Vilnius University Press

Contents lists available at Vilnius University Press

Politologija
2020/1, vol. 97, pp. 123–129

ISSN 1392-1681 eISSN 2424-6034
DOI: https://doi.org/10.15388/Polit.2020.97.5

Recenzijos / Reviews

45 požiūriai į dirbtinio intelekto grėsmes ir galimybes – ko tikėtis?

Ieva Skurdauskaitė

Vilniaus universiteto Tarptautinių santykių ir politikos mokslų instituto doktorantė
El. paštas: ieva.skurdauskaite@tspmi.vu.lt

Panašiu metu išleistos dviejų autoritetingų technologijų temos mąstytojų redaguotos knygos pagaliau susistemintai atsveria kasdienes, dažnai nepamatuotai optimistiškas arba, priešingai, hiperbolizuotai gąsdinančias publicistines antraštes apie dirbtinio intelekto vystymą[1]. Tiek Martinas Fordas, tiek Johnas Brockmanas ne vienerius metus nagrinėja technologijų temą[2] ir už tai jau yra susilaukę nemenko dėmesio[3]. Šįkart jie knygose „Architects of Intelligence: The Truth

1 Ford Martin, *Architects of Intelligence: The Truth About AI from the People Building It*, Packt Publishing: UK, 2018 m. lapkričio mėn.
Brockman John, *Possible Minds: Twenty-Five Ways of Looking at AI*, New York: Penguin Press, 2019 m. vasario mėn.

2 Pavyzdžiui, tarp naujausių darbų yra Ford Martin, *Rise of the Robots: Technology and the Threat of a Jobless Future*, Basic Books: New York, 2015, ir Brockman John, sud., *What to Think About Machines That Think: Today's Leading Thinkers on the Age of Machine Intelligence*, Harper Perennial: New York, 2015.

3 M. Fordas yra išleidęs 3 knygas technologijų tema, o 2015 m. jo knyga „Rise of the Robots" laimėjo „*Financial Times* and McKinsey Business Book of the Year Award". J. Brockmanas yra literatūros agentas, kuris specializuojasi mokslinėje literatūroje, yra įkūręs organizaciją „Edge Foundation", vienijančią plataus profilio mokslininkus, išleido daugiau kaip 25 mokslinius ir intelektualius klausimus keliančias knygas.

about AI from the People Building it“ ir „Possible Minds: Twenty-Five Ways of Looking at AI“ savo kėdę užleidžia 45[4] šiuo metu populiariausiems ir neabejotinai reikšmingiausiems dirbtinio intelekto kūrėjams ir mąstytojams, o patys atlieka puikiai šį lauką išmanančio klausėjo ir redaktoriaus vaidmenis.

Pasirinkti mąstytojai, abiejų knygų tematika ir forma iš esmės ir nulemia tai, kodėl, kalbant apie anksčiau išleistą knygą, norisi su ja sieti ir antrąją. Dirbtinio intelekto grėsmių ir galimybių tema vis dar yra karšta ir itin kompleksiška, apimanti ne vieną žmogaus gyvenimo sferą, todėl knygose pateiktos mintys yra lyg viena dirbtinio intelekto vystymo stotelė – apibendrinimas to, kas jau padaryta, aptarimas to, kas vyksta dabar, ir prognozės ateičiai. Šiandien niekas negali tiksliai nusakyti dirbtinio intelekto vystymo padarinių, tačiau M. Fordas ir J. Brockmanas pasistengė surinkti tuos, kurie dėl savo profesijos, darbų ir patirties gali pasidalyti svariausiomis įžvalgomis. Tai ir populiariausio pasaulyje dirbtinio intelekto vadovėlio autorius Stuartas Russellas, savo darbais apie egzistencines grėsmes ir superintelektą išgarsėjęs filosofas Nickas Bostromas, „DeepMind“ įkūrėjas Demis Hassabis, Stanfordo universiteto instituto „Human-Centered Artificial Intelligence“ vadovė Fei-Fei Li, kompiuterių mokslininkas ir filosofas Judea Pearlas, fizikas ir kosmologas Maxas Tegmarkas, robotistas Rodney Brooksas, kompiuterių mokslininkė Barbara Grosz ir kiti. Dauguma mąstytojų pabrėžia, kad dirbtinio intelekto klausimai nėra vien techniniai ir technologiniai, tačiau daugiausia kalbinami tiksliųjų mokslų atstovai. Nenuneigiant jų įžvalgų, neapleidžia mintis, jog, įtraukus daugiau socialinių, teisės ir humanitarinių mokslų atstovų, diskusija įgautų naujų įžvalgų ir gylio. Ypač kai kalbama dirbtinio intelekto reguliavimo, etikos, galimų tarptautinių santykių varžybų temomis. Šį trūkumą iš dalies kompensuoja į knygą „Possible Minds“ įtraukti mąstytojai, dar kartą primindami, kad dirbtinio intelekto vystymo ir pritaikymo klausimai yra klausimai apie žmogų ir jo galimybes, suvokimą ir socialinį pasaulį.

[4] Iš tiesų 48, bet 3 autoriai pasirodo abiejose knygose: Stuart Russell, Judea Pearl ir Rodney Brooks.

Kita vertus, knygos forma pačius autorius šiek tiek apriboja detaliau aptarti dirbtinio intelekto nulemtas šiuolaikinio socialinio pasaulio aktualijas. M. Fordas savo knygai pasirinko interviu žanrą, taip palikdamas daug erdvės kalbintų mąstytojų reakcijoms ir pasakojimams. Čia netrūksta klausimų apie asmenines patirtis, tačiau tai nenukreipia žvilgsnio nuo diskusijos dirbtinio intelekto tema, netgi priešingai – kiek įmanoma, leidžiama skaitytojui įsijausti į kūrėją ir jo kontekstą. Be to, dirbtinio intelekto tema yra techniškai kompleksiška, jos žodynas suprantamas tik šios srities atstovams, tad betarpiško interviu metu išlaikomas balansas tarp mokslinių žinių ir nesudėtingos terminijos pasakojimo, priartinančio šias idėjas prie platesnio skaitytojų rato.

J. Brockmano redaguota knyga yra kiek oficialesnė, nes susideda iš trumpų 25 mąstytojų esė pasirinktu dirbtinio intelekto tyrimų ir taikymo klausimu. Tiesa, knygos pagrindą sudaro matematiko, kibernetikos pradininko ir šiuolaikinius dirbtinio intelekto tyrimus įkvėpusio mokslininko Roberto Wienerio darbai („Cybernetics“, o vėliau ir „The Human Use of Human Beings“), todėl pateiktas esė vertėtų priimti kaip plačias interpretacijas, pasitelkiant asmeninius tyrimus ir interesų sritis. Dėl to knygos pavadinimas, kuriame nėra nuorodų į R. Wienerio idėjų rinkinį, skaitytojus šiek tiek klaidina ir temos prasme apriboja pačius autorius. Kita vertus, tai suteikia daugiau objektyvumo, diskusijas „nuleidžia ant žemės“ ir dirbtinio intelekto vystymą leidžia apžvelgti retrospektyviai. Sujungę abiejų knygų pajėgas, gauname įvairiapusę dirbtinio intelekto klausimų apžvalgą, aktualią tiek šioje srityje tiesiogiai dirbantiems profesionalams, tiek šiuolaikinių technologinių diskusijų pulsą bandantiems užčiuopti entuziastams.

Dirbtinio intelekto galimybės ir grėsmės

M. Fordas interviu struktūruoja pagal tris pagrindines temas: 1) dirbtinio intelekto ir robotikos įtaką darbo rinkai ir ekonomikai; 2) bendrąjį dirbtinį intelektą ir 3) dirbtinio intelekto progreso grėsmes. Autorius jau yra parašęs knygą apie robotiką ir ekonomiką, tad nenuostabu, kad

pirmajai temai skirta nemažai vietos. Visi kalbinti pašnekovai pripažįsta, kad dirbtinis intelektas yra daugelio paskirčių technologija ir neabejotinai keis (ir keičia) darbo rinką. Kai kurie dirbtinio intelekto poveikį lygina su elektra, tačiau visi iš esmės sutaria, kad tai, kiek dirbtinis intelektas padarys žalos ir naudos ekonomikai, priklausys ne nuo pačios technologijos savaime, o nuo socialinės sistemos. Tai, visų pirma, reiškia adekvačias investicijas į žmonių edukaciją ir persikvalifikavimą, dėmesio sutelkimą į tas sritis, kuriose šiuo metu dirbtinis intelektas gali duoti daugiausia naudos – žemės ūkį ir mediciną. Technologijų įtaka ekonomikai nėra nauja tendencija, todėl ir šįkart nevertėtų priešinti technologijų ir žmonių, nes profesijos, kurių išnykimui gresia didžiausias pavojus, yra karjeros kopėčių apačioje. Be to, kuriamos darbo vietos kompleksiškesnių įgūdžių srityje, technologiniai sprendimai žmones išlaisvina iš rutininio darbo, didina darbo efektyvumą ir laiko kokybę, sprendžiama pensinio ir darbinio amžiaus žmonių santykio didėjimo problema. Nuomonės daugiausia išsiskiria, kai klausimai pasisuka į vieną iš siūlomų sprendimų – universalias bazines pajamas. Vieni mąstytojai tvirtina, kad nėra kitos alternatyvos, lieka tik klausimas, kaip ir kada prie to prieisime, kiti, priešingai, tikina, kad tai nėra aktualiausia tema, nes dar negalime suvokti realaus dirbtinio intelekto įtakos darbo rinkai masto. Jis padėtų spręsti atlyginimų lygybės klausimus, tačiau tada būtų nuvertinama darbo prasmė, kuri taip pat apima žmogaus orumą, savigarbą, bendruomeniškumą. Anksčiau rutininį darbą atlikę žmonės turėtų paskatą mokytis, tačiau vis tiek liktų techninio darbo, kurį reikėtų atlikti žmogui.

Faktiškai visų M. Fordo ir J. Brockmano kalbintų mąstytojų pamatuotas ir ramus požiūris į bendrąjį dirbtinį intelektą ir apskritai naujausias technologijas išskiria juos iš garsiai rėkiančių antraštininkų ir gąsdintojų. Remiantis M. Tegmarko bendrojo dirbtinio intelekto „judėjimų“ klasifikacija – skaitmeniniai utopistai, technoskeptikai ir naudingo dirbtinio intelekto judėjimas[5] – aptariamus mąstytojus galima priskirti paskutinei grupei. Šiuo atveju manoma, kad abejonės

[5] Tegmark Max, *Life 3.0: Being Human in the Age of Artificial Intelligence*, UK, Penguin Books, 2018, p. 31.

yra naudingos, nes dirbtinio intelekto saugumo tyrimai ir diskusijos padidina teigiamų pasekmių tikimybę. Nekalbama radikaliomis kategorijomis (žmonės *vs* technologijos), pripažįstama, kad technologijų įsiliejimas į mūsų gyvenimą yra neišvengiamas, bet – gera žinia – reguliuojamas pačių žmonių. Iš esmės remiamasi ir S. Russello į žmogų orientuoto dirbtinio intelekto idėja – intelektualios technologijos yra naudingos tada, kai jų veiksmai atitinka žmonių, o ne jų pačių tikslus[6]. Žmogaus įsitraukimo principas (angl. *human-in-the-loop*) dominuoja bendrojo dirbtinio intelekto apmąstymuose ir iš dalies pagrindžia tai, kodėl mąstytojai nepriešina žmonių ir technologijų. Grėsmę kelia ne intelektas savaime, o jo autonomija, todėl, esant visuotiniam susitarimui dėl žmogiškosios kontrolės masto, įmanoma pasiekti didžiausią dirbtinio intelekto naudą žmonijai.

Nors dauguma mąstytojų į dirbtinio intelekto vystymą žiūri pamatuotai, realiame pasaulyje nėra lengva priimti visuotinius susitarimus, ypač kai kalba pakrypsta apie svariai tarptautinę galią didinančias technologijas. Pagrindinės įvardijamos grėsmės apima tarptautines varžybas dirbtinio intelekto srityje, autonominius ginklus, kuriuos siūloma vertinti kaip masinio naikinimo ginklus, taip pat rinkodarinę manipuliaciją, pavyzdžiui, rinkėjais ir balsais. Taip pat užkabinamas Vakarų pasaulyje didelio susidomėjimo ir kritikos sulaukęs pagal rasę, lytį ar lytinę orientaciją diskriminuojančių algoritmų klausimas, žmogaus ir dirbtinio intelekto vertybių suderinamumo problema. Pirmas žingsnis sprendimo link – efektyvus teisinis reguliavimas, kuris kol kas gerokai atsilieka nuo sparčiai besivystančių technologijų. Tačiau šios grėsmės apima kur kas daugiau nei teisę – jos nukreipia į etines ir moralines dilemas. Teisiniai reguliavimai reikalingi ne technologijoms, o žmonėms: kaip efektyviai naudoti dirbtinio intelekto technologijas? Kokius žmogaus ribotumus gali papildyti technologijos? Kur yra pasitikėjimo technologijomis riba? Kiek atsakomybės joms perleisti, ypač susidūrus su gyvybės ir mirties klausimu karo zonoje? Atsakymai į šiuos ir daugelį kitų klausimų formuos tolesnį

6 Russell Stuart, *Human Compatible AI and the Problem of Control,* UK, Penguin Random House, 2019, p. 11.

mūsų suvokimą apie žmogaus orumą, jo galimybes, tapatybę ir gyvenimo vertę. Knygose kalbinami daugiausia tiksliųjų mokslų atstovai, todėl šie klausimai yra tik iškeliami, bet ne tyrinėjami. Kita vertus, jie nurodo šiuolaikinę dirbtinio intelekto etikos kryptį.

J. Brockmano knygoje detaliau ar abstrakčiau aptariamos panašios temos, tačiau dėl platesnės autorių profesinės įvairovės įtraukiama daugiau požiūrių. Pavyzdžiui, populiariojoje žiniasklaidoje ir pasisakymuose galima išgirsti teiginį, kad dirbtinis intelektas niekada nepakeis žmogaus kūrybiškumo ir tai yra jo unikalusis bruožas. Meno kuratorius, kritikas ir meno istorikas Hansas Ulrichas Obristas aprašo su dirbtiniu intelektu dirbančių menininkų mintis ir sutinka, kad ši technologija suteikia naujų galimybių kūrybos srityje, tačiau neneigia, kad mes labai daug tikimės iš dirbtinio intelekto ir jį suabsoliutiname, todėl būtų pravartu jo nesieti tik su teigiamais rezultatais[7]. Psichologijos profesorė Alison Gopnik savo ruožtu siūlo atkreipti dėmesį į vaikus – tai, kaip mokosi keturmečiai, gali pasufleruoti, kaip toliau vystyti dirbtinį intelektą. Dar negalime tiksliai nustatyti, iš kur kyla kūrybinis vaikų mąstymas, kita vertus, matome, kad, priešingai nei dirbtinis intelektas, jie yra aktyvūs informacijos priėmėjai, pasaulį pažįsta per žaidimus ir smalsumą, yra kultūriškai angažuoti ir socialūs[8]. Nors šiuo metu tokius kriterijus suteikti dirbtiniam intelektui dar neįmanoma, tai nurodo bendrojo dirbtinio intelekto kūrimo sąlygas.

Dirbtinio intelekto ir žmogaus santykis

Iš pateiktų požiūrių susidaro nuomonė, kad: 1) technologija yra tai, ką žmogus iš jos padaro; 2) dirbtinio intelekto vystymo logika ateina iš žmogaus pažinimo. Nieko neįtikėtinai inovatyvaus ir radikalaus nepasakyta, tačiau, iš pirmo žvilgsnio, techniškoje temoje dar kartą atkreiptas dėmesys į socialinį žmogaus gyvenimą.

Dirbtinio intelekto vystymo iššūkiai ir keliamos grėsmės kelia nemažai klausimų apie mus pačius – kaip mes mąstome, bendrauja-

7 Obrist Hans Ulrich, „Making the Invisible Visible: Art Meets AI“, *Possible Minds*, p. 206–218.

8 Gopnik Alison, „AIs Versus Four-Year-Olds“, *Possible Minds,* p. 229–230.

me, mokomės, jaučiame ir kuriame. Jei norime sukurti žmogaus lygį siekiantį dirbtinį intelektą, pirma turime sau atsakyti į klausimus: kas apibrėžia žmogų ir jo tikslus? Akademiniame pasaulyje vis aktyviau kyla būtent šios krypties diskusijos, per pastaruosius keletą metų prie garsiausių pasaulio universitetų įsikūrė į žmogų arba jam kylančias grėsmes orientuoti specializuoti tyrimų institutai (pavyzdžiui, Oksfordo universiteto „Future of Humanity Institute“ ar Stanfordo universiteto institutas „Human-Centered Artificial Intelligence“). Juose tiriamas žmogaus ir technologijos santykis, technologijų naudojimo etika, keliami egzistencinės grėsmės, naujų teisinių normų kūrimo, diskriminacijos, pasitikėjimo technologijomis ir atsakomybės klausimai. Apskritai, technologijų ir žmogaus santykio klausimas nėra dirbtinio intelekto fenomenas, tik iškyla naujai ir žymi suaktyvėjusias žmogaus etikos diskusijas ir paradoksus. Viena vertus, primenamas žmogaus unikalumas – sunkiausiai atkartojamas jo mąstymas ir moralė. Kita vertus, dirbtinis intelektas šiuo metu vis dar yra šališkas ir diskriminuojantis, nes mokosi iš žmogaus teikiamos informacijos.

Naujos teisinės normos, žmogaus mokymasis ir technologijų etika yra tos temos, kurias pravartu sekti artimiausiu metu ir kurios atlieps tolesnę dirbtinio intelekto tyrimų eigą. Jos taip pat nurodys, ar, kaip teigia Bryanas Johnsonas, dirbtinis intelektas iš tiesų yra geriausias mums nutikęs dalykas po pjaustytos duonos[9].

Literatūra

Brockman John, sud., *Possible Minds: Twenty-Five Ways of Looking at AI,* New York: Penguin Press, 2019.

Brockman John, sud., *What to Think About Machines That Think: Today's Leading Thinkers on the Age of Machine Intelligence,* New York: Harper Perennial, 2015.

Ford Martin, *Architects of Intelligence: The Truth About AI from the People Building It*. UK, Packt Publishing, 2018.

Ford Martin, *Rise of the Robots: Technology and the Threat of a Jobless Future*, New York: Basic Books, 2015.

Russell Stuart, *Human Compatible AI and the Problem of Control,* UK, Penguin Random House, 2019.

Tegmark Max, *Life 3.0: Being Human in the Age of Artificial Intelligence,* UK, Penguin Books, 2018.

9 Interviu su B. Johnson. Kn. *Architects of Intelligence*, p. 515.