https://doi.org/10.24101/logos.2022.70
Gauta 2022 10 11

ROLANDAS BARTKUS
Kauno technologijos universitetas, Lietuva
Kaunas University of Technology, Lithuania

KANTO ETIKA IR MORALAUS DIRBTINIO INTELEKTO GALIMYBĖS SĄLYGOS

Kantian Ethics and Conditions of Possibility of Moral Artificial Intelligence

SUMMARY

Kantian ethics tries to rationalize morality and bring it solely to reason as "autonomous ethics of principles". The supremacy of the pure reason with respect to the sense body is marked by duty – a practical act, performed according to the law of morality. Computer principles are grounded in deontic logic (or deontological ethics). Can information systems be adapted to implement moral reasoning? What are the possibilities of reflecting moral principles or the essence of moral law in an environment of artificial intelligence? Do conditions of possibility of moral artificial intelligence exist? The answer is being searched for by investigating artificial intelligence in the light of Immanuel Kant's metaphysical reasoning as well as rethinking the categories of moral philosophy in the context of artificial intelligence.

SANTRAUKA

Kanto etika kaip „autonominė principų etika" bando racionalizuoti dorovę, suvesti vien į protą. Grynojo proto viršenybę juslinio kūno atžvilgiu žymi pareiga – praktinis poelgis, atliekamas pagal dorovės dėsnį. Kompiuterijos principai pagrįsti deontine logika (arba deontologine etika). Ar informacinės sistemos (gali būti) pritaikytos moraliniams samprotavimams įgyvendinti? Kokios galimybės atspindėti moralės principus dirbtinio intelekto aplinkoje? O dorovės dėsnio esmę? Ar egzistuoja moralaus dirbtinio intelekto galimybės sąlygos? Atsakymų ieškome tyrinėdami dirbtinį intelektą Immanuelio Kanto metafizinių svarstymų požiūriu ir persvarstydami moralės filosofijos kategorijas dirbtinio intelekto kontekste.

RAKTAŽODŽIAI: dirbtinis intelektas, Kanto etika, pareiga, dorovės dėsnis, meistriškumo taisyklės.
KEY WORDS: artificial intelligence, Kantian ethics, duty, moral law, rules of expertise.

Juk iš tikrųjų moralės dėsnis [...] mus perkelia į tokią prigimtį, kurioje grynasis protas sukurtų aukščiausiąjį gėrį, jei tik jis turėtų jį atitinkantį fizinį sugebėjimą (Kantas 1997: 60).

ĮVADAS

Šiuolaikinės dirbtinio intelekto sistemos gali būti tiek pritaikytos, tiek ir nepritaikytos atsižvelgti į moralę priimant sprendimus, todėl išlieka neetiškų sprendimų, netgi keliančių grėsmę žmogiškumui ar žmonijai, rizika (Amilevičius 2017: 36). Kiekvienu atveju aplinką atpažįstanti ir tikslo siekianti saviadaptyvi (savireguliatyvi) dirbtinio intelekto sistema pasireiškia kaip racionalus agentas (ten pat). Dirbtinis intelektas geba savarankiškai optimizuoti ir kurti specialias programas pagal uždavinio pobūdį.

„Kanto etika – autonominė principų etika“, bandymas dorovę „redukuoti vien į protą, racionalizuoti“ (Anilionytė 2006: 56–57). „Kantas ieško gryno dorovės pagrindimo galimybės, o grynas, jo nuomone, gali būti tik protas, tad jis ir esąs dorovės principų šaltinis“ (ten pat: 57). Kiekvieną moralų poelgį įgalina tik gryna (empiriškai nesąlygota) valia, savo maksimomis atitikdama „visuotinį ir būtiną“ moralės principą arba „kategorinį imperatyvą“ kaip „besąlygišką reikalavimą, arba dėsnį“ (Miniotaitė 1988: 142). Taigi valia, pareiga, privalėjimas yra proto kuriami produktai, o „gėrio galima ieškoti ten, kur pasirodo protas, ir ne vien pasirodo, bet ir užima visą elgesio situaciją. [...] Protas yra tiek dorovės dėsnių kūrėjas, tiek ir jų vykdytojas“ (Anilionytė 2006: 57–58). Kitaip tariant, etika suprantama kaip dorovės metafizika.

Kategorinio imperatyvo (moralės dėsnio) racionalumo samprata pasiūlo prielaidą, kad turėtų būti galimybė sukurti dirbtinį intelektą, veikiantį pagal racionalaus elgesio algoritmus ir kartu paklūstantį dorovės metafizikai. Tiesa, sprendžiant šį klausimą, tenka persvarstyti Immanuelio Kanto moralės filosofijos kategorijas dirbtinio intelekto kontekste, kiek ir kaip jis pajėgus atitikti klasikinės vokiečių filosofijos pradininko formuluojamus etinio turinio reikalavimams.

Iš pradžių dirbtinis intelektas vadintas *mašininiu intelektu*, *mašininiu mąstymu*; lig šiol ginčytina jo analogijos reikšmė žmogaus mąstymui (protui, sampročiui, sąmonei). Apie žmogaus sąmonę ir dirbtinį intelektą dažniausiai kalbama ne jų prigimties bendrumo (arba ontologinio atitikmens), bet principų, požymių, funkcijų – analoginio panašumo – požiūriu. Juk suprasdami, kad veidrodyje pačių (tikrųjų) daiktų nėra, iš jų atspindžių galime spręsti apie judėjimą ir formas. Taigi nors dirbtinio intelekto kontekste tokios Kanto dorovės metafizikos sąvokos kaip paskata, suinteresuotumas, maksima, pagarba, noras, laisvė, laimė etc. tiesiogiai neaptinkamos, stebimi jų pėdsakai (atspindžiai). Tačiau jeigu dirbtinio intelekto aplinkoje nėra galimybės atspindėti pačios dorovės dėsnio esmės, paneigiama šiuo principu paremta moralaus dirbtinio intelekto (agento) idėja.

DEONTOLOGINĖ ETIKA DIRBTINIO INTELEKTO STRUKTŪROJE

Duali – protingoji ir juslinė – žmogaus prigimtis pasireiškia pareigos ir polinkio priešprieša. Pareiga – praktiniu poelgiu, atliekamu pagal dorovės dėsnį ir pašalinančiu „visus iš polinkio kylančius determinantus“, – žymima grynojo proto viršenybė juslinio kūno atžvilgiu (Kantas 1987: 100). Jei valia atitiktų grynąjį dorovės dėsnį, ji būtų š v e n t a, ir pareigos dėsnį pakeistų š v e n t u m o dėsnis (ten pat: 101). Kadangi žmogus kaip sukurtoji būtybė šventos valios neturi (ten pat: 103), mūsų santykis su moralės dėsniu determinuojamas p a g a r b a dėsniui ir pagarbumu pareigai (ten pat: 102). Kokios dirbtinio intelekto galimybės atspindėti šiuos moralės principus?

Kompiuterio duomenų bazę sudaro laukai (apriorinė forma potencialiam turiniui) ir įrašai (formos turinys). Kanto įvardintų „laisvės kategorijų gėrio ir blogio atžvilgiu“ lentelėje (ten pat: 85) galima įžvelgti „dorovės failo“ struktūrą. Bendri, universalūs principai – kiekybės, kokybės, santykio ir modalumo kategorijų grupės, kiekviena apimanti po tris kategorijas (taigi dvylika iš viso) (ten pat: 85–90) – yra tarsi „atraminė konstrukcija“ steigti moralaus dirbtinio intelekto duomenų bazės laukų struktūrą. Suformuotus (arba tebeformuojamus) laukus užpildo įrašai; beje, panašiai žmogaus intelektas apdoroja (registruoja ir klasifikuoja) patyrimo medžiagą.

Programos agentas, nevykdantis užduoties, būtų prastas agentas, o kompiuteris, nepaklūstantis pareigai, – prastas kompiuteris. *Deontologinė* (gr. *δέον* – pareiga, prievolė, *λογος* – mokslas) *etika* kyla iš būtinybės. Čia veikia vienintelis paklusnumo dėsniui, pareigai, procedūrai, taisyklei principas. Kompiuteriu palaikoma (angl. *computer-supported*) kompiuterio etika konstruojama pradedant deontinės logikos (arba deontologinės etikos) principais (Hoven, Lokhorst 2002: 280–287). Deontiniai modeliai, realizuoti kompiuterinėmis programomis, virsta informacinėmis sistemomis, pritaikytomis moraliniams samprotavimams įgyvendinti (ten pat). Aukščiausiam tikslui – visuotiniam ir privalomam dėsniui – subordinuojant visas užduotis (veiklas) programinėje aplinkoje, būtų išvengta žmogui būdingų motyvų, kylančių iš proto ir kūno priešpriešos ir pasireiškiančių subjektyviu dėsnio bei maksimos santykiu[1].

Kuriant moralinius principus atspindinčias (arba juos įgyvendinančias bei palaikančias) informacines sistemas, derinami deontinės, epistеminės[2] ir veiksmo[3] logikos operatoriai ir gaunami mišrūs sprendiniai; tokiu būdu konstruojamos hibridinės[4] sistemos (ten pat: 284–287). Iškyla nuoseklumo ir pirmenybės (Wright 2022: 225), laisvės, tikėjimo, pasitikėjimo ir priežiūros (sekimo), moralės ir teisės prieštaravimai (konfliktai) (ten pat: 224). Ieškoma sprendimo mechanizmų šiems prieštaravimams įveikti: einama norminimo (įsakmių prievolių) keliu arba tenkinamasi vien aprašomuoju pobūdžiu (ten pat: 233–234), kuriamos lanksčios normatyvinės sistemos, įgalinančios sklandaus išplėtimo, suspaudimo, peržiūrėjimo funkcijas (ten pat: 235), bandoma pritaikyti daugiareikšmę logiką (ten pat: 233) ir kt. Galiausiai pripažįstama, kad neįmanoma

sukurti „etiškų“ mašinų, veikiančių pagal žmogiškos etikos principus, – nepasiteisina „moralių mašinų eksperimentas“ (ten pat: 235).

Moralių mašininių agentų sukūrimo idėja iš etikos perkeliama į teisę, tikintis, kad teisinei sistemai adaptuota deontinė logika išspręs pareigų konfliktus (ten pat). Juk, regis, išleidus įstatymus „Privaloma sustoti prie *Stop* kelio ženklo“ ir „Negalima sustoti prie karinės bazės“, tiesiog nelogiška būtų *Stop* ženklą pastatyti būtent tokioje vietoje (ten pat: 233). Tačiau ar šitaip realizuojant pareigos etiką gali pasireikšti Kanto dorovės dėsnio esmė?

DOROVĖS DĖSNIO IŠŠŪKIAI DIRBTINIAM INTELEKTUI

Kanto suformuluotas visuotinis ir būtinas dorovės (moralės) dėsnis kaip kategorinis imperatyvas įsako: „*elkis tik pagal tokią maksimą, kuria vadovaudamasis, tu kartu galėtum norėti, kad ji taptų visuotiniu dėsniu*“ (Kantas 1980: 51–52). Antroji dėsnio formuluotė reikalauja: „*elkis taip, kad nei savo asmenyje, nei kieno nors kito asmenyje niekada žmogaus nepanaudotum vien kaip priemonės, o visada kaip tikslą*“ (ten pat: 62). Abi formuluotes apima trečioji: „elkis pagal maksimas nario, kuris nustato visuotinius dėsnius galimai tikslų viešpatijai“ (ten pat: 75).

Kategorinis imperatyvas draudžia neuniversalizuojamas maksimas, žmogaus sudaiktinimą, išorinę prievartą (Miniotaitė 2006: 21–22). Kiekvieną pareigą subordinuojant aukščiausiam tikslui – kategoriniam imperatyvui, visuotiniam ir privalomam dorovės dėsniui, – iškyla problema. Ją yra parodęs pats Kantas: remiantis jo siūlomu moralumo kriterijumi, neįmanoma sukurti visiems vienodai tinkančią normatyvų sistemą (ten pat: 22). Pavyzdžiui, egzistuoja prieštaringas atsakymas į klausimą „Ar pateisinama savižudybė dėl tėvynės gerovės?“ Tobula savisaugos pareiga savižudybę draudžia, o tobula pareiga gerbti žmones – leidžia. „Išlikti gyvam reiškia pažeminti žmogaus orumą savo asmenyje, o pasielgti priešingai – matyti savyje tiktai priemonę“ (Кант 1965: 360–361, cit. iš: Miniotaitė 2006: 22). Kantas pasirinkimą redukuoja į polinkio ir pareigos susidūrimo schemą (Anilionytė 2006: 60). Kategorinis imperatyvas pajėgus efektyviai spręsti pareigos ir polinkio konfliktus, bet anaiptol ne kiekvieną pareigų konfliktą (Miniotaitė 2006: 22). Jeigu meluoti draudžiama, tai kaip atsakyti naciui, kai šis klausia, kur slepiasi žydai (Schmidt 2002: 209)? Ir atsakydamas į netobulos pareigos vykdymo privalėjimo klausimą, Kantas vienu atveju teigia, jog nevykdymas yra netoleruotinas, kitu – „kad pareiga gerbti kitus reikalauja toleruoti tobulos pareigos nevykdymą“ (Miniotaitė 2006: 21).

Kantas nurodo p a g a r b o s moralės dėsniui jausmą, atsirandantį „intelektualiniu pagrindu“ (arba tiesiog *moralinį* jausmą) (Kantas 1997: 93). Nors „subjektas neturi jokio p i r m e s n i o jausmo, kuris linktų į moralumą“ (ten pat: 94–95), bet „pagarba yra poveikis jausmui, taigi protingos būtybės juslumui“ (ten pat: 95). Viena vertus, „ji turi kaip prielaidą šį juslumą“, ir antra – „pagarba dėsniui yra ne dorovės paskata, bet pati dorovė“ (ten pat). Pagarbos jausmas yra

pajauta, kurios „priežastis glūdi *grynajame praktiniame prote*“ (kursyvas mano. – *R. B.*), skatinanti moralės dėsnį „paversti mūsų maksima“ (ten pat).

> Pagarba visada reiškiama tik asmenims ir niekad nereiškiama daiktams. Pastarieji gali sužadinti mūsų polinkį [...], bet niekad negali sužadinti pagarbos. [...] Žmogus man taip pat gali būti meilės, baimės ar stebėjimosi, netgi nuostabos objektas, tačiau dėl to jis netampa pagarbos objektu. [...] Aš galiu pridurti: paprastam, nekilmingam žmogui, kurį aš matau esant tokio doro būdo, kokio neįsisąmoninu paties savęs, lenkiasi mano dvasia [...]. Jo pavyzdys man rodo dėsnį, sutriuškinantį mano savimaną, kai aš jį gretinu su savo elgesiu ir matau, jog veiksmu įrodyta, kad dėsnio buvo laikomasi, taigi įrodytas jo įvykdomumas. Aš netgi galiu būti įsisąmoninęs, kad esu toks pat doras, ir vis dėlto pagarba išlieka (ten pat: 95–96).

Per pagarbos jausmą mes suvokiame privalomąjį dorovės dėsnio galiojimo pobūdį. Taigi Kantas, dorovės dėsnio įrodymui pasitelkdamas ne kokį nors dedukcinį ar indukcinį samprotavimą, bet tam tikrą savaiminio įrodinėjimo rūšį – moralinį intuityvizmą, išeina už grynojo racionalizmo ribų (Schönecker 2022: 175–176). Kategorinis („tvirtai įsitvirtinęs“) imperatyvas niekaip neišskaičiuojamas (ten pat: 175). Tiesa, moralinis sentimentalizmas Kantui svetimas: moralinis jausmas parodo ne kategorinio imperatyvo turinį (ką privalome atlikti ir ko išvengti; tai jau yra proto prerogatyva), bet kategorinį dorovės dėsnio privalomumą (ten pat: 176). Tokia grynajame praktiniame prote glūdinti praktinė priežastis (pagarba) yra neprieinama kompiuterio sąrangai (ten pat); taigi neprieinamas ir pats dorovės dėsnis.

Spontaniškumo, galinčio pradėti veikti savaime, idėją Kantas įvardija transcendentaline laisvės idėja (Kantas 1987: 387). „Žmogus, kad ir priklausydamas jutimais suvokiamam pasauliui, kur viskas griežtai determinuota, vertina savo ir kitų elgesį, tarsi jis būtų laisvas ir atsakingas“ (Miniotaitė 1988: 147). Dorovinės laisvės galimybė kyla iš žmogus priklausymo dviem pasauliams – jusliškai suvokiamam ir suvokiamam protu (ten pat). „*Kategoriškas* privalomumas sudaro apriorinį sintetinį teiginį dėl to, kad prie mano valios, kurią veikia jutiminiai troškimai, prisideda dar idėja tos pačios valios, bet priklausančios intelekto pasauliui, grynos ir pačios sau praktinės valios“ (Kantas 1980: 94). Tik abiejų komponentų – *aprioriškumo*, glūdinčio grynajame prote ir jį atitinkančioje grynojoje valioje, ir *sintetiškumo*, kylančio iš empirinės valios santykio su jusline tikrove, – *sintezė* įgalina apriorinį sintetinį teiginį.

Sugebėjimas rinktis tik tai, ką protas pripažįsta geru, yra dorovinė laisvė (Miniotaitė 1988: 146), o „laisva valia ir dorovės dėsniams paklūstanti valia yra tas pat“ (Kantas 1987: 85). Taigi svarbiausia protingos būtybės valios savybė yra jos autonomija arba laisvė. Kokias laisvės galimybės sąlygas suteikia dirbtinio intelekto aplinka?

Esminis dirbtinio intelekto bruožas yra griežta sankiba (susiliejimas) su aplinka (imanencija), būdinga gyvūno prigimčiai. Žmogus pasižymi kūrybinio jautrumo aplinkai ir jos objektiškumo refleksijos pusiausvyra (transcendencija). Dėl to ir gyvūnai, ir kompiuteriai negali juoktis, verkti, meluoti ar apgaudinėti.

Dirbtinis intelektas visuomet yra imanentiškuose mikropasaulio spąstuose. Dirbtinio intelekto mikropasaulio kūrimu siekiama sukurti uždarą virtualių objektų, savybių ir ryšių sritį (domeną) apibrėžiant komandų reikšmes ir sudarant sąlygas deramai reaguoti uždaroje aplinkoje (Beavers 2002: 67).

Žmogaus ir dirbtinio intelekto sprendimų skirtumai išryškėja analizuojant jų *klaidų* ir *klaidų vertinimo* (kurie sprendiniai yra arba gali būti klaidingi) *pobūdį*. Dirbtinį intelektą gali suklaidinti objektų panašumas, dydis ar ryškumas (Schlicht 2022: 22). Atsitinka taip, kad žmonės priskiriami goriloms, o liūtas – bibliotekai (beje, su dideliu patikimumo laipsniu) (ten pat: 22–23). Ir dirbtiniam intelektui trūksta anaiptol ne kokių nors gebėjimų (klasifikuojant kategorijas, skiriant požymius ar nustatant kriterijus), bet visuminio tikrovės modelio (visuminės sampratos) (ten pat: 26–27). Vyksta „kažkas, kas labai skiriasi nuo žmogaus suvokimo“ (ten pat: 23).

Žmogus bet kada gali suklysti ir, kita vertus, sugeba priimti tinkamus sprendimus nenumatytais atvejais bei juos koreguoti. Įsivaizduokime situaciją. Vandenyne skęsta vaikas, ant kranto – jo motina, kojos prispaustos nepajudinama betono plokšte, rankose kibirėlis. Motina gelbsti vaiką – semia kibirėliu vandenį, nors žino, kad vandenyno neišsems. Mamos užduotį perleidus dirbtiniam intelektui, šis, įvertinęs sąlygas ir galimybes, pateiktų sprendinį: „Išgelbėti negalima.“ Jei žmogus gyventų pagal instinktus ar įpročius, jo elgesys būtų iš esmės nuspėjamas kaip kompiuterio ar tam tikros rūšies gyvūno. „Kartais žmogus elgiasi visiškai nelogiškai gamtos dėsnių ar kompiuterinės logikos požiūriu, ir tik tame glūdi žmogaus didybė, esminė žmogaus ir gamtos skirtis“ (Bartkus 2008: 12).

„Atsakymas į klausimą „Ką aš privalau daryti?“ nepriklauso nuo atsakymo į klausimą „Ką aš galiu žinoti?“; tai, kas privalo būti, neišvedama iš to, kas yra“ (Miniotaitė 1988: 149). Nepriklausomai nuo tikrovės aprašo, patyrimo ar patarimo, pats žmogus savo protu nusprendžia, ką jam daryti.

> Kantui žmogaus moralinė pozicija yra konstruktyvi, tačiau nesukonstruojama. Moralė – tai amžinas tapsmas. Kantas pabrėžia, kad žmogaus moralumą galima suprasti tik kaip pradžios neturintį judėjimą pirmyn. Objektyviai jis yra nepasiekiamas idealas. Žmogaus dorovinė pozicija niekada negali būti ramybės būsenoje. Jei dorumas tampa įpročiu, žmogus praranda laisvę (ten pat: 160).

Suprantame, kad išdėstytų sąlygų, išreiškiančių žmogaus ir dirbtinio intelekto specifiką dorovės dėsnio požiūriu, jau pakanka paneigti moralaus dirbtinio intelekto sukūrimo galimybę.

DESTRUKTYVI DIRBTINIO INTELEKTO REALYBĖ

Iš technologijų progreso laukdami kuo didesnio dirbtinio intelekto panašumo į mus, deja, patys su juo supanašėjame. Šiuolaikinės technologijos objektyviais, beasmeniais, „išoriniais“ pakaitalais „industrializuoja“, „protezuoja“, „schemizuoja“ sąmonę išstumdamos individualų, vidinį, laiko tėkmės sąlygotą santykį su tikrove (Vidauskytė 2021: 74–75). Užuot įgyvendinęs mūsų norus

ar įkūnijęs moralinius samprotavimus, dirbtinis intelektas „vagia“ mūsų žinias ir gebėjimus, kūrybinius pajėgumus, netgi norus ir valią (ten pat: 72). Nusitvėrę mokymosi visą gyvenimą idėją jau nesugebame užaugti ir subręsti, susidaryti išbaigto pasaulio vaizdo (ten pat).

Svarstėme teorinę galimybę dirbtinio intelekto aplinkoje išvengti motyvų, kylančių iš proto ir kūno priešpriešos. Nors programavimo industrijos „pagauta“, jos kontroliuojama ir pažeminta sąmonė išgyvena „troškimų demotyvaciją“ (ten pat: 76), tai neturi nieko bendro su Kanto moralės etikos nurodytu įsisąmonintu ir savanorišku asmeninių motyvų suvaldymu iš pagarbos dorovės dėsniui. Realybė priešinga: šiuolaikinės technologijos kaip tik tarnauja vartotojiškai „libido ekonomikai“, technika užpildo troškimų erdvę (ten pat)[5].

Naująjį moderniosios civilizacijos „racionalumą“ yra taikliai apibūdinęs Hansas Georgas Gadameris:

> Dabar protinga tėra rasti teisingas priemones iškeltam tikslui pasiekti, netikrinant, ar protingi patys tikslai. Todėl moderniosios civilizacijos aparato racionalumas galų gale yra racionali beprotybė, savotiškas priemonių sukilimas prieš tikslų viešpatiją, trumpai sakant, išlaisvinimas to, ką visose gyvenimo srityse vadiname „technika“ (Gadamer 1999: 24).

Antropotechninė visų minėtų dabarties negandų priežastis yra *meistriškumo taisyklių* suabsoliutinimas. Realiais pavidalais ir milžiniškais mastais išaugusių grėsmių šaknis parodyta *Dorovės metafizikos pagrinduose*:

> Kiekvienas mokslas turi kokią nors praktinę dalį, kurią sudaro užduotys, nurodančios, kad koks nors tikslas mums yra galimas, ir imperatyvai, nurodantys, kaip tą tikslą galima būtų pasiekti. Tokius imperatyvus apskritai galima vadinti meistriškumo [...] imperatyvais. Ar tikslas protingas ir geras, – toks klausimas čia visai nekeliamas, o kalbama tik apie tai, ką reikia daryti, *kad* jį pasiektume (Kantas 1980: 43–44).

Meistriškumo *taisyklės* ir išmintingumo *patarimai* iš esmės skiriasi nuo dorovės *įsakymų (dėsnių)* (Kantas 1980: 45).

> Pirmieji ir antrieji imperatyvai negali būti praktiniai dėsniai, nes juose formuluojamas tikslas susijęs su jusline žmogaus prigimtimi. Tai empiriniai principai. Tokiuose principuose pasirinkimas remiasi įsivaizduojamu objektu ir tuo jausmu, kurį sukelia jo pasiekimas (malonumo ar nemalonumo jausmu). Apie jokį įsivaizduojamą objektą negalima *a priori* žinoti, ar jis sukels malonumą, ar ne (Miniotaitė 1988: 145).

Praeityje madingas techninių mokslų lozungas „Technines problemas galima išspręsti pačia technika“ šiandien nekelia pasitikėjimo. Racionalios technikos, euristikos[6] ar analogijos[7] priemonės nepajėgios išspręsti technikos sukurtų problemų. Žinomas informatikų posakis „*Garbage in, garbage out*“ („Šiukšlės – įėjimas, šiukšlės – išėjimas“). Egzistuoja ir tokia grėsmių formuluotė: „Pažeidimas – įėjimas, katastrofa – išėjimas“ (Инхетвен 1989: 362–363).

IŠVADOS

Kompiuterinių principų pagrindimas deontine logika (arba deontologine etika) pasiūlo moralaus dirbtinio intelekto galimybės prielaidą. Deja, tokių informa-

cinių sistemų, kuriose visi veiksmai būtų subordinuoti dorovės dėsniui konstravimas yra problemiškas.

Kategorinį dorovės dėsnio privalomumą parodo pagarbos moralės dėsniui jausmas. Ši pajauta kyla iš grynojo praktinio proto ir skatina moralės dėsnį paversti mūsų maksima. Dirbtinis intelektas įkalintas imanentiškuose mikropasaulio spąstuose. Jis neturi transcendentalumo galimybės, be kurios negali būti pagarbos jausmo, pasirinkimo laisvės ir kartu dorovės.

Dorovinė laisvė yra sugebėjimas rinktis tai, ką protas pripažįsta geru, – turėdama galimybę apsispręsti „už" ir „prieš", laisva valia *pasirenka* paklusti dorovės dėsniui ir sutampa su dorovės dėsniams paklūstančia valia.

Dorovės dėsnis galioja kategoriškai, bet veikia reguliatyviai. Mėginant išpildyti jo reikalavimus (vykdant pareigas), susiduriama su neapibrėžtumu ir / ar neužbaigtumu. *Apsispręsti* būtina net egzistuojant klaidos rizikai.

Destruktyvios šiuolaikinių technologijų apraiškos demonstruoja, kaip dirbtinis intelektas užbaigia technikos priemonių sukilimą prieš tikslų viešpatiją. Šis reiškinys kyla iš meistriškumo taisyklių suabsoliutinimo. Pats dirbtinis intelektas kaip techninis produktas paklūsta empirijos dėsniams arba meistriškumo imperatyvams, kitaip tariant, atspindi ir produkuoja meistriškumo taisyklių pasaulį.

Dirbtinio intelekto aplinkoje nėra galimybės atspindėti dorovės dėsnio esmės, t. y. neegzistuoja moralaus dirbtino intelekto galimybės sąlygos.

Literatūra

Amilevičius Darius. 2017. Dirbtinis intelektas ir besiformuojančių technologijų etika, *Naujasis židinys-Aidai* 5: 31–36.

Anilionytė Loreta. 2006. Kanto etikos monologinis formalizmas, *Logos-Vilnius* 46: 56–63.

Bartkus Rolandas. 2008. Sistema ir dvasia: žmogiškosios tvarkos kalėjimas ir išlaisvinanti Dievo malonė, *Soter-Kaunas* 28 (56): 7–22.

Bartkus Rolandas. 2019. Mokslinių paradigmų kaita Th. S. Kuhno ir P. Engelmeierio teorijose, *Logos-Vilnius* 98: 64–75.

Beavers Anthony F. 2002. Phenomenology and Artificial Intelligence, Moor H., Bynum T. W. (eds.). *Cyberphilosopphy. The Intersection of Philosophy and Computing*: 66–77. Oxford: Blackwell Publishing Ltd.

Gadamer Hans-Georg. 1999. *Istorija. Menas. Kalba.* Vertė A. Sverdiolas. Vilnius: Baltos lankos.

Hoven Van den Jaroen, Lokhorst Gert-Jan. 2002. Deontic logic and computer-supported computer ethics, Moor H., Bynum T. W. (eds.). *Cyberphilosophy. The Intersection of Philosophy and Computing*: 280–289. Oxford: Blackwell Publishing Ltd.

Kantas Imanuelis. 1980. *Dorovės metafizikos pagrindai*. Iš vokiečių k. vertė K. Rickevičiūtė. Vilnius: Mintis.

Kantas Imanuelis. 1982. *Grynojo proto kritika*. Iš vokiečių k. vertė R. Plečkaitis. Vilnius: Mintis.

Kantas Imanuelis. 1987. *Praktinio proto kritika*. Iš vokiečių k. vertė R. Plečkaitis. Vilnius: Mintis.

Miniotaitė Gražina. 1988. Kanto etinės idėjos šiuolaikinėje moralės filosofijoje. Degutis A. (sud.). *I. Kanto filosofijos profiliai*: 141–159. Vilnius: Mintis.

Miniotaitė Gražina. 2006. Moralinės tolerancijos samprata Kanto praktinėje filosofijoje, *Logos-Vilnius* 48: 14–23.

Schlicht Tobias. 2022. Minds, Brains, and Deep Learning: The Development of Cognitive Science Through the Lens of Kant's Approach to Cognition. Kim H., Schönecker D. (eds.). *Kant and Artificial Intelligence*: 3–38. Leck: CPI books GmbH.

Schmidt Elke Elisabeth. 2022. Kant on Trolleys and Autonomous Driving. Kim H., Schönecker D. (eds.). *Kant and Artificial Intelligence*: 189–221. Leck: CPI books GmbH.

Schönecker Dieter. 2022. Kant's Argument from Moral Feelings: Why Practical Reason Cannot Be Artificial. Kim H., Schönecker D. (eds.). *Kant and Artificial Intelligence*: 169–188. Leck: CPI books GmbH.

Stiegler Bernard, Neyrat Frédéric. 2012. Interview: From Libidinal Economy to the Ecology of the Spirit, *Parrhesia: A Journal of Critical Philosophy* 14: 10, cit. iš: Vidauskytė Lina. 2021. Dirbtinis intelektas: visuomenės infantilizacija ir bejėgiškumas, *Logos-Vilnius* 109: 76.

Vidauskytė Lina. 2021. Dirbtinis intelektas: visuomenės infantilizacija ir bejėgiškumas, *Logos-Vilnius* 109: 71–77.

Wright Ava Thomas. 2022. Rightful Machines. Kim H., Schönecker D. (eds.). *Kant and Artificial Intelligence*: 223–237. Leck: CPI books GmbH.

Горохов В. Г. (сост.), перевод с немецкого и английского Ц. Г. Арзаканяна, В. Г. Горохова, Ю. Б. Тупталова, А. О. Сейдалиной. *Философия техники в ФРГ*: 354–363. Москва: Прогресс.

Инхетвен Рюдигер. 1989. Эвристика и аналогии в технических науках, Арзаканян Ц. Г.,

Кант Иммануил. 1965. *Сочинения в шести томах* 4 (2). Асмус В. Ф., Гулыга А. В., Ойзерман Т. И. (ред.). Москва: Мысль: 360–361, cit. iš: Miniotaitė Gražina. 2006. Moralinės tolerancijos samprata Kanto praktinėje filosofijoje, *Logos-Vilnius* 48: 22.

Nuorodos

[1] Tiesa, šventumo dėsnis dirbtinio intelekto (kaip sukurtosios būtybės kūrinio) terpėje negalėtų galioti. Jei žmogus neturi šventos valios (Kantas 1987: 103), tai kompiuteris apskritai neturi valios (Schönecker 2022: 184).

[2] Episteminė (gr. *ἐπιστήμη* – pažinimas, žinios) logika – modalinės logikos atšaka, susijusi su žiniomis ir tikėjimu. Modalinė (lot. *modus* – būdas, matas) logika nagrinėja teiginių kombinacijas atsižvelgdama į jų modalumą: būtinybę, galimybę ir negalimybę (Hoven, Lokhorst 2002: 284–285).

[3] Veiksmo logika – modalinės logikos šaka, semantiškai artima pačiai modalinei logikai ir temporalinei (laikinumo, dinaminei) logikai, tyrinėjančiai galimus pasaulius su tikrais jų tarpusavio ryšiais. Pritaikyta dinaminei logikai kompiuterijos moksle. Pagal operatorių (*pasirūpinti, kad...*) veiksmo logika dar vadinama priežiūros arba sekimo logika (Hoven, Lokhorst 2002: 285–286).

[4] Hibridinės (mišrios) loginės sistemos sujungia kelių sričių operatorius. Deontika pritaikoma veiksmams, susiejami veiksmo ir episteminiai operatoriai, ir visa pritaikoma kompiuterinei etikai. Loginės sistemos sudaro kompiuterinio aprūpinimo pagrindą (Hoven, Lokhorst 2002: 286–287).

[5] Anot Bernardo Steiglerio, „*τέχνη* [...] sudaro libido“ (Stiegler 2012: 10, cit. iš: Vidauskytė 2021: 76, vertimas mano. – *R. B.*).

[6] Euristika (gr. *εὑρισκειν* – rasti) – atradimo menas, turintis senas tradicijas: sušukęs „eureka“, Archimedas įvardijo atradimo idėją. Tai mokslas, tiriantis kūrybinę veiklą, siekiantis suprasti bendruosius kūrybos dėsningumus, išskirti įvairius kūrybos pjūvius, kurie padėtų jai tiksliau ir nuosekliau apibūdinti įvairius kūrybos aktus. Daugiau žr.: Bartkus 2019: 70, 75.

[7] Analogija (gr. *αναλόγια* – atitikimas) – skirtingų daiktų, reiškinių, funkcijų, sąvokų panašumas; samprotavimas (pažinimo ir kalbėjimo būdas), kai pagal vienus objektų panašumo požymius sprendžiama apie kitus arba kai koks nors dalykas nusakomas tik iš santykio su kitu (jau pažintu) dalyku.