta metavertė tampa privačia kompanijų nuosavybe, išvalgų ir pelno šaltiniu, nebe bendro naudojimo dalimi. Kai tuo tarpu buvo tikėtasi, kad tai greit taps bendruoju turtu tik kita forma. Kitaip tariant, turime grobstomą viešąją erdvę, taip pat mažėjančią erdvę, kuriose duomenys nebūtų renkami. O kompanijos tampa vis didesniais žinių perdavimo kanalais. Duomenys lemia DI pasaulio interpretaciją, jie informuoja ir performuoja modelius per apmokymo procesą, tačiau būtent DI kūrimo užsakovų prioritetai lems, kaip ši vizija materializuosis (Crawford 2021, 121). Todėl DI kūrimas – tai socialinė ir politinė intervencija.

Pirma susijusi su duomenimis problema yra duomenų, kurių pagalba bus mokomas DI, atranka. Michał Boni pasidalino pastebėjimais dėl to, kaip DI atėjimas į rinką atrodo praktiškai ir koks technologijų santykis su verslu skirtingomis vystymo stadijomis. Pasak jo, kurdami DI sistemas

turime atsižvelgti į du aspektus. Pirmiausia, kokių duomenų reikia norint pradėti dirbti su mašininio mokymu, gilioju mokymu, kaip paruošti DI sistemą darbui. Šių duomenų, modelio kokybė yra susijusi su *ex-ante* vertinimais. Duomenų tikrinimas, ar jie tinkami, baimė, pavyzdžiui, kad būtų palaikyta lyčių pusiausvyra, žmonių lygybė ir panašiai, kad būtų išlaikytas balansas, susijęs su visų reikalavimų vykdymu, yra šios stadijos uždaviniai. Taigi kalba eina apie duomenų kokybę. Antras aspektas, klausimas, kokius modelio algoritmus paruošime – ar jie bus paprasti, su galimybe pažinti ir atpažinti efektus, ar itin aukšto lygio, susiję su visiškai nežinomomis veiklomis. Žinia, literatūroje šia tema yra algoritmų tipų taksonomija – „baltoji dėžė“, „pilkoji dėžė“, „juodoji dėžė“, „sargyba“ ir „singularumas“. Pastarasis reiškia, kad algoritmas gali pakartotinai tobulėti ir tokiu būdu pasiekti singularumą. Šio tipo algoritmo intelektas yra aukštesnis ir stipresnis už žmogaus intelektą. Kalbant apie etines problemas skirtingose DI ciklo stadijose (sukūrimo, naudojimo ir reguliavimo), pasak jo, panašu, kad jos kyla dėl „juodosios dėžės“ problemos, kai algoritmas pasižymi naujais bruožais, todėl sunku arba neįmanoma paaiškinti jo charakteristikos. Taigi, jo teigimu, jau „pačioje pradžioje turime žinoti, kokių nenumatytų pasekmių galime tikėtis kalbant apie vieną ar kitą algoritmą.“ Vis tik su kitais iššūkiais susiduriame, kai kalbame apie *ex-ante* modelį. Šis modelis, pasak politiko, leidžia kalbėti, kad yra jau turimas DI sistemos prototipas, sukurtas remiantis minėtais principais ir taisyklėmis. Iššūkis yra jo veikimą patikrinti realioje aplinkoje. Šis etapas svarbus, nes parodo, kokių nenumatytų padarinių gali sukelti DI sistema pramoninėje sferoje, pavyzdžiui, tam tikrų sveikatos priežiūros sistemos problemų. Politikas priminė, kad kai prieš 3-4 metus prasidėjo diskusija apie DI, verslui buvo aišku, kad jei eisime šiuo keliu, pirmiausia reikia koncepcijos, tada patikrinti savo principus, esamas taisykles, tada paruošti prototipą, funkcionuojantį realioje aplinkoje. Verslo požiūriu tai buvo ilgas procesas, kuomet pirmiausia patikrinama, kaip į produktą reaguoja rinka. Taigi tai apima ir ilgas diskusijas su įmonėmis dėl to, ko tikėtis ateityje, kada bus patvirtintas reglamentas. Šis *ex-ante* modelis, pasak Boni, kėlė klausimą, kaip įmonė turėtų dirbti situacijoje, kai DI sistema atnaujinama. Ar ji turėtų grįžti ir praeiti visą patikros kelią iš naujo, ar pats modelis turėtų būti atnaujinamas greičiau, o sertifikato turėjimas leistų toliau veikti rinkoje. Tai ligšiol tebėra diskusijų objektas. Politiko nuomone, reguliavimas kaip paskutinis etapas yra susijęs su *ex-ante* poveikio vertinimo ir *ex-post* priežiūros bei stebėsenos deriniu. Bendrai, tikslas yra įtraukti visus etikos principus jau DI sistemos sukūrimo procese ir atlikus atitikties įvertinimą, įgyti sertifikuotą ir pradėti veikti rinkoje. Visi aukštos rizikos produktai ir paslaugos, anot jo, turėtų būti įtraukti į *ex-post* stebėseną, kad sužinotume daugiau ir pamatytume, kas vyksta (žr. priedą nr. 2).

Tad duomenų atranka yra rimta problema. Tačiau ne mažiau rimta problema yra susijusi su tuo koku būdu gaunami šie duomenys. Žvelgiant istoriškai, iki atsirandant mašininiam mokymui ir duomenų mokslui, taikomoji matematika, statistika, kompiuterių mokslas nebuvo laikomi žmogaus tyrimų formomis. Pirmaisiais DI dešimtmečiais nors ir buvo naudojami žmonių duomenys į tyrimus

buvo žiūrima kaip į keliančius nedidelę riziką individui ir bendruomenei. Matematika ir kompiuterių mokslas buvo laikomi kur kas abstraktesniais nei psichologijos ar biomedicinos intervenciniai eksperimentai. Ir nors nuo 1980-ųjų imta DI taikyti realaus pasaulio situacijoms, pavyzdžiui, numatant kriminalinių nusikaltėlių tolesnį elgesį, DI tyrimai vis dar laikyti kaip keliantys minimalią riziką (Crawford 2021, 115-116). Tai neturėtų stebinti, žinant tokio mąstymo prielaidą – vieši duomenų rinkiniai kelia mažą riziką, todėl nepatenka į etikos sritį. Ši prielaida tebeveikia DI tyrėjų protuose (Metcalf, Crawford 2016). Iš to galime spręsti, kad kažkuria prasme tos ankstesnės prielaidos apie mašininį mokymą atsiliko nuo šiandienos progreso technologijų srityje, nuo praktikos.

Yra nemažai suinteresuotų industrijų, institucijų, individų, palaikančių kolonialistinį požiūrį į duomenis, t.y. tebe Palaikomas valios galia grindžiamą duomenų nusavinimas, naudojimas be apribojimų, jų interpretavimas be konteksto. Stebime tendenciją, kad duomenų ekstrakcijos kultūra tik auga, nepaisant problemų susijusių su privatumu, etika ir saugumu. Šiandien žinoma, kad DI sistemose yra kaupiami mūsų visų rūšių bioduomenys tam, kad būtų atrasti modeliai ir daromi vertinimai.

Privatumo pažeidimų problema masiniu lygiu pasimato garsiai nuskambėjusiu atveju, kuomet Didžiosios Britanijos Karališkojas NHS fondas 2015 m. susitarė su „Google“ dukterine įmone „DeepMind“ susitarimo pasekoje pasidalino 1,6 mln. pacientų duomenų įrašais. Šie duomenys buvo panaudoti modelių mokymai, pažeidžiant privatumo teises (Reuter 2021). Kitaip tariant, neatsiklausiant žmonių naudoti jų asmeniniai duomenys. Bylos tyrėjai apibendrina, kad „inovacijos kaina neturėtų lemti erozijos privačioms teisėms.“ Vis tik tai nei pirmas, nei paskutinis privatumo teisių pažeidimo atvejis.

Jau minėtas Vincentas C. Mülleris atkreipia dėmesį ir į kitą su duomenų nusavinimo problema susijusią problemą – manipuliaciją. Nemokamos paslaugos internete apmokamos mūsų asmeniniais duomenimis, kuriuos kruopščiai surenka internetinės kompanijos. Didžiojo 5-tuko kompanijų – Amazon, Google/Alphabet, Microsoft, Apple, Facebook – pagrindinė duomenų rinkimo dalis yra grindžiama apgaule, išnaudojant žmogiškąsias silpnybes, skatinant atidėliojimą, sukeltą priklausomybę ir manipuliuojant (Sax 2021). Mülleris pastebi, kad jau seniai egzistuoja įstatymai, griežtai kontroliuojantys azartinius žaidimus ar sukeliančių priklausomybę medžiagų pardavinėjimą, tačiau vis dar nėra reikalingų įstatymų, ribojančių manipuliacijas internete (Müller 2020). Šių manipuliacijų tikslas dvejetainis: pirmas, priversti žmogų vėl ir vėl naudoti siūlomą produktą, ką taiko Facebook, Youtube, Amazon ir panašiai, tam, kad duomenų srautas nemažėtų. Tai aktuali problema, ir aišku, kad taip naudotis DI nėra etiška. Kaip pastebi John Tasioulas, „daugelis DI naudojimo platformose atvejų, ypač socialinės žiniasklaidos platformos, yra skirti patraukti žmonių dėmesį, išlaikyti juos platformoje kuo ilgiau. Tai nėra labai geras tikslas. Taigi svarbiausias klausimas yra, ar DI naudojami tikslai yra orientuoti į tai, kad padėti žmonėms, ar užsidirbti daug pinigų pritraukiant

jų dėmesį. <...> Tai pats elementariausias klausimas“ (žr. priedą nr. 1). Antra su šiuo klausimu susijusi problema yra ta, kad jau surinkta informacija manipuliatyviai naudojama priverčiant žmogų padaryti tam tikrą sprendimą (nusipirkti produktą, kurio jam galbūt nereikia, atiduoti savo balsą partijai X ir panašiai). DI jau įgalina šiuos procesus kur kas efektyviau.

Taip, šiuo atveju duomenys nėra pagrobiami, bet jie yra surenkami naudojant manipuliacines technikas. Tai prilygsta pasakymui, kad žmogus atiduoda kazino savo pinigus savanoriškai.

Svarbu suprasti, kad duomenys yra pagrindinis DI informacinis resursas, iš kurio kuriamas galutinis produktas, tačiau tuo pačiu metu duomenys yra ir jo pritaikymo objektas. Jeigu sritį kaip tai, kaip DI tvarkys duomenis, dar bandoma reguliuoti, tai pats duomenų gavimo procesas, vykstantis DI sukūrimo etape, toli gražu nėra toks skaidrus kaip galbūt norėtūsi. Tai sritis taip pat reikalaujanti etinio reguliavimo.

5.3. EKOLOGIJOS PROBLEMA

Viena iš esminių problemų DI sukūrimo apima ekologiją. Kate Crawford nagrinėja Tesla Gigafactory, didžiausios pasaulyje ličio baterijų gamyklos atveį. Juo parodo, kad kalbėdami apie inovacijas, kaip kad Tesla kuriamus elektromobilius, pamirštame, kokia materiali jėga jas varo ir kokie to kaštai gamtai (Crawford 2021, 29). Apskaičiuota, kad Tesla kasmet sunaudoja daugiau nei 28 tūkstančius tonų ličio hidroksido, t.y. pusę viso planetos suvartojamo kiekio (Bloomberg 2018). Kaip daugelis parodė, elektromobilis nėra švari technologija, kovojant su anglies dvideginio išmetimu (Bansal, 2021). Akumuliatorių tiekimo grandinės kasyba, lydymas, eksportas, surinkimas ir transportavimas daro didelį neigiamą poveikį aplinkai ir bendruomenėms, kurias paveikė jos degradacija. Norint įkrauti elektrinį automobilį energiją reikia paimti iš tinklo, o šiuo metu tik 1/5 visos JAV elektros gaunama iš atsinaujinančių energijos šaltinių (EIA 2021). Crawford teigia, kad prekyba ir skaičiavimai remiasi būtent baterijomis, būtent įkraunamos ličio jonų baterijos yra būtinos mobiliesiems įrenginiams ir nešiojamiesiems kompiuteriams, namuose naudojamiems skaitmeniniams asistentams ir duomenų centro atsarginei energijai. Jomis taip pat remiasi internete veikiančios prekybos platformos nuo bankininkystės iki mažmeninės prekybos ir akcijų rinkos sandorių. Kadangi jos yra vienkartinio naudojimo, atitarnavusios tampa atliekomis ir taip prisideda prie planetos taršos tokiose vietose kaip Gana, Malaizija ar Pakistanas. Ji taikliai prideda teoretiko Jussi Parikka mintį, kuris siūlo galvoti apie medijas kaip apie Žemės tęsinį (Crawford 2021, 31). Apmąstydami žiniasklaidą ir technologijas kaip geologinius procesus, galime apsvarstyti neatsinaujinančių išteklių, reikalingų technologijoms, išsekvojimą. Kiekvienas DI sistemos objektas yra sukurtas naudojant elementus, kuriems susidaryti žemėje prireikė milijardų metų.

Duomenų centrai yra vieni didžiausių elektros energijos vartotojų pasaulyje (Energy Innovation 2020). Kompanijos, siekdamos efektyvumo ir skaičiavimo maksimizavimo (*compute maximalism*), paraleliai naudojant vis daugiau lustų bei kuo mažesnių ekonominių išlaidų, veikia netausodamos aplinkos ir daro žalą klimatui, kuri bus vis labiau jaučiama. Kadangi apie klimato kaitą kalbėti būtina siekiant išlaikyti įvaizdį viešojoje erdvėje, didžiosios technologijų kompanijos teigia, kad kuria švarias technologijas, kurios neprisideda prie taršos, kitos – kad ketina vis daugiau investuoti į aplinką švarinančias technologijas, t.y. siūlo spręsti su klimatu susijusias problemas naudodamos DI kaip problemų sprendimo priemonę. Pavyzdžiui, Apple ir Google teigia esančios neutralios anglies dioksido išmetimo atžvilgiu (tai reiškia, kad jos kompensuoja savo anglies emisiją pirkdamos kreditus), o Microsoft žada iki 2030 m. tapti neigiamu anglies dioksido išmetimui, t.y. užtikrinti procesą, kurio metu anglies dioksido dujos pašalinamos iš atmosferos ir izoliuojamos ilgą laiką (Smith 2020). Tačiau realybė pateikia kitokį vaizdą. Šiai dienai pasaulio skaičiavimo (*computational*) infrastruktūros išskiria tokį anglies dvideginio kiekį, kuris prilygsta aviacijos pramonės šakoms. Dar labiau sunerinti verčia prognozės. Apskaičiavimai skiriasi, bet, pavyzdžiui, tokių tyrėjų kaip Lofti Belkhir ir Ahmedas Elmeligi teigimu, technologijų sektorius iki 2040 m. sudarys 14 procentų pasaulio šiltnamio efektą sukeliančių dujų, tuo tarpu viena Švedijos komanda prognozuoja, kad vien duomenų centrų elektros energijos poreikis iki 2030 m. padidės maždaug penkiolika kartų (Jones 2018). Kitas pavyzdys galėtų būti žvelgiant į natūralios kalbos apdorojimo, vadinamus NLP, modelius. Autorė Crawford remiasi Emmos Strubell knyga, kurioje pateiktais duomenimis, naudojant vieną NLP modelį ir leidžiant DI modelių skaičiavimus, veikiančius šimtus tūkstančių valandų, išmetama daugiau nei 660 000 svarų anglies dioksido, o tai prilygsta 5 dujomis varomiems automobiliams per visą jų eksploatavimo laiką (įskaitant gamybą) arba 125 skrydžiams iš Niujorko į Pekiną (Crawford 2021, 42). Iš to galima spręsti, kad duomenų ekonomika, vykdydama savo veiklą, lieka abejinga aplinkai. Skaičiuojama, kad nuo 2012 m. vienam DI modeliui apmokėti naudojamų skaičiavimų skaičius kasmet išaugo dešimt kartų (Hernandez, Brown 2020). Galime spręsti, kad atsižvelgiant į skaičiavimo pajėgumus, reikalingus DI modeliams kurti, planetai labai brangiai kainuoja mūsų noras eksponentiškai didinti greitį ir tikslumą.

Tai, kad DI sukūrimas susijęs su konkrečiais žemės resursais, su elektros vartojimu bei su mase atliekų daro jį vienu iš svarių faktorių prisidedančių prie ekologinės krizės. Šiuo atveju ES dokumentuose, reglamentuojančiuose DI veiklą, apie žalą ekologijai DI sukūrimo metu tiesiog nėra kalbama. Atrodytų, kad turime bendrą ES ekologinę viziją. Tačiau, kaip pasirodo, ji turi būti koreguojama, juk net toks nekaltas dalykas kaip duomenų saugyklos pradeda daryti vis didesnę įtaką klimato kaitai.

6. ETINIO DI KŪRIMO PROCESO REGULIAVIMO PROBLEMA GALIOS CENTRŲ KONKURENCIJOS KONTEKSTE

Taigi, kaip matome, yra aibė etinių problemų, susijusių su DI kūrimu. Maža to, apie jas nėra nutylima – šia tema vyksta plati diskusija. Tačiau kol kas negalima pasakyti, kad situacija iš principo gerėtų. Tokių pokyčių nematyti ne vien dėl to, kad DI yra *brand new* dalykas. Daugybėje jau „tradicinių“ sferų, nešančių solidų pelną, kūrimo procesas taip pat yra ganėtinai silpnai reguliuojamas arba šie reguliavimai yra apeinami. Iš kitos pusės, tuo atveju, kai kokio nors objekto sukūrimas yra susijęs su politinės galios stiprinimu (kaip kad, pavyzdžiui, buvo ir yra su branduoliniu ginklu), etinis jo kūrimo reguliavimas taip pat neduoda tų rezultatų, kurių norėtūsi. Bandant pažiūrėti į šį reiškinį iš teorinės perspektyvos norėtūsi sugrįžti prie pačioje darbo pradžioje pristatytos galios koncepcijos bei susieti DI kūrimo etinio reguliavimo problemas su šiomis teorinėmis prielaidomis.

Galima prisiminti ne vieną pavyzdį, demonstruojantį galios centrų požiūrį į galios įrankius. Kaip kad Crawford analogija su kasybomis arba sugalvoti dar aibę pavyzdžių. Bet visi šie atvejai apimtų ekonominės galios stiprinimą. Juk kaip pastebi Timas Fountaine, Brianas McCarthy ir Tamimas Saleh, „Technologijos, įgalinančios dirbtinį intelektą, pavyzdžiui, kūrimo platformos ir apdorojimo galia bei duomenų saugojimas, sparčiai tobulėja ir tampa vis labiau prieinamos. Atrodo, kad atėjo laikas įmonėms pasinaudoti dirbtiniu intelektu. Iš tiesų, manome, kad DI per ateinančią dešimtmetį pasaulio ekonomiką papildys 13 trilijonų dolerių“ (Fountaine, McCarthy, Saleh 2019). Tad kalbama apie milžinišką pelną.

Šiuo atveju, kaip matome, etinis reguliavimas yra būtinas. Kaip pastebi Johnas Tasioulas, kalbėdamas apie dialogą tarp privataus sektoriaus ir valstybinių institucijų, „dialogo neturėtų būti tiek daug. Šioje srityje įmonėms suteikėme per daug galios“ (žr. priedą nr. 1). Tai, pasak jo, įvyko dėl dviejų priežasčių: dėl to, kad technologijų įmonės per ilgai įtikinėjo žmones, kad jos skiriasi nuo kitų kapitalistinių organizacijų, kurdamos moralių veikėjų įvaizdį. Vis tik šiandien yra aiškiau supranta, kad tai taip pat pelno siekiantys subjektai. Jie tiesiog rūpinasi savo pelno didinimu, turi daug galios, daug išteklių ir veikia visame pasaulyje. Antroji priežastis, pasak profesoriaus, kodėl jiems suteikiama tiek daug galių, yra ta, kad „daugelis įstatymų leidėjų nežino apie techninius pokyčius. Taigi jie jaučia, kad negali spręsti, kas vyksta, mano, kad turi leisti didelėms technologijoms reguliuoti save, nes mes nesuprantame šios srities“. Pasak jo, abi priežastys yra blogos. Taigi „vyriausybės turėtų palaikyti dialogą su visais, tačiau iš esmės vyriausybių darbas yra valdyti“. Deja, pasak Tasioulas, leidome didžiosioms technologijų įmonėms per daug laisvai nustatinėti darbotvarkę, įskaitant nusavintą žodžio „etika“ vartojimą. Taigi „etika“ šiandien reiškia savanorišką savireguliaciją. Profesorius užginčijo, kad etika nėra tik savanoriškas savireguliacijavimas, veikia, kad tai yra mūsų

pagrindinės vertybės, kuria būtina kartais primesti per privalomą teisę (žr. priedą nr. 1). Taigi dialogas vyksta per teisinį reikalavimą laikytis šių normų.

Tad atrodo, kad problemą galima išspręsti jeigu valstybė pradės kontroliuoti verslą. Tačiau problema yra ta, kad valstybė – tai dar vienas galios centras, kuris irgi turi būti kontroliuojamas. Pažiūrėjus į politikos pasaulį pamatysime tą patį paveikslą. Taip, pavyzdžiui, branduolinio ginklo sukūrimas nepaisant visų bandymų griežtai jį kontroliuoti arba apskritai sustabdyti dažnai buvo ir yra vykdomas nepaisant jokių reguliavimų (CTBTO 2012). Iš dalies dėl to Paskutiniojo Teismo dienos Laikrodžio rodyklės 2020 metais buvo arčiausiai vidurnakčio per visą šios mašinos egzistavimo laiką. Šiuo atveju vien potencialus pavojus persveria visą realią branduolinės energijos naudą.

Tačiau ne visi linkę matyti valstybėje galios centrą, kuris galės pasinaudoti DI. Kaip teigia profesorius Kristinnas R. Thórissonas, paklaustas kiek, jo nuomone, politiniai veiksniai įtakoja tai, kaip šiandien daromas DI, jis atsakė, kad: „Ne tiek ir daug“ (žr. priedą nr. 3). Pasak jo, valdžios institucijos tik dabar deda pastangas suprasti, kas yra DI, ką jis gali padaryti, koks gali būti šalutinis poveikis ir kokį neigiamą poveikį jis gali turėti. Profesorius pripažįsta, kad dėl ankstyvos „taikomojo DI“ stadijos net sunku tai įvertinti. Jis kalbėjo apie tai, kad dar galime bent pusę amžiaus DI bus svarbiausia technologija: „Pramonės revoliucijų istorija rodo, kad vyriausybės yra labai reakcingos, kai reikia reguliuoti technologijas, jų taikymą, valdymą ir naudojimą. DI dar mažiausiai 50 metų karaliaus tarp įtakingiausių technologijų – daugeliu atžvilgių tai tikrai yra kitas informacinių technologijų „lygis“, todėl tai, ką matome dabar, jokių būdu nėra jo viršūnė“. Atkreipė dėmesį į tai, kad be potencialų masinį naikinimą galinčių pasiekti DI produktų, politinis elitas nėra toks vieningas: „Politikai ir politinės institucijos vis dar tik pradeda vertinti, ar DI vertas jų laiko, ar ne. Išskyrus „robotus žudikus“ – autonominius mirtinus ginklus – dar turime pakankamai laiko nuspręsti, kaip reaguoti (žr. priedą nr. 3).

Tačiau šitą „akademinį“ požiūrį užginčija buvęs ES europarlamentaras Michał Boni. Jis pastebi, kad jau galima kalbėti apie DI veikimą „Realpolitik“ sąlygomis, kai tai siejama, pavyzdžiui, su dezinformacija, kai algoritmų sekos organizuoja dezinformacijos veiksmus, o iš kitos pusės, tuos veiksmus galime atpažinti naudodami DI. Tokiu būdu, pasak jo, „DI galima panaudoti prieš demokratiją, prieš kai kurias visuomenei svarbias vertybes. Be to, gali būti įmanoma organizuoti algoritmus, kad būtų sukurtas klaidingas šališkumas. Pavyzdžiui, gilios netikros naujienos.“ Vis tik net toks neigiamas DI panaudojimas politikos srityje, pasak politiko neleidžia DI laikyti galia *per se*, o veikia „galia, kuri naudoja DI kai kuriose srityse“ (žr. priedą nr. 2).

Di yra galios įrankis ir kaip toks jis kuriamas ne tam, kad galios centras kažkam tarnautų, bet tam, kad kiti tarnautų galios centrui ir stiprintų jo galią. Šiuo atveju, kaip matėme, tai, ką sako kompanijos atstovai ar politikai gali būti gryna retorika, kurios tikslas vienas – neleisti visuomenei apriboti galios centrą. Tad DI klausimas šioje vietoje toli gražu nėra naujas. Apskritai galima teigti,

kad daugelis DI problemų bet kokių atveju yra tik bendro pobūdžio problemos, kaip pavyzdžiui, privatumo ar diskriminacijos problemos. Jas turėjome iki DI, tačiau jos taip pat ryškios srityje.

DI žada dominavimą ir ekonomikoje, ir politikoje, jis apjungia savyje ir „švelniąją“, ir „kietąją“ galias. Iš vienos pusės, būdamas idealiu pasaulinio *maštabo* panoptikumo įgyvendinimu jis leidžia efektyviai rinkti informaciją ir dar efektyviau ja naudotis. Kur buvo ir ką mėgsta Misteris X, ką jis pirko savo draugei ir koks yra jo santykis su tėvais, ką jis gėdingai slepia nuo savo draugų ir artimųjų – iš vienos pusės, tai gali būti naudojama atitinkamų diskursų ir reklamų kūrimui, iš kitos, – padaro fizinę individo kontrolę be galo lengvą. Kariuomenė ir policija taip pat supranta, kad DI panaudojimas žada didžiules permainas jų srityje. Ginklo panaudojimas bus dar efektyvesnis jau vien dėl to, kad DI prognozės toli peržengia kelių štabo analitikų sugebėjimus.

DI žada galios centrą dar didesnių galios stiprinimą, todėl laukti, kad šioje situacijoje patys galios centrai pradėtų imtis realių savikontrolės priemonių DI proceso kūrimo metu būtų kiek naivu. Kaip matėme, problemos susijusios su etiniu reguliavimu prasideda jau DI supratimo stadijoje. Jeigu mokslininkas su patosu sako, kad jo nedomina politika, kad jis tarnauja mokslui ir tik mokslui, jis tuo pačiu metu tampa lengvai kontroliuojama marionete. Šia prasme galima net būtų teigti, kad visa mokslo dėl paties mokslo koncepcija yra tik dar vienas iš diskursų, tik vienas iš *soft power* būdų priversti mokslininką daryti tai, kas yra naudinga galios centrui.

Galios centras, žiūrėdamas į DI kaip į universalų galios įrankį, yra suinteresuotas turėti jį kuo greičiau ir taip laimėti konkurencinėje kovoje. Būtent todėl iškyla tokios problemos kaip duomenų nusavinimo atvejai arba diskriminuojantys programų algoritmai. Tam, kad būtų galima iš anksto išvengti tokių problemų, reikėtų patį DI kūrimo procesą padaryti žymiai lėtesniu. Tačiau, kaip buvo sakoma, tai gresia pralaimėjimu. Ta pati priežastis priverčia galios centrus atlaidžiai žiūrėti ir į problemas susijusias su žaliavomis, reikalingomis DI sukūrimui, ir su žala daroma ekologijai.

Šiuo atveju tik ženklus visuomenės įsitraukimas į DI kūrimo procesą gali pakeisti požiūrį į jį kaip į galios įrankį. Demokratizuotas DI, nustojęs būti galios centrų įrankiu, gali pasitarnauti visuomenei. Iš vienos pusės, tai išspręstų duomenų nusavinimo problemą – kiekvienas galėtų pats spręsti, ar jo duomenys gali būti naudojami DI mokymuisi. Iš kitos pusės, visuomenė galėtų iš karto kreipti DI kūrimo procesą jai naudinga linkme. Turint omenyje, kad kiekvieno visuomenės nario interesas yra stiprinti savo autonomiją bei silpninti galios centrų įtaką, galima numanyti, kad vietoje to, kad būtų kuriamos programos ir robotai, atimantys iš žmonių darbo vietas, daugiau dėmesio būtų skirta socialiniams ir ekologijos klausimams. Bet šioje vietoje jau pasigirsta utopijos natos. Kaip pastebi Ioannis Pitas, AUTH Informatikos katedros profesorius, DI ir informacijos analizės (AIIA) direktorius, komentuodamas DI prieinamumą, dabartinė padėtis yra neteisinga, žymi nelygybę ir diskriminaciją: „trečiojo pasaulio šalys yra labai nepalankioje padėtyje. Moterys taip pat neturi prieigos prie įrankių, ypač kai kalbama apie darbus, žinote, kad jos negauna aukštų pozicijų.“

Kalbėdamas apie reikalingus pokyčius, akcentavo ilgą perspektyvą: „jei norime tai padaryti tvariai, geriausias būdas yra visus Edukuoti, įskaitant moteris ir neturtingus žmones, taip pat iš skurdesnių šalių. Paklaustas, kaip visuomenė galėtų prisidėti prie sprendimų priėmimo, jis referavo į sąmoningumo didinimą: „žmonės nežino, kas vyksta. Su mūsų visuomene tai darosi vis sudėtingiau. Žmonės neturi gero išsilavinimo, o jo reikia norint suvokti problemas. Jie turėtų žinoti, kad, pavyzdžiui, jų duomenys turi vertę, kad privatumas yra vertybė, kurią reikia saugoti. Taip jie būtų labiau įsitraukę.“ Vis tik profesorius konstatavo, kad nors pokyčiai ir diskusijos apie DI būtinos jau dabar, viskas nėra taip paprasta: „tai nėra lengva. Šiuo metu yra keletas mažumų grupių, kurios garsiai apie tai kalba, bet to nepakanka. Taigi problemos, kurias aptarėme, iš tikrųjų nėra plačios visuomenės reikalas. Tarp politikų kyla tam tikras ažiotažas, bet to nepakanka. Technologijų vystymasis toks greitas, ir jei nesiimsime veiksmų, pavyzdžiui, dezinformacija gali sugriauti demokratiją“ (žr. priedą nr. 4). Taigi dabartinė padėtis neatitinka progreso masto ir nors noras padidinti visuomenės įsitraukimą į DI kūrimo procesą gali padėti susilpninti galios centrų galią bei padėti paskiram individui jaustis kiek labiau autonomišku, yra komunikuojamas, rimtiems socialiniams pokyčiams reikalingas stipresnis politinis balsas.

Tokiu būdu galios centrai, žiūrintys į DI kaip į savo galios stipriniomo įrankį, nėra suinteresuoti, kad jo sukūrimo procesas būtų etiškai reguliuojamas, nes tai susilpnintų jų įtakos mastą. Iš kitos pusės, visuomenė, norinti apsaugoti savo laisvę, privalo įsitraukti į DI kūrimo procesą. Padaryti tai nėra lengva. Galbūt norint griežtai reguliuoti šią procesą iš tikrųjų reikalingas stiprus politinis balsas. Tačiau visuomenės įsitraukimas nebūtinai turi implikuoti didelį apsiskaitymą DI srityje ar griežtų priemonių taikymą. Pradžiai pakaktų jau vien to, kad iš anksto būtų apibrėžti DI sukūrimo proceso tikslai bei uždaviniai.

IŠVADOS

1. Darbe buvo prieita prie išvados, kad tai, kaip apibrėžiamas DI, tiesiogiai koreliuoja su jo sukūrimo procesu. Kalbant apie DI jis gali būti apibrėžiamas grynai techniškai, neįtraukiant platesnio konteksto, taip pat ir politiškai – kaip vienas iš galios įrankių. Apibrėžiant DI formaliai, galima pasakyti, kad tai sistema arba mašina, kuri geba imituoti žmogaus proto veiksmus, tam kad atliktų tam tikras užduotis ir gali palaipsniui mokytis naudodamasi gauta informacija. Tokio apibrėžimo pilnai pakanka norint bent apytiksliai suprasti, ką galima padaryti DI pagalba. Tačiau DI, kaip buvo parodyta, neegzistuoja vakuume, jį kuria ir juo naudojasi žmonės, kurie turi skirtingus tikslus. Turint omenyje, kad DI yra įrankis, galintis padėti įgyventi šiuos tikslus, buvo įvestas politinis DI kaip galios įrankio apibrėžimas. Žiūrint pro šito apibrėžimo prizmę, buvo artikuliuotas DI kaip kūno ir galios centrų įrankio konceptas.
2. Analizuojant etinio DI sukūrimo stadijos reguliavimo problemą galios centrų konkurencijos kontekste buvo prieita išvados, kad iš vienos pusės DI tampa svarbiu įrankiu šioje kovoje, iš kitos – būtent tai sulėtina etinio reguliavimo procesus. Juk galia suprasta kaip jėga nukreipta į save pačią, t.y. kaip noras didinti save pačią kitų pagalbą yra sąvoka priešinga etikai. Juk etinis elgesys didina individo autonomiją, o galios veiksmai ją slopina. Suprantant DI kaip universalų galios įrankį pasimato koreliacija tarp etinio reguliavimo stokos jo sukūrimo stadijoje ir galios centrų noro panaudoti DI savo tikslams.
3. Analizuotų ES ir NATO DI strategijų atžvilgiu buvo prieita prie išvados, kad nors iš vienos pusės jose reiškiamas susirūpinimas DI panaudojimo reguliavimu, iš kitos – kur kas atlaidžiau žiūrima į jo sukūrimo proceso kontrolę. Vakarų pasaulyje suvokiama, kad DI kelia grėsmę pamatinėms demokratinėms vertybėms, todėl daug kalbama apie DI panaudojimo reguliavimą. ES ir NATO DI strategijoje kalbama apie tinkamą DI panaudojimą gynybos tikslams, apie demokratinių vertybių apsaugą, apie šališkumo mažinimą, atsiskaitomybės mechanizmų užtikrinimą. Nemaža dalimi ES ir NATO tikslai etinio DI panaudojimo reguliavimo klausimu sutampa. Tačiau tuo pačiu metu praleidžiamas etinio reguliavimo DI sukūrimo stadijoje klausimas.
4. Buvo prieita prie išvados, kad ši situacija tiesiogiai susijusi su tuo, kad DI vystymas įtraukia į save eilę skirtingų etapų, todėl tai, kas vyksta DI sukūrimo proceso metu, iki šiol

yra sunkiai atsekamas ir reguliuojamas procesas. Tai liečia ir žmonių išnaudojimą trečiojo pasaulio šalyse, ir ekologines problemas, ir atsakomybės klausimus. Studijų šių tema vis dar trūksta.

5. Analizuojant mokslininkų atsakomybės už DI daromas klaidas klausimą, buvo pastebėta koreliacija tarp nenoro prisiimti atsakomybę ir grynai techninio DI apibrėžimo. Suprantant DI vien kaip „kodą“ atitinkamai ir visos jo klaidos suprantamos kaip grynai techninės, o ne etinės. Toks DI supratimas implikuoja, kad į jo daromus sprendimus žiūrima kaip į techninę problemą. Iš kitos pusės, kompanijos suprantančios DI kaip galios stiprinimo įrankį, linkusios pažeidinėti klausimus susijusius su duomenų privatumo, medžiagų kilmės reikalingų DI sukūrimui ir ekologijos klausimais.

6. Etikos ignoravimas DI sukūrimo proceso metu yra tiesiogiai susijęs su požiūriu į DI kaip į galios įrankį. Šis įrankis žada didžiulį pranašumą prieš kitus varžovus. Būtent todėl į DI sukūrimą investuojami milijardai bei atlaidžiai žiūrima į susijusius su jo sukūrimu kaštus. Etinis reguliavimas DI sukūrimo proceso metu sulėtintų šį procesą, tačiau tai yra vienintelis būdas paversti DI iš galios įrankio visuomenės gyvenimą tobulinančiu įrankiu. Turint omenyje, kad kiekvieno visuomenės nario interesas yra stiprinti savo autonomiją bei silpninti galios centrų įtaką, galima numanyti, kad vietoje to, kad būtų kuriamos programos ir robotai, atimantys iš žmonių darbo vietas, daugiau dėmesio būtų skirta socialiniams ir ekologijos klausimams. Tokiu būdu DI sukūrimo proceso etinis reguliavimas iš vienos pusės susilpnintų galios centrus, iš kitos – padėtų visuomenei apginti pamatines demokratines vertybes.

LITERATŪROS SĄRAŠAS

1. Analytics Insight, 2021. "Artificial intelligence investment by top 10 countries", *Analytics Insight*, January 14, 2021. <https://www.analyticsinsight.net/artificial-intelligence-investment-by-top-10-countries/> [žiūrėta 2022 01 01].
2. Aneesh, Aneesh, 2002. „Technologically Coded Authority: The Post-Industrial Decline in Bureaucratic Hierarchies“. Stanford: Stanford University.
[http://web.stanford.edu/class/sts175/NewFiles/Algocratic Governance.pdf](http://web.stanford.edu/class/sts175/NewFiles/Algocratic%20Governance.pdf) [žiūrėta 2021 12 07].
3. Azure, 2022. „Artificial intelligence (AI) vs machine learning (ML)“. *Microsoft*.
<https://azure.microsoft.com/en-au/overview/artificial-intelligence-ai-vs-machine-learning/#introduction> [žiūrėta 2022 01 02].
4. Balazs, Rapi, 2019. „New Commission President Promises Legislation for Ethical AI“. <https://medium.com/@fourthquadrant/new-commission-president-promises-legislation-for-ethical-ai-3fa73cfc5d43> [žiūrėta 2022 01 01].
5. Bansal, Tima, 2021. „How Green is Tesla, Really?“. *Forbes*, May 13, 2021.
<https://www.forbes.com/sites/timabansal/2021/05/13/how-green-is-tesla-really/?sh=73535bc41576> [žiūrėta 2022 01 01].
6. Bauman, Zygmunt, 2000. *Liquid Modernity*. Cambridge: Polity Press.
7. Belkhir, Lofti, Elmeligi, Ahmedas, 2018. „Assessing ICT global emissions footprint: Trends to 2040 & recommendations“. *Journal of Cleaner Production*, Volume 177, March 10, 2018.
<https://doi.org/10.1016/j.jclepro.2017.12.239> [žiūrėta 2021 12 07].
8. Bloomberg, 2018. „Chinese lithium giant agrees three-year pact to supply Tesla“. *Bloomberg*, September 21, 2018.
<https://energy.economictimes.indiatimes.com/news/power/chinese-lithium-giant-agrees-three-year-pact-to-supply-tesla/65900761> [žiūrėta 2022 01 01].
9. Bostrom, Nick, 2014, *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
https://books.google.be/books?id=7_H8AwAAQBAJ&pg=PA1&hl=nl&source=gbs_toc_r&cad=3#v=onepage&q&f=false [žiūrėta 2021 11 15].
10. Bowker, Geoffrey, Star Susan Leigh, 1999. *Sorting Things Out: Classification and its Consequences*. Cambridge: The MIT Press.
https://books.google.be/books/about/Sorting_Things_Out.html?id=FFT5DwAAQBAJ&print

- [sec=frontcover&source=kp_read_button&hl=en&redir_esc=y#v=onepage&q&f=false](#)>
[žiūrėta 2021 09 02].
11. Brownlee, Jason, 2019. „14 Different Types of Learning in Machine Learning“. *Machine Learning Mastery*, November 11, 2019. <https://machinelearningmastery.com/types-of-learning-in-machine-learning/> [žiūrėta 2022 01 02].
 12. Burke, Jason, 2017. „Congo Violence Fuels Fears of Return to 90s Bloodbath“. *The Guardian*, June 30, 2017. <https://www.theguardian.com/world/2017/jun/30/congo-violence-fuels-fears-of-return-to-90s-bloodbath> [žiūrėta 2021 12 07].
 13. Butler, Judith, 1997. *The Psychic Life of Power: Theories in Subjection*. Stanford: Stanford University Press.
 14. Cave, Stephen, 2019. „To Save Us from a Kafkaesque Future, We Must Democratise AI“. *The Guardian*, 4 January 2019. <https://www.theguardian.com/commentisfree/2019/jan/04/future-democratise-ai-artificial-intelligence-power> [žiūrėta 2021 12 07].
 15. Church, Clare, Crawford, Alec, 2018. „Green Conflict Minerals“. *IISD*, August 2018. <https://www.iisd.org/story/green-conflict-minerals/> [žiūrėta 2021 12 07].
 16. Copeland, Jack, 2021. „artificial intelligence“. *Britannica*, December 14, 2021. <https://www.britannica.com/technology/artificial-intelligence> [žiūrėta 2021 10 13].
 17. Council on Foreign Relations, 2021. „U.S.-Russia Nuclear Arms Control 1949-2021“. *Council on Foreign Relations*. <https://www.cfr.org/timeline/us-russia-nuclear-arms-control> [žiūrėta 2021 12 15].
 18. Crawford, Kate, 2021. *The Atlas of AI*. Yale: Yale University press.
 19. CTBTO, 2012. „Nuclear Testing“. *CTBO*. <https://www.ctbto.org/nuclear-testing> [žiūrėta 2021 12 15].
 20. DELL, 2018. „Responsible Minerals Sourcing Report“, <https://i.dell.com/sites/doccontent/corporate/corp-comm/en/Documents/sourcing-report18.pdf> [žiūrėta 2021 12 07].
 21. EIA, 2021. „Electricity explained: Electricity in the United States“. *EIA*, March 18, 2021. <https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php> [žiūrėta 2022 01 01].
 22. Energy Innovation, 2020. „How Much Energy Do Data Centers Really Use?“. *Energy Innovation*, March 17, 2020. <https://energyinnovation.org/2020/03/17/how-much-energy-do-data-centers-really-use/> [žiūrėta 2021 12 07].

23. European Commission, 2018a. Artificial Intelligence for Europe. *European Commission*. Brussels, April 25, 2018. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN> [žiūrėta 2021 12 15].
24. European Commission, 2018b. Declaration of Cooperation on Artificial Intelligence. *European Commission*, Brussels, April 10, 2018. <https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/eu-declaration-cooperation-artificial-intelligence> [žiūrėta 2021 12 07].
25. European Commission, 2019a. Ethics Guidelines for Trustworthy Artificial Intelligence. *European Commission*, Brussels, April 8, 2019, 15. <https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf> [žiūrėta 2021 12 07].
26. European Commission, 2019b. Policy and investment recommendations for trustworthy Artificial Intelligence. *European Commission*, Brussels, June 26, 2019. <https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence> [žiūrėta 2021 12 07].
27. European Commission, 2020. White Paper on Artificial Intelligence: A European Approach to Excellence and Trust (COM (2020) 65 Final). *European Commission*, Brussels, February 19, 2020. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065&from=EN> [žiūrėta 2021 12 07].
28. European Commission, 2021a. Coordinated Plan on Artificial Intelligence 2021 Review“. *European Commission*, Brussels, April 21, 2021. <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review> [žiūrėta 2021 12 15]
29. European Commission, 2021b. Proposal for a Regulation on artificial intelligence. *European Commission*, Brussels, April 21, 2021. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF [žiūrėta 2021 12 15]
30. European Commission, 2021c. A European approach to artificial intelligence. *European Commission*, Brussels, September 27, 2021. <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> [žiūrėta 2021 12 07].
31. European Commission, 2021d. Excellence and trust in artificial intelligence“. Brussels. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_en [žiūrėta 2022 01 05].

32. Faulkner, Philip, Lawson, Clive, Runde, Jochen, 2010. „Theorising technology“. *Cambridge Journal of Economics*, Volume 34, Issue 1, January 2010, 1–16. <https://doi.org/10.1093/cje/bep084> [žiūrėta 2022 01 03].
33. Foucault, Michel, 1978. *The History of Sexuality*. New York: Pantheon Books.
34. Fountaine, Tim, McCarthy, Brian, Saleh, Tamim, 2019. „Building the AI-Powered Organization“. Harvard: Harvard Business Review, July-August 2019. <https://hbr.org/2019/07/building-the-ai-powered-organization> [žiūrėta 2022 01 01].
35. Gee, Alastair, 2018. „San Francisco or Mumbai? UN Envoy Encounters Homeless Life in California“. *The Guardian*, January 22, 2018. <https://www.theguardian.com/us-news/2018/jan/22/un-rapporteur-homeless-san-francisco-california> [žiūrėta 2021 12 08].
36. Hern, Alex, 2016. „Stephen Hawking: AI will be 'either best or worst thing' for humanity“. *The Guardian*, October 19, 2016. <https://www.theguardian.com/science/2016/oct/19/stephen-hawking-ai-best-or-worst-thing-for-humanity-cambridge> [žiūrėta 2022 01 03].
37. Hernandez, Danny, Brown, Tom, 2020. „Measuring the Algorithmic Efficiency of Neural Networks“. NY: Cornell University, May 5, 2020. <https://openai.com/blog/ai-and-efficiency/> [žiūrėta 2021 12 07].
38. Ho, Vivian, 2019. „Homelessness surges in San Francisco while tech's richest grow richer“. *The Guardian*, May 17, 2019. <https://www.theguardian.com/us-news/2019/may/16/san-francisco-homeless-increase-tech-ipo> [žiūrėta 2021 12 07].
39. Hoang, Lê Nguyễn, Mhamdi, El Mahdi El, 2019. *Le fabuleux chantier : Rendre l'intelligence artificielle robustement bénéfique*. Les Ulis: EDP Sciences.
40. Jones, Nicola, 2018. „How to Stop Data Centres from Gobbling Up the World's Electricity“. *Nature*, September 12, 2018. <https://www.nature.com/articles/d41586-018-06610-y> [žiūrėta 2022 01 04].
41. Kantas, Imanuelis, 1978. *Praktinio proto kritika*. Vilnius: Mintis.
42. Kapoor, Chaynika, Yadav, Ankit, Adlakha, Apoorva, 2015. „Introduction to Artificial Intelligence“. *International Journal for Sereach in Applied Science & Engineering Technology*, Volume 3, Issue IV, April 2015, <https://www.ijraset.com/files/serve.php?FID=2122> [žiūrėta 2022 01 02].
43. Kirsch, Jonathan, 2000. „Mining San Francisco's Dark Past“. *Los Angeles Times*, February 23, 2000. <https://www.latimes.com/archives/la-xpm-2000-feb-23-cl-1522-story.html> [žiūrėta 2022 01 02].

44. Kissinger, Henry, 2018. „How the Enlightenment Ends“. *The Atlantic*, June 2018 Issue, <https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/> [žiūrėta 2021 12 08].
45. Lee, Nicol Turner, Resnick, Paul, Barton, Genie, 2019. „Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms“. *Brookings: Center for Technology Innovation*, May 22, 2019. <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/#footnote-10> [žiūrėta 2021 10 03].
46. Leyen, Ursula von der, 2019. A Union that Strives for More. My Agenda for Europe. *European Commission*, Brussels. https://ec.europa.eu/info/sites/default/files/political-guidelines-next-commission_en_0.pdf [žiūrėta 2021 12 07].
47. Lingevičius, Justinas, 2021. „NATO diskusiją apie mirusias smegenis keičia dirbtinis intelektas“. *LRT*, November 17, 2021. <https://www.lrt.lt/naujienos/nuomones/3/1535372/justinas-linge vic ius-nato-diskusija-apie-mirusias-smegenis-keicia-dirbtinis-intelektas> [žiūrėta 2022 12 03].
48. Lingevičius, Justinas, Jakniūnaitė, Dovilė, 2021. „Skaitmeninė geopolitinė konkurencija dirbtinio intelekto amžiuje: Jungtinių Amerikos Valstijų, Kinijos, Europos Sąjungos vizijos“. Vilnius: Vilniaus Universitetas, Vol. 76 (2021): Darbai ir Dienos: 77-78, 86. <https://ejournals.vdu.lt/index.php/DiD/article/view/2945/1793> [žiūrėta 2022 01 05].
49. Littlefield II, William J., 2019. „The Human Skills AI Can't Replace“. *Quillette*, September 25, 2019. <https://quillette.com/2019/09/25/the-human-skills-ai-cant-replace/> [žiūrėta 2022 01 02].
50. Liu, Shanhong, 2021. „Private investments in AI by geographical area 2020“. *Statista*, April 15, 2021. <https://www.statista.com/statistics/1226538/ai-private-investments-by-area/> [žiūrėta 2021 12 07]
51. Magnuson, Stew, 2021. „U.S. Startups Seek to Claw Back China's Share of 'Technology Minerals' Market“. *National Defence Magazine*, September 7, 2021. <https://www.nationaldefensemagazine.org/articles/2021/9/7/us-startups-seek-to-claw-back-chinas-share-of-technology-minerals-market> [žiūrėta 2021 12 07].
52. Maknickienė, Nijolė, Urbonavičiūtė, Amanda, 2020. „Investavimo akcijų rinkoje naudojant giliojo mokymosi modelius tyrimas“. Vilnius: Technika. <https://www.semanticscholar.org/paper/INVESTAVIMO-AKCIJU-RINKOJE-NAUDOJANT-GILIOJO-Maknickienė-Urbonavičiūtė/41f3089ef25bbd7bc9e4e127c8caa97dceee092d> [žiūrėta 2021 11 15].

53. Mbembe, Achille, 2017. „*Critique of Black Reason*“. Durham: Duke University Press.
<https://criticaltheory.berkeley.edu/wp-content/uploads/2019/09/achille-mbembe-critique-of-black-reason.pdf> [žiūrėta 2022 12 03].
54. Mbembe, Achille, 2017. „*Critique of Black Reason*“. Durham: Duke University Press.
<https://criticaltheory.berkeley.edu/wp-content/uploads/2019/09/achille-mbembe-critique-of-black-reason.pdf> [žiūrėta 2022 01 03].
55. McAfee, Andrew, Brynjolfsson, Erik, 2012. „Big Data: The Management Revolution“. Harvard: Harvard Business Review, October 2012. <https://hbr.org/2012/10/big-data-the-management-revolution> [žiūrėta 2022 01 01].
56. Metcalf, Jacob, Crawford, Kate, 2016. „Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide.“ *Big Data & Society*, June 2016.
<https://doi.org/10.1177/2053951716650211> [žiūrėta 2022 01 01].
57. Müller, Vincent, 2020. „Ethics of Artificial Intelligence and Robotics“. *The Stanford Encyclopedia of Philosophy*, Summer 2021 Edition, ed. Edward N. Zalta,
<https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/> [žiūrėta 2021 11 15].
58. NATO, 2021. Summary of the NATO Artificial Intelligence Strategy.
https://www.nato.int/cps/en/natohq/official_texts_187617.htm [žiūrėta 2021 12 07].
59. Nye, Joseph, 2004. „Soft Power and American Foreign Policy“, *Political Science Quarterly*, vol. 119, Nr. 2, 2004, 255-270.
60. Noble, Safiya Umoja, 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
https://books.google.be/books?id=AzlDDwAAQBAJ&pg=PA1&hl=nl&source=gbs_toc_r&cad=2#v=onepage&q&f=false [žiūrėta 2021 10 07].
61. Parietti, Guido, 2018. „On the Concept of Power“. NY: Columbia University.
<https://www.proquest.com/openview/693dd98f074ee08bdc7a14e0818fccf4/1?pq-origsite=gscholar&cbl=18750> [žiūrėta 2022 01 03].
62. Parsons, Talcott, 1963. „On the Concept of Political Power“, *Proceedings of the American Philosophical Society*, vol. 107, Nr. 3, 1963, 232-262.
63. Putnaitė, Nerija, 1998. „Dorybės ir laimės ryšys Kanto „Praktinio proto kritikoje“, *Problemos*, vol. 52, 1998, 128-145.
64. Rainie, Lee, Anderson, Janna, 2017. „Code-Dependent: Pros and Cons of the Algorithm Age“. Pew Research Center, February 8, 2017.
<https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/> [žiūrėta 2021 11 15].

65. Rainie, Lee, Anderson, Janna, 2017. „Code-Dependent: Pros and Cons of the Algorithm Age“. *Pew Research Center*, February 8, 2017.
<https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/> [žiūrėta 2021 11 15].
66. Reuter, Elise, 2021. „Google's DeepMind faces suit over access to NHS patient records“. *Medcitynews*, September 30, 2021. <https://medcitynews.com/2021/09/googles-deepmind-faces-suit-over-access-to-nhs-patient-records/> [žiūrėta 2021 12 07].
67. Sabo-Walsh, Stefan, 2017. „The hidden risks of batteries: Child labor, modern slavery, and weakened land and water rights. *Green Tech Media*, March 29, 2017.
<https://www.greentechmedia.com/articles/read/green-battery-revolution-powering-social-and-environmental-risks#gs.0fX6d=s> [žiūrėta 2021 12 07].
68. Sax, Marijn, 2021. *Between Empowerment and Manipulation*. Amsterdam: University of Amsterdam, Information Law Series, Volume 47, 2021, 6.
https://www.ivir.nl/publicaties/download/Sax_INFO_47-1.pdf [žiūrėta 2021 08 08].
69. Scott, Aileen, 2021. „Difference Between Algorithm and Artificial Intelligence“. *Data Science Central*, July 6, 2021.
<https://www.datasciencecentral.com/profiles/blogs/difference-between-algorithm-and-artificial-intelligence> [žiūrėta 2021 11 15].
70. Serackis, Artūras, Jonkus, Artūras, 2014. „Pavienių žodžių atpažintuvo algoritmų derinimas ir eksperimentiniai tyrimai triukšmingoje aplinkoje“. Studentų mokslinė praktika 2014 : konferencijos pranešimų santraukos, II dalis. Projekto „Studentų mokslinės veiklos skatinimas“ leidinys. Vilnius : Lietuvos mokslo taryba, p. 237-239.
http://studentai.lmt.lt/DOKUMENTAI/KONFERENCIJOS/LMT_knyga-II_d.pdf. [žiūrėta 2021 11 15].
71. Smith, Brad, 2020. „Microsoft will be carbon negative by 2030“. Microsoft, January 16, 2020. <https://blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/> [žiūrėta 2021 12 07].
72. Smith, Tony, 2010. „Technological change in Capitalism: some Marxian themes“. *Cambridge Journal of Economics*, Volume 34, Issue 1, January 2010, 203–212, <https://doi.org/10.1093/cje/bep048> [žiūrėta 2022 01 03].
73. Stanevičius, Justinas, 2019. „Tarptautinis bendradarbiavimas dėl mirtinų autonominių ginklų: Justinių Tautų atvejis“. Vilnius. <https://talpykla.elaba.lt/elaba-fedora/objects/elaba:87356338/datastreams/MAIN/content> [žiūrėta 2022 12 03].
74. Taylor, Christine, 2021. “Structured vs. Unstructured Data”. *Datamation*, May 21, 2021.
<https://www.datamation.com/big-data/structured-vs-unstructured-data/> [žiūrėta 2021 11 15].

75. Tatum, Malcolm, 2012. „What is Machine Perception“. *EasyTechJunkie*, October 3, 2012. <https://www.easytechjunkie.com/what-is-machine-perception.htm> [žiūrėta 2022 01 02].
76. Tegmark, Max, 2016. „Benefits & Risks of Artificial Intelligence“. Future of Life Institute. <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/> [žiūrėta 2022 01 03].
77. The National Security Commission on Artificial Intelligence, 2021. Final Report. <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf> [žiūrėta 2021 12 15]
78. Thomas, Rob, 2020. „How AI Is Driving The New Industrial Revolution“. *Forbes*, March 4, 2020. <https://www.forbes.com/sites/ibm/2020/03/04/how-ai-is-driving-the-new-industrial-revolution/?sh=7e84c3ae131a> [žiūrėta 2021 12 07].
79. USA. Commodity Futures Trading Commission, 2010. *Dodd-Frank Act*. <https://www.cftc.gov/LawRegulation/DoddFrankAct/index.htm?> [žiūrėta 2021 12 07].
80. World Health Organization, 2002. „Drugs and Money“. Ed. M.N.G. Dukes, F.M. Haaijer-Ruskamp, C.P. de Joncheere, A.H. Rietveld. Amsterdam: IOS Press. https://www.euro.who.int/__data/assets/pdf_file/0011/96446/e79122.pdf [žiūrėta 2021 12 15].
81. Zuboff, Shoshana, 2019. „Surveillance Capitalism and the Challenge of Collective Action“. *New Labor Forum*, 28(1), pp. 10–29. doi: [10.1177/1095796018819461](https://doi.org/10.1177/1095796018819461) [žiūrėta 2021 11 15].
82. ТАСС, 2017. „Путин: лидер по созданию искусственного интеллекта станет властелином мира“. *ТАСС*, Сентябрь 1, 2017. <https://tass.ru/obschestvo/4524746> [žiūrėta 2021 12 07].
83. Хайдеггер, Мартин, 2018. „Европейский нигилизм“. *Нищие и пустота*. Москва: Алгоритм, pp. 73-280.

1 PRIEDAS

Pokalbis su John Tasioulas,
Oksfordo universiteto AI etikos instituto direktoriumi

1. To which extent do you think that the political factors influence how AI is done today?

It depends on what we call “political”. All technology involves value judgements. When we develop technology and think of how it is going to work, it involves a value judgement. An interesting question then is whether those value judgements will be made by political actors, but to a large extent they are made by corporations. In this way we allow political decisions to be made by corporations and this is an issue in a democracy.

2. Do you think people implement their own values in AI?

The problem is that a lot of people are presenting political questions as technical issues, so as economic issues are often presented as technical issues for bankers. But they are not because they have serious impacts on people’s well-being. So, it applies in technology. People try to present moral issues, like should we have this technology, or how should it operate as a technological issue. the same thing with covid. People who say I am following the science, but science can’t tell you how to address Covid, because it implies value judgements. Like, who bears the cost, who bears the different policies. Science can’t answer to these questions. Thus, we have to be very careful. The use “technical” as a way to push people out, saying this is just for experts, but it is just a way to give privilege to their own values. People while creating an AI not necessarily implement their own values, but they definitely implement values – maybe theirs bosses’ values, maybe peer group values, they might implement thus values shared by white man. So, it is not necessarily their personal values, but the values of some group. Usually whoever pays for something, get its values implemented.

3. How would you judge the importance given to ethical questions in AI?

Regarding the importance given to ethical questions in AI, the question is, is it the right ethics, and are the right people making a decision? However, a lot of people making a decision are powerful unaccountable economic actors, like think tank. Ethics is unavoidable. Whatever choice we make it is an ethical choice, so you can’t say .. maybe they got ethical values. If you are going to choose to use an AI for facial recognition, that is an ethical choice. It implies ethical values. If you choose to use autonomous weapons that again is an ethical choice. So it is always going to be ethics. The question is, weather it is a good ethics or bad ethics and who is making these choices.

Regarding the western approach on AI and ethics, there was an Oxford event on “EU’s AI proposal” – Ewa Keilish?. It is important to have a global approach to AI regulation, because it is in many ways a global problem, but a first step should be to get consensus at least in democratic countries. So, there are two things: first, to reach an agreement amongst democracies, second, can we agree to more minimal conditions on countries like China would expect. There are obstacles here because EU has a different approach to regulation than America. It could be that the UK, who has links with both, but is not part of the EU, maybe can bring about some kind of consensus between these two blocs – the EU and the US. The US is very resistant to have all encompassing regulation like the EU has got. So, one of interesting questions is do you have to try to have all encompassing regulation like the EU is proposing, or do you have more targeted regulation about specific matters, like facial recognition or social scoring.

4. What would be the most important ethical issues related to the development of AI?

Most important ethical issues related to AI development. A lot of issues in AI are just general problems anyway, like issues of privacy or discrimination. We already have those problems, and they are also present in AI. However, the unique problem that AI presents is humans no longer being in charge, humans surrendering their decision-making capacity to automated systems. That could be dangerous in many ways. Obvious things like people losing work or decisions that should be taken democratically, are not taken so, but taken by automated systems. But also, just how do we relate to each other in the community. Do you want a robot making a decision about whether you are getting a job, a robot making a decision whether you go to prison? These are very serious questions. We never had to answer these questions before, because this never has been a non-human decision maker. Now that there is, I think the question is, how much of our decision-making power is acceptable for us to give up to these systems in order to get some other benefits.

The particular ethical issue that could be related to the creation stage of AI is privacy/unfairness/discrimination. Why are we creating this AI in the first place? Is it a good goal it is trying to achieve? A lot of AI, for example, used on platforms, especially social media platforms, are there to grab people’s attention, to keep them on the platform for as long as possible. That is not a very good goal. So, the most fundamental question is, whether the objectives that are using AI, are using it to help people, reach genuine goals, or is it to make a lot of money getting people’s attention, getting people to buy stuff that they don’t need. That is the most basic question.

On material elements. It is part of that. If you think that AI exists somewhere in the cloud, you might be surprised that it has some environmental impacts. But before you ask about the material impact, you have to ask, why do I want to use this AI in the first place? What is the goal that I am trying to achieve? And then only if I know my goal and my goal is valuable, then I can answer is my goal justify the material impact its creation can have.

5. Are ethical questions of equal matter at creation, use and regulation stage?

Whether at creation, use and regulation stage ethical issues matter the same depends from one AI system to another, depends on what we are talking about. In some cases, there might be serious question, you might ask whether you should have such a system at all. For example, autonomous weapon systems. The creation might be the important issue, because you might think is there a really strong argument against this. In other cases, you might say it is perfectly fine to have an AI system that helps you making appointments at the doctor, but then you have to be careful whether it doesn't discriminate against woman, people at colour, etc. So, it depends on the system in question. You can't say that one stage is more important in general.

6. What do you think about the dialogue between private companies and the public sector (institutions)?

Regarding the dialogue between the private sector and public institutions, there shouldn't be a dialogue so much. We have given companies too much power in this domain. For two reasons: one, because for too long tech companies convinced people that they are different from other capitalist organisations, that they were somehow good. Now people understand more clearly that they are profit seeking entities. They are not churches, they are not special, morally good. They are just concerned on maximising their profits, and they have a lot of power, a lot of resources, and operate globally. The second reason, why they are given a lot of power is that a lot of legislators are ignorant of technical developments. So they feel they can't judge what's happening, they feel they have to give big tech to regulate themselves, because we don't understand this area. These are two very bad reasons. So, governments should be in dialogue with everyone, but fundamentally governments job is to govern. Unfortunately, we allowed the big tech so set the agenda too much. Including, we allowed them to steal the word "ethics". So that "ethics" means voluntarily self-regulation. But ethics isn't just voluntarily self-regulation. It is about our fundamental values. And sometimes we impose our fundamental values through binding law. So, the dialogue is through legally requiring respecting these norms.

7. What is the place of ecology in the debate of ethics of AI? Should it have a bigger importance in the debate?

The place of ecology. How AI affects the environment, the development, especially in poor countries, we haven't thought about it enough. And also, we didn't think about enough how it might help us to deal with climate crisis. So, again the question is, is AI used for a good purpose or not. All this ethics being made to get people to buy stuff maybe could be used to get solutions to climate change. But that would require governance to step forward because it is not something that would necessarily deliver profit. To make something for public good requires governance, because government's help us achieve things that the market can't help us achieve.

2 PRIEDAS

Pokalbis su Michał Boni,

buvusiu Europos Parlamento nariu ir Martens centro vyresnioju moksliniu bendradarbiu

1. To which extent do you think that the political factors influence how AI is done today?

We need to talk about how AI is used. Because on the one hand, you can use AI as a tool to change, for example, the medical system. So, these would be some advantages. Of course, if you are talking about power of AI of Realpolitik, it is often related for example to disinformation, when sequences of algorithms organize disinformation actions, and on the other hand, we can recognize those actions using AI. So, you can use AI against democracy, against some values, important for the societies. And also, it could be possible to organize algorithms to create false bias. For example, deep fake news. All those require using AI and algorithms. So, it can serve the negative in political area, but on the other hand, I think that we can consider AI not a power per se, but as a power by using AI in some options/areas.

2. Do you think people implement their own values in AI?

In my opinion, technology is neutral, and it is basic. But in many situations' technology can be used as not neutral. For example, we want to use remote face recognition and it can also allow to recognize emotions of humans. So, it means that technologically we can make it make because it is possible. But on the other hand, it should be forbidden. Because it is against human values and integrity. The European debate about those pyramids threats/risks from the proposal presented by European Commission. There are some unexempted risks and those practices of ai are prohibited. For example, when AI is used by people in some purposes for diminution on without understanding how important fundamental rights are, so it is the example of using AI for some purposes which are against human rights. Everything depends on how we are using it and the most important is that in EU we are working on regulation based on this risk-based approach. It means that we try to. Describe what is possible and what shouldn't be possible because we want to keep the human control over technology. The question is whether we allow to use technology against people or we are keeping some rules. // we need to remember that it is an American book based on experience of America. The EU is much more oriented to standardization and principals, this is the reason we are trying to prepare the regulations, legislations. Of course, when we are discussing in EU of implementing some ex-ante assessments (from EU commission proposal on AI legislation), it means that creators, developers, employers of AI firstly at the beginning should check if that proposed system agrees with EU values, ethical

principles. On the one hand we can have the first creators who would like to incorporate their own values, but when we are discussing with Americans in the document signed by EU and US representatives on 29th September during the trade and technology council. It was clear that both sides agree that the partners should work on common standards global standards. Standards mean that some standards relate to keeping values and principals should be incorporated in AI functionalities at the beginning. It means that privacy by design, ethics by design. This is the way to avoid the situation in which individual values will be stronger done by creator and implement to AI system. I don't know what kind of values many creators of AI present, but on the other hand this is something like accessibility of data to teach and use this machine learning to prepare AI to function. Of course, it depends on the quality of data and the balance of used data. From the statistical point of view, would you use data related only to the man view of the world, without women. So, woman have discrimination. And this discrimination in algorithms will be replicated. So, I think that it is possible that people have create AI against common values but on the other hand, the debate on standards, norms, regulations is crucial but will limit this kind of behaviour.

3. How would you judge the importance given to ethical questions in AI?

In my opinion its sufficient. Level 8/9. It is important that we are talking about ethical aspects of the functioning of AI as you remember it was established by high level expert group of AI in 2019. It was presented the view by German ethics group of AI. I think so that it's clear that all documents try to show how important ethical aspects are, but on the other hand, this is a problem, how to transfer ethical understanding and ethical principles into practical solutions. This is the key challenge. It means that ok, if we want to have AI with respect to ethical norms and principals. Firstly, we need to assess what kind of threats some AI models and systems can bring to us. This is a basis for this idea of considering this pyramid of risk. Because it means 85 percent of applications also based on AI, simple systems of AI, there is no need to make stronger checking because they are not intrusive, no possibility to create some incident for humans in physical aspects or ethical/moral aspects. But we have those 15 percent, and we are trying to access, and this model of ex ante is crucial, as you now during the consultation of AI strategy some civic organisations presented the view that when we are working on ex-ante model and impact assessment we need to use (it is a model/set of procedures) human rights impact assessment. Suggestions done by civic society was that it should be addressed to all AI systems. It should be done by third part piece – outsider of company, which are creating ai systems and models. In the final version of the proposal, there are final points important for human rights impact assessment, because this ex-ante assessment is related to the quality of data, how the set of data was created. It can create discriminatory model centre behaviour of AI or not. So,

sensitivity how to use AI in this ethical mode is very strong in EU. But of course, there are some expectations coming from the civil society that should be stronger. In my view, the proposal created this model in the proper way, the high level into account ethical issues. On the other hand, what is the added value, this is the transfer, keeping fundamental rights in systems of AI and transferring this to the consumer rights. Finally, when there is an incident, it is the example of unethical behaviour of robot or AI/sequence of algorithms, if it will be done so it should exist (and there is also a discussion how to organise it) some kind of redress mechanism, so me as users should know how to differ my opinion, what institutions to address and so on. I think the only concern is that I am not sure if the institutional framework after implementation after improving this legislation, how institutional framework will be properly ready. For example, there is a proposal of creation the European council of AI. With representatives of European Commission and European member states should participate. In my view, to the council there should also be invited representatives of the business and from the science side, to check all those issues. The last point, it was also raised explainability principal. Our fears are often come from nonunderstanding how AI is working, and this principal should give users the proper information. Shouldn't be related to technical information, because we are not prepared to understand, but general mechanism should be clear. In addition to checking point, ex-ante model should be in which we can check if it is a threat to fundamental rights this usage of AI.

4. What would be the most important ethical issues related to the development of AI?

This is a human integrity in all aspects. It means that we have our physical integrity and mental. For example, if AI could be used as model for recognition for our emotions, and as they are related to our mentality, psycho aspects. This is a question, what for some institutions would like to know about my emotions, this is the brake of integrity of my personality, this human integrity is crucial. The second issue is related to discrimination and the lack of quality. Discrimination related to the race, colour of my skin, gender, should be related to many aspects of discrimination. Of course, when we are discussing about this ethical aspect it is also related to our right not to be under surveillance, our right to know with whom we are talking. In the proposal done by European Commission there is a level in this model of risk pyramid, the level of the legit risks and there is a level of requirement that it should be addressed that all those conversational chat house should present themselves to the people for example, when they are used in marketing cause. So we have right to know with whom we are in relation. On the other hand, when we are talking about ethical issues, it is our right to access in the fair system. For example, if there are some general rating systems, some scoring of assessing people, as China propose, it is against our integrity and also as we know this is the question of avoiding dependency. If systems of AI can create some dependency of us, this is the question of senior people

with disabilities, and it should also be addressed to the children when they are in touching relation with robots, AI. I think that in general sense integrity of us as humans in physical and psychological and mental aspects with all kinds of rights which are belonging to us, we should describe in Universal declaration of human rights and all those coming after declaration.

5. Are ethical questions of equal matter at creation, use and regulation stage?

It should be. Because when you are working to create AI systems, you need to consider two aspects. First, what kind of data is needed to start with machine learning, deep learning, to prepare AI system to work. the quality of this data, model is related to this ex-ante assessments. Checking of the data if they are proper, fear, keeping for example gender balance, keeping all kinds of balance related to fulfil those requirements, related to the equality of people and so on. The quality of data. Second, what kind of model algorithms you can prepare – if it will be simple with the possibility to know and recognise the effects, or it super high level when it is related to the completely unknown activities. You know in the literature of this topic there is the taxonomy of algorithm types - white box, grey box, black box, sentinel, and singularity. Singularity means that algorithm is capable of recurse itself improvement, the algorithm in this way can achieve the singularity. Intelligence of this kind of algorithm is higher and stronger than intelligence of human. But in the question, I see more the problem of black boxes, in which algorithm exhibits emergent proprieties making it difficult or impossible to explain its characteristics. So, I think that at the very beginning, we need to know what kind of unintended consequences we can have starting with this kind of algorithm. After that, asking what does mean this ex-ante model, we have the prototype of AI system, it was established due to those principals and rules. We have prototype, and we need to check the functioning of the prototype in the real environment. The purpose of this action is important because it can show that this AI system can bring some unintended consequence in the industrial activities creating some problems in health care system. We should remember, when the discussion of AI started 3-4 years ago it was clear for the business that if we will go this way, first , we need to have the concept, then check to our principals, rules, existing rules, third, we need to prepare prototype, forth, to create the prototype functioning in the real environment and it should show us what kind of unintended consequences we can have. From the business point of view, it was the way in which the process of starting and knowing to the market with product/devises will require longer time, because you need to have this stage of analysis, how prototype is working. So, it is also the discussion with some companies, when we will have in the future, when regulation will be approved. This ex-ante model how the company should work in the situation in which the solution, the system of AI should be updated. If it should come back and go all this way, or it should be quicker and faster model of updating, having the

certification and be able and used in the market. This is under discussion. In my view, the last point, is related to the combination of ex ante impact assessment and ex post surveillance and monitoring. So, the idea is to prepare and incorporate all those ethical principals during the process of preparation of AI system and after conformity assessment with the possibility to be certified and starting at the market, all those high level risk products and services should be ex-post monitoring to know more and to see what is going on, and if we have all those checks at the beginning if we have unintended consequences or not.

6. What do you think about the dialogue between private companies and the public sector (institutions)?

It should be of course. On the other hand, when we described this AI council, I think that we need cooperation of all partners. If in this proposal you have representatives of member states and the European commission, without business and science – there is a lack of dialogue. Of course, I think that what is important is to create the plotter leathers of cooperation companies, business, which erective the areas of preparation of AI systems, but you know we need public institutions for the cooperation. But they should be independent. And the fact that the Polish government decided not to improve some conclusions that the European Council related to AI because the world “gender” was used. I could say that this government is crazy. Government should consider all principals using AI. This is the question of using AI and public institutions are using AI models, systems, algorithms, important how they are using it. If they want to detect emotions from the faces, as it was proposed by some border guards, but it is against the fundamental rights. Regarding AI, the same obligations the same focus ethical issues should be applicable to the companies, and we have this ex-ante model. It should be done by the government, by the public institutions.

7. Would you consider AI technology equally accessible — used by all?

We are in the middle of the process, of course it depends on the level of complication of AI. Some high-level AI with big possibilities to analyse data. But when we are using our smartphone devices, we are using AI. There are algorithms and so on. So, we are in the middle. But are we aware of the consequences over rights, and the debate on the privacy issues are related to this, like DMA, targeted advertisement. If we want to allow for the models using algorithms which are very useful for targeted advertisement, or should we use contextual advertisement, which is not so invasive towards our data. So we are in the middle, and the society I think should have influence. Society has an influence. Civic society, consumer organisations are influential and shaping the solution of the legislation of AI. Of course, not every proposal is well considered whether work in the Commission on the legislation last

2 years ago. But I think we have the possibility and if we have this Council of AI, so I think that via this high-level expert group, council, we us consumers inform if something is going wrong. The discussion about this redress mechanism, to give consumers clear procedures that they have some objections address to the functioning model of AI, so the formula in which consumers, society can present its view.

8. What is the place of ecology in the debate of ethics of AI? Should it have a bigger importance in the debate?

I don't want to say that AI be ecological because it's clear it should be. It is related to the energy it uses. And it does not depend on AI system, but on people and industry, which uses it. All those solutions which are changing our attitudes, in which we are going from the traditional model of contacts, organising conferences, travel, changing it to the mobile model. Smartphone allows to exchange ideas being in different parts of the world, it is efficient. If we are discussing some kind of digital doings, some prototypes, solutions, digital tools, using AI, it means that it can prepare some solutions – industrial solutions, prototypes in digital area, it must more effective, cost effective. It will bring to us solutions which are more oriented to climate goals. There are many possibilities, if we want to disseminate knowledge about us, what kind of carbon sign, we are using something, putting some signs, it is possibility. Recommendation done by UNESCO 3 weeks ago, related to AI, one of this, use impositive sense of AI to some common goods and values like fight with climate change, prepare technical solution for measuring this carbon sign. So, there are many possibilities.

3 PRIEDAS

Pokalbis su Kristinn Thórisson,
Islandijos intelektualijų mašinų instituto vykdomuoju direktoriumi, Reikjaviko universiteto
Kompiuterių mokslų mokyklos docentu

1. To which extent do you think that the political factors influence how AI is done today?

Not so much. Governments, law-making bodies, and other institutions affecting rules and regulations are just now coming to grips with what AI is, what it can do, what the side-effects can be, and what are the negative impacts that it can have. In fact, because we are at such an early stage of applied AI, most of these have not even come to light. The history of industrial revolutions shows that governments are very reactionary when it comes to regulating technologies, their application, management, and use. AI technology will reign for at least another 50 years among the most influential technologies – in many ways it really is the next “level” of information technology – and therefore what we are seeing now is not at all its peak, by any measure. Politicians and political bodies are still in the early stages of assessing whether or not AI is worth their time. With the exception of “killer robots” – autonomous lethal weapons – we still have plenty of time to decide how to react.

2. Do you think people implement their own values in AI?

If by that you mean that the ethics of AI are left to its developers, then I’d have to say ‘to some extent,’ sure. We must not forget that current AI is just like any other new technology: Its uses and applications are determined by people at companies that provide particular services and products. When they affect the bottom line positively, their use tends to be strengthened. Ethical concerns are always an afterthought because market mechanisms do not implicitly put ethics over income. Since most new applications of AI break new ground in some ways, since they have not been done before, developers have very few places to turn to except their own intuition.

3. How would you judge the importance given to ethical questions in AI?

For contemporary AI, fundamentally, ethical questions are just as important – but also not any more important – as ethical questions related to other technologies. When the application of AI crosses new thresholds and invades new territories it is important to highlight these, especially because much of the latest AI technology is so new and its limitations and unintended side-effects have no precedent.

The real fun will start when we have AI that starts to become more autonomous – that is, AI that has a sense of self and context, and can reason more like human beings.

4. What would be the most important ethical issues related to the development of AI?

There are some serious questions related to “killer robots” – machines that are programmed to act without any oversight on decisions that can end human life. These are highly relevant already, since modern technology allows for the constructions of such concoctions well beyond the ethical concerns posed by for instance land mines. Although other kinds of weaponry, such as chemical and biological warfare, are still bigger and deeper, those have to some extent regulatory frameworks that helps keep them in check. AI has none, and comes right on the heels of those kinds of technologies in terms of potential harm and threat. This issue is urgent and highly relevant already to our near and far future of human life on this planet. As for other ethical issues, nothing comes close; however, there are some misconceptions about AI that have negative side effects when ignored, such as how classifiers such as deep neural network make decisions. A so-called “self-driving car” typically has deep neural networks for interpreting data captured by sensors, and hand-coded control software for turning these into actions that the car takes. This arrangement cannot be anything more than garbage-in, garbage-out: If the deep network interpolates incorrectly a new situation, based on data it was trained on, there is nothing that the system can do – no understanding of context, uncertainty, or knowledge about the reliability (or unreliability) of its own actions. The side-effects of this have already been shown: People have died for this reason. Contemporary AI is really dumb – many like to think of it as not really being intelligence at all. The future of AI research may change this drastically, but this is the situation currently that we are facing.

5. Are ethical questions of equal matter at creation, use and regulation stage?

Not at all. Basic research continues as it has for the past 70 years or so in AI. As it should be – there is no good reason to restrict research in general information technology by law, any more than basic research in other engineering disciplines, because basic research is the generation of new knowledge. The side effects of research, however, like the application of new technologies that basic and applied research has enabled, now, that is a totally different matter. Foreseen negative side effects should of course be avoided, and if ethical concerns of any importance play into that, well, then, legal measures can and should be taken in many cases. The challenge is that much of new technology, especially that which touches on or can affect complex systems like societies, brains, biological systems, ecosystems, etc., may have side effects that are extremely difficult to foresee, or whether the benefits outweigh

the risks. Regulation of AI technologies is, as the technology itself, in its infancy. It is important that we don't move too quickly, because negative side effects of premature and inappropriate regulation may do more harm than good.

The only place where, in my opinion, we currently need to move quickly is on a ban of autonomous lethal weapons: History presents plenty of evidence, as does current evaluation of the potential harm of such technologies, to warrant aggressively limiting the development and deployment of such systems. History also presents plenty of useful methods for making such a ban work, and stick, once agreed on. But it's a challenge when some of the world's largest nations seem to be unwilling to take part.

6. What do you think about the dialogue between private companies and the public sector (institutions)?

Private companies are not in the business of assessing negative social side effects of their products or services. There is little motivation for them to disclose – or even engage in – analysis of risks and negative social impact of AI. Governments and public institutions must therefore go at it without relying on market forces. Currently there seems to be little interest in an active dialog between these “sides” of society on the topic of AI, possibly because to most people AI still seems rather abstract and unimportant in the larger picture, especially in the midst of a global pandemic. But the importance and impact of AI is – and will be – vastly greater.

7. Would you consider AI technology equally accessible — used by all? Could you consider the public could contribute to decision-making to ethics of AI? If yes, how?

AI has the potential to be equally accessible and used by citizens in all societies, provided that the level of education allows it. Primarily AI rests on information technology infrastructure, and as such doesn't require much else except know-how. The public can contribute to the discussions about the ethics of AI in many ways, including sharing experiences with AI-based technologies, products, and services, voicing their opinion via open letters to government bodies, and by educating themselves in its development, for instance via participation in AI conferences. Both scientists and government bodies can decide to actively involve the general public in such discourse via similar methods.

8. What is the place of ecology in the debate of ethics of AI? Should it have a bigger importance in the debate?

There is a general misconception that AI is somehow as dangerous to the world at large as many other threats that have been around longer, such as chemical spills, toxic waste, genetic engineering, deforestation, mining, etc. The fact is that contemporary AI is vastly less dangerous than any of these. Biology already has mechanisms for exponential self-replication and expansion – as has become uncomfortably clear with the COVID pandemic. AI has nothing of the sort. Chemical and biological warfare, oil spills, etc., are vastly more dangerous from an ecological perspective than AI, now and for the foreseeable future. That doesn't mean that AI cannot be used to speed up some of those other dangers to feverish levels, but the danger does not stem from the AI per se, in any such case, and going after the AI on those grounds would be a fool's errand.

9. Do you think we could improve the ethics of AI by better data governance? How?

As far as the civilian side of AI deployment is concerned, data governance is certainly a key factor in its ethical use. Preventing unethical or gray-area application of contemporary AI could for the most part be addressed with effective data governance. I say "contemporary" because the modern AI technologies that systems primarily depend on are very data-hungry, and generally don't work, or are simply not feasible, without availability of large amounts of data. There is reason to believe that future AI systems, however, will not be so data hungry, and thus to make our policies future-proof they must not be boxed too much in with the current state of the art in AI technologies; these will change and therefore our laws and regulations must look beyond these, to the principal foundations of the ethical concerns and civilian rights that we consider threatened.

10. Would you first look for a technical solution to the bias problem in the field of AI? Do you think it is now time to think about the ethics of AI? Or to have just better AI? Do you think solution to technical problems could improve the ethics of AI?

All of the above, frankly. We cannot afford to bet on any single one of these coming to fruition – we must hedge our bets. This is certainly more "expensive" in terms of time, energy, cost, etc., but it will significantly reduce risk. We must assume that progress in applied AI will march forward – and not simply in incremental steps but also take leaps and bounds over some periods of time. If we are too timid in thinking about the consequences of these new technologies we risk going down avenues that may be difficult to backtrack on. This means that we must allow ourselves a certain amount of "sci-

fi thinking” when discussing these issues. For this to be possible, however, the general public must get more educated in these technologies – we cannot continue to discuss “AI” as if dystopian futures like depicted in some movies are equally likely to general spying by oppressive governments; the former is still science fiction and will be for a long time, the latter is already happening. Without some framework to discuss the likelihood of the various outcomes and risks, there will continue to be a confusion about what kind of future we need to prepare for when looking 10 years ahead versus 30 or 50 years; the latter is extremely difficult to see, the former is absolutely not. Governments must listen to scientists on these issues, and fund studies in the public’s interest. The government of Iceland, for instance, has funded courses on AI for the general public, based on the long-standing experience of the Finnish government of doing the same. In many ways AI is very much like any other information technology, and so we can transfer know-how and regulation frameworks from general IT domains when dealing with it. In other aspects, however, AI is unlike anything we have seen before. The effects of automation through AI are likely to be similar to the automation of muscle power in the industrial revolution: We will be able to automate things that were impossible to automate before. A thorough review of ethics should be on the menu of any culture and government, but even more than that, so should a review of the fundamental mechanisms on which we have built markets and the global economy. The effects and impacts of AI are likely to be deep and wide.

4 PRIEDAS

Pokalbis su Ioannis Pitas,

AUTH Informatikos katedros profesoriumi, DI ir informacijos analizės (AIIA) direktoriumi

1. To which extent do you think that the political factors influence how AI is done today?

There are many political issues that are involved, sometimes in a negative way. For example, disinformation and fake news are created. Regarding governmental policies (though it depends on the country), some governments try to develop AI surveillance, or use it in social media. In most cases, as the advances are so fast, governments have to formulate policies, so they run after the events.

2. Do you think people implement their own values in AI?

Technology is not neutral at all. For example, Facebook uses AI for its own profit, other use AI for their own interest, for example, to propagate disinformation. So for both purposes. It depends on the social needs, person needs.

3. How would you judge the importance given to ethical questions in AI?

Ethical questions have big importance, because you should produce tools that advance humanity in general and wealth. EU uphold those values, however, in a rather wrong way, because they foresee only punishments, they ignore that personal data in the AI produce wealth. So they should be balanced. It is a good thing to protect private rights, but on the other hand, we should become richer. Because if Europe doesn't become richer, and China, for example, becomes, then Europe would be poor lady, that is very ethical, but has no wealth. Of course, in the past, in the US there was kind of wild West, there was no regulation, and many companies like Facebook benefited. New regulations come in, but it is kind of late.

4. What would be the most important ethical issues related to the development of AI?

There are many issues, for example, to me the biggest – data wealth internationally, owned by people in social media, who pose their data. It is stolen by several big companies, because the only word they get is that these people have a Facebook account. Massively steal personal property. Of course, sometimes this information is misused. // Essentially, at the creation stage, unless you do something on purpose, you just create data every day, and this data is in any case owned.

5. Are ethical questions are of equal matter at creation, use and regulation stage?

Regulation – is not something very different. When you talk about data, essentially you create your data, and you store and distribute them. The biggest problem is in distribution. Regarding the creation, unless you create disinformation on purpose, it is the distribution of data that is right now problematic. Distribution, in other words, communication data, the way data is communicated it is almost for free. The other thing is that they are miscommunicated. For example, fake data is miscommunicated. It is a major issue that has to stop.

6. What do you think about the dialogue between private companies and the public sector (institutions)?

They should talk to each other because it is for the benefit to everybody. Before companies did whatever they wanted, it is a domain that is highly monopolized. So, even big social media companies had now regulation. Right now, even in the US there is a big pressure to put regulation on them. For example, in the protection of minors, for example, parents should protect their children by not allowing them to stay too many hours online. Any social media companies like YouTube or Facebook, if they see that somebody fairly long stays online for long hours, they can cut him/her off automatically. And this should be done by regulation. It is not in line with companies' interest, but they will be forced to do it. So, there should be a dialogue first, but if it is not fruitful then there are already regulators in the US trying to apply antimonopoly laws and brake Facebook into pieces. In EU we don't have much such big social media companies, even though the market is big. Regarding AI research in companies, EU companies produce AI products, but not to the extent that US does. Restrictions should be done, but so that don't harm the creation of wealth.

7. Would you consider AI technology equally accessible — used by all? Could you consider the public could contribute to decision-making to ethics of AI? If yes, how?

That is certainly wrong. As it now we have strong dismiss, because for example, third world countries are heavily disadvantaged. Also women don't have access to tools, especially when it comes to jobs, you know that they don't get good managed positions. This happens in the area of AI and this should change. We live in a flat of information/ knowledge. If we want to do this in a sustainable way, the best way is that everyone is educated, including women and poor people, also from poorer countries. In order for public to contribute to decision making in AI, there should be awareness. People don't know what is going on. With our society it becomes ever more complicated. Because people don't have good education, while firstly they should be aware of the problems. Aware that for example that their data have value, that privacy is a value that should be protected. They should start being engaged.

But it is not easy. Right now there are some minority groups, who are very vocal, but it is not enough. So problems we discussed are not really a matter for general public. There is some hype among politicians, but it is not enough. Developments are so fast, and if we don't act fact disinformation for example, can destroy democracy. We saw this in the Trump era. Europe now with Covid. We should act fast.

8. What is the place of ecology in the debate of ethics of AI? Should it have a bigger importance in the debate?

Ecology is related to AI, but not directly to ethics. Ecology related to AI in many aspects, like how AI tools should be used to improve the environmental situation. The questions is, which are the best actions to protect it. And the AI could help on this, because the environment is a complicated system, so you have to model it well and start acting. So far, not so much of an action. Of course, there is a lot of regulation going on, including carbon cut emissions. But when it comes to trying to influence the environment we don't have the capacity to do it. China for example tried to remove the cloud, to have nice blue day in Beijing when they celebrated the 100th anniversary of the communist party. Weather it is something good or not – it is another story. If you start doing these things to influence the environment, you should know what you are doing. And so, the AI can help modeling the environment before starting the action. AI is not neutral that to train it you need a lot of energy. For Blockchain technology for example you need much more energy. This result in a lot of carbon print. And much could be done to produce the green AI. Used by everybody, it is an expensive excavation procedure, there should be balance. Raw materials are used by everybody, it is an expensive excavation procedure, and it is polluting. So there should be a balance there. We cannot live without rare earths. So far the only reason that this is sustainable, it is because China is not environmental friendly, and they produce cheap rare earths if they start implement good ecology rules for their excavation it would be much more expensive. In the near future, however, it will be forced to use more eco-friendly technology in their excavation.

9. Do you think we could improve the ethics of AI by better data governance? How?

I think by better data regulations you could improve the ethics of AI, with more clear and transparent rules, that are to the good of many people. In the case of AI and data it is a global phenomenon. You don't govern, you try to regulate it somehow.

10. Would you first look for a technical solution to the bias problem in the field of AI? Do you think it is now time to think about the ethics of AI? Or to have just better AI? Do you think solution to technical problems could improve the ethics of AI?

If you don't try to do it in the correct way, it can be biased, but it is not by design, this is by accident. Pure practice is in collecting data and managing data. Then you have biased AI systems. But this can be corrected providing that those who develop the system are aware of the problem and willing to solve it. But if they don't want to solve it really want to create a system that discriminating, you cannot do anything about it. For example, if you want to hire people, but for a reason you don't like women you can create system that discriminate against women, nothing can be done because those who designed the system, did this on purpose. If accidentally the system is discriminating women, but you don't have enough data, you can easily correct the system behaviour. Sometimes it is not easy because the groups that are discriminated, are not well represented, they don't give data out. Like gipsy's or Roma people, if they are discriminated and at the same time not willing to give out their data for training, then this will never stop. So, if it is a technical bias, you can easily spot it, if it is social bias, than you need political action to solve it, you need regulations. // NATO's approach is not an exact Western response to regulation needs. Right now it is European response to regulation needs. We have GTPR and an AI Act, my doubts are weather they go the correct way. Also it seems that Canada, which is advanced in AI, they have the same spirit with Europeans. The US was less proactive in regulation so far, but they see the real danger here, they felt it already with storming the Capital, so there is support now from both parties, who want to regulate. So there is a way here and it will come out soon. And it is going to be disruptive, because they see a pressing need here. Maybe a European way. But I don't think that a EU's way is the best of it, well-balanced. It is still good though as it covers the need for AI ethics, but for example it created problems for collecting research data which is unfair. It is only the AI ethics that will suffer from European universities and research. In United Nations strategy I know there is a reference to lethal autonomous weapons. So there is an effort to control it.

SUMMARY

Karolina Rimkute. *The political struggle for artificial intelligence as an ethical issue*: Master Thesis in International Relations and Diplomacy / Scientific adviser prof. T. Janeliūnas. Vilnius: Vilnius University Institute of International Relations and Political Science, 2022. – 72 p.

Keywords: *Artificial Intelligence, Politics, Ethics, International Relations, Power*

This paper examines the problem of the relationship between artificial intelligence (AI) and ethics in international relations. AI is a technology that is significantly changing our daily lives, and it is expected to solve a number of problems facing modern society. AI at the same time is seen as a technology that threatens fundamental democratic values. That is why the EU is actively trying to control the use of AI. The axis of the work revolves around the fact that although the main Western organizations (EU, NATO) allocate considerable human and financial capital to regulate the use of AI, at the same time there is not enough talk about the problems related to the development of AI at the creation stage. The paper reveals the main ethical problems related to the development of AI: discriminatory classification algorithms, the problem of data acquisition, the links between AI and the ecological crisis. It is argued that ignoring ethics in the process of creating an AI is directly related to the approach to AI, which allows it to be seen as a universal tool of power in a political context. This understanding of AI makes it one of the priority goals in the competitive struggle of major global centers of power. Ethical regulation of this process is not good for any of the centers of power, so instead of solving real problems, we are often confronted with only supposedly “ethical” rhetoric of politicians and the business world.