



Briuselis, 2019 04 08
COM(2019) 168 final

**KOMISIJOS KOMUNIKATAS EUROPOS PARLAMENTUI, TARYBAI, EUROPOS
EKONOMIKOS IR SOCIALINIŲ REIKALŲ KOMITETUI IR REGIONŲ
KOMITETUI**

Pasitikėjimo į žmogų orientuotu dirbtiniu intelektu didinimas

1. ĮVADAS. EUROPOS DIRBTINIO INTELEKTO STRATEGIJA

Dirbtinis intelektas gali atnešti daug gero mūsų pasauliui: jis gali padėti gerinti sveikatos priežiūrą, mažinti energijos suvartojimą, didinti automobilių saugumą ir suteikti ūkininkams galimybę efektyviau naudoti vandenį ir gamtos išteklius. Dirbtinis intelektas gali būti naudojamas aplinkos ir klimato kaitai prognozuoti ir finansinės rizikos valdymui gerinti. Be to, jis suteikia priemonių gaminti prie mūsų poreikių pritaikytus produktus taip, kad susidarytų mažiau atliekų. Dirbtinis intelektas taip pat gali padėti nustatyti sukčiavimo atvejus ir grėsmes kibernetiniam saugumui ir sudaro teisėsaugos tarnyboms sąlygas efektyviau kovoti su nusikalstamumu.

Dirbtinis intelektas gali būti naudingas visai visuomenei ir ekonomikai. Tai strateginė technologija, kuri šiuo metu sparčiai plėtojama ir naudojama visame pasaulyje. Tačiau dirbtinis intelektas kelia ir naujų su darbo rinkos ateitimi susijusių uždavinių, taip pat teisinių ir etinių klausimų.

Kad šie uždaviniai būtų sprendžiami ir kad būtų kuo geriau išnaudotos dirbtinio intelekto teikiamos galimybės, Komisija 2018 m. balandžio mėn. paskelbė Europos strategiją¹. Strategijoje nustatoma, kad plėtojant dirbtinį intelektą svarbiausia – žmogus, t. y. **dirbtinis intelektas turi būti orientuotas į žmogų.** Pasiūlytos trys veiklos kryptys: didinti technologinį ir pramoninį ES pajėgumą ir skatinti diegti dirbtinį intelektą visose ekonomikos srityse, pasirengti socialiniams ir ekonominiams pokyčiams ir užtikrinti tinkamą etikos ir teisinę sistemą.

Siekdama įgyvendinti dirbtinio intelekto strategiją, **Komisija kartu su valstybėmis narėmis parengė Suderintą dirbtinio intelekto planą².** Šio 2018 m. gruodžio mėn. pristatyto dokumento tikslas – užtikrinti sinergiją, sutelkti duomenų, kurie yra daugelio dirbtinio intelekto prietaikų išteklius, ir didinti bendras investicijas. Siekiama skatinti tarpvalstybinį bendradarbiavimą ir sutelkti visus subjektus, kad per artimiausią dešimtmetį viešosios ir privačiosios investicijos pasiektų **bent 20 mlrd. EUR** per metus³. Komisija padvigubino programoje „Horizontas 2020“ numatytas investicijas į dirbtinį intelektą ir planuoja iš programos „Europos horizontas“ ir Skaitmeninės Europos programos lėšų kasmet investuoti 1 mlrd. EUR – visų pirma remti bendras duomenų erdves sveikatos, transporto ir gamybos sektoriuose, didelius eksperimentinius objektus, pavyzdžiui, pažangiąsias liginines bei automatizuotų automobilių infrastruktūrą, ir strateginę mokslinių tyrimų darbotvarkę.

Siekdama įgyvendinti tokią bendrą strateginę mokslinių tyrimų, inovacijų ir diegimo veiksmų darbotvarkę, Komisija suintensyvino **dialogą su visais atitinkamais** pramonei, mokslo tiriamosioms institucijoms ir valdžios institucijoms atstovaujančiais **suinteresuotaisiais subjektais.** Labai svarbi bus ir Skaitmeninės Europos programa – pagal ją sukurti skaitmeninių inovacijų centrai, sustiprinta bandymų ir eksperimentavimo infrastruktūra, sukurtos duomenų erdvės ir mokymo programos padės užtikrinti, kad dirbtinis intelektas būtų prieinamas ir visų valstybių narių mažosioms ir vidutinėms įmonėms.

Europa garsėja saugiais ir kokybiškais produktais, todėl šia reputacija pagrįstas etiškas jos požiūris į dirbtinį intelektą didina piliečių pasitikėjimą skaitmenine plėtra ir turėtų padėti užtikrinti Europos dirbtinio intelekto įmonių konkurencinį pranašumą. Šio komunikato tikslas

¹ COM(2018) 237.

² COM(2018) 795.

³ Siekdama padėti įgyvendinti šį tikslą, Komisija pasiūlė Sąjungai būsimo 2021–2027 m. programavimo laikotarpiu investicijoms į dirbtinį intelektą finansuoti kasmet skirti bent 1 mlrd. EUR iš programos „Europos horizontas“ ir Skaitmeninės Europos programos lėšų.

– pradėti nuodugnaus bandymo etapą, kuriuo kuo įvairesni suinteresuotieji subjektai bandytų praktiškai įgyvendinti dirbtinio intelekto plėtojimo ir naudojimo etikos rekomendacijas.

2. PASITIKĖJIMO Į ŽMOGŲ ORIENTUOTU DIRBTINIŲ INTELEKTŲ DIDINIMAS

Europos dirbtinio intelekto strategijoje ir koordinuotame plane aiškiai nurodoma, kad **pasitikėjimas – būtina sąlyga siekiant užtikrinti, kad dirbtinis intelektas būtų orientuotas į žmogų**: dirbtinis intelektas nėra savitikslis, jis – žmogaus reikmes turinti tenkinti priemonė, kurios galutinis tikslas – didinti žmonių gerovę. Siekiant šio tikslo **turėtų būti užtikrintas dirbtinio intelekto patikimumas**. Plėtojant dirbtinį intelektą turi būti visapusiškai vadovaujamosi vertybėmis, kuriomis grindžiamos mūsų visuomenės.

Sąjunga grindžiama **šiomis vertybėmis: pagarba žmogaus orumui, laisve, demokratija, lygybe, teisine valstybe ir pagarba žmogaus teisėms**, įskaitant mažumoms priklausančių asmenų teises⁴. Šios vertybės yra bendros visų valstybių narių visuomenėms, kuriose vyrauja pliuralizmas, nediskriminavimas, tolerancija, teisingumas, solidarumas ir lygybė. Be to, visoje Europos Sąjungoje žmonėms suteiktos asmeninės, pilietinės, politinės, ekonominės ir socialinės teisės yra išdėstytos viename dokumente – **ES pagrindinių teisių chartijoje**.

Europos Sąjungoje sukurta **tvirta reglamentavimo sistema**, kuri bus pasaulinis pavyzdys plėtojant į žmogų orientuotą dirbtinį intelektą. Bendruoju duomenų apsaugos reglamentu užtikrinamas aukštas asmens duomenų apsaugos lygis; jame reikalaujama taikyti priemones, kuriomis būtų užtikrinta pritaikytoji ir standartizuotoji duomenų apsauga⁵. Laisvo ne asmens duomenų judėjimo reglamentu pašalinamos laisvo ne asmens duomenų judėjimo kliūtys ir užtikrinamas visų kategorijų duomenų tvarkymas visoje Europoje. Neseniai priimtas Kibernetinio saugumo aktas padės didinti pasitikėjimą interneto erdve – to paties tikslo siekiama ir siūlomu E. privatumo reglamentu⁶.

Nepaisant to, dirbtinis intelektas kelia naujų uždavinių, nes sudaro sąlygas kompiuteriams mokytis ir priimti bei įgyvendinti sprendimus be žmogaus įsikišimo. Netrukus tai taps standartine daugelio rūšių prekių ir paslaugų – nuo išmaniųjų telefonų iki automatizuotų automobilių, robotų ir internetinių prietaikų – funkcija. Tačiau algoritmų priimti sprendimai gali būti pagrįsti neišsamiais, taigi nepatikimais duomenimis, jie gali būti neteisėtai paveikti per kibernetinius išpuolius, būti neobjektyvūs ar tiesiog klaidingi. Todėl, taikant šią besivystančią technologiją be jokios atodairos, kultų problemų ir piliečiai nenorėtų jos pripažinti ir naudoti.

Vietoj to dirbtinio intelekto technologija turėtų būti plėtojama daugiausia dėmesio skiriant žmogui ir taip pelnant visuomenės pasitikėjimą. Tai reiškia, kad dirbtinio intelekto prietaikos turėtų atitikti ne tik teisės aktus, bet ir etikos principus. Be to, turėtų būti užtikrinta, kad jomis naudojantis būtų išvengta netyčinės žalos. Visais dirbtinio intelekto plėtojimo etapais turėtų būti atsižvelgiama į lyčių, rasų ar etninės kilmės, religijų ar tikėjimų, negalių ir amžiaus

⁴ ES taip pat yra JT neįgaliųjų teisių konvencijos susitariančioji šalis.

⁵ Reglamentas (ES) 2016/679. Bendruoju duomenų apsaugos reglamentu užtikrinamas laisvas asmens duomenų judėjimas Sąjungoje. Jame pateiktos nuostatos dėl vien automatizuotu duomenų tvarkymu pagrįsto sprendimų priėmimo, be kita ko, profiliavimo. Atitinkami asmenys turi teisę būti informuoti apie automatizuoto sprendimų priėmimo priemonių taikymą ir gauti prasmingą informaciją apie loginį automatizuoto sprendimų priėmimo pagrindą, taip pat apie tokio duomenų tvarkymo reikšmę ir numatomas pasekmes asmeniui. Tokiais atvejais jie taip pat turi teisę reikalauti žmogaus įsikišimo, pareikšti savo požiūrį ir užginčyti sprendimą.

⁶ COM(2017) 10.

įvairovę. AI prietaikos turėtų suteikti daugiau galių piliečiams ir paisyti jų pagrindinių teisių. Jų tikslas turėtų būti ne padaryti žmonių gebėjimus nereikalingus, o juos stiprinti. Be to, turėtų būti užtikrinta, kad šiomis prietaikomis galėtų naudotis neįgalieji.

Todėl būtina parengti esama reglamentavimo sistema pagrįstas **etikos gaires**, kuriomis vidaus rinkoje turėtų vadovautis dirbtinio intelekto kūrėjai, tiekėjai ir naudotojai, kad visose valstybėse narėse būtų užtikrintos vienodos etinės sąlygos. Šiuo tikslu Komisija įsteigė iš įvairių suinteresuotųjų subjektų sudarytą **Aukšto lygio ekspertų grupę dirbtinio intelekto klausimais**⁷ ir pavedė jai parengti dirbtinio intelekto etikos gaires ir rekomendacijų dėl platesnės dirbtinio intelekto politikos. Tuo pat metu buvo sukurtas **Europos dirbtinio intelekto aljansas**⁸ – atvira įvairių suinteresuotųjų subjektų platforma, prie kurios yra prisijungę daugiau kaip 2 700 narių ir kurios tikslas – suteikti daugiau Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais darbui svarbios informacijos.

2018 m. gruodžio mėn. Aukšto lygio ekspertų grupė dirbtinio intelekto klausimais paskelbė pirmąjį etikos gairių projektą. Po **konsultacijų su suinteresuotaisiais subjektais**⁹ ir susitikimų su **valstybių narių atstovais**¹⁰ Dirbtinio intelekto ekspertų grupė 2019 m. kovo mėn. pateikė Komisijai peržiūrėtą dokumentą. Iš iki šiol gautų atsiliepimų matyti, kad suinteresuotieji subjektai apskritai palankiai vertina praktinį gairių pobūdį ir jose dirbtinio intelekto kūrėjams, tiekėjams ir naudotojams pateikiamas konkrečias rekomendacijas, kaip užtikrinti patikimumą.

2.1. Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais parengtos gairės dėl patikimo dirbtinio intelekto

Šiame komunikate nurodomos Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais parengtos gairės¹¹ visų pirma grindžiamos Europos mokslo ir naujų technologijų etikos grupės ir Pagrindinių teisių agentūros veikla.

Gairėse nurodoma, kad norint užtikrinti, jog dirbtinis intelektas būtų patikimas, būtini trys elementai: 1) jis turėtų atitikti teisės aktus; 2) jis turėtų atitikti etikos principus; 3) jis turėtų būti patvarus.

Atsižvelgiant į šiuos tris elementus ir į 2 skirsnyje nurodytas Europos vertybes, gairėse nustatomi septyni pagrindiniai reikalavimai, kuriuos turėtų atitikti dirbtinio intelekto prietaikos, kad būtų laikomos patikimomis. Gairėse taip pat pateikiamas vertinimo klausimų sąrašas, padedantis patikrinti, ar šie reikalavimai įvykdyti.

Septyni pagrindiniai reikalavimai yra:

- žmogiškasis veiksnys ir žmogaus atliekama priežiūra;
- techninis patvarumas ir saugumas;

⁷ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

⁸ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

⁹ Per šias konsultacijas gauta atsiliepimų iš 511 organizacijų, asociacijų, įmonių, mokslo tiriamųjų institucijų, asmenų ir kitų subjektų. Gautų atsiliepimų santrauka pateikiama adresu https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf.

¹⁰ Valstybės narės palankiai įvertino ekspertų grupės darbą – 2019 m. vasario 18 d. priimtose Tarybos išvadose, be kita ko, atkreipiamas dėmesys į artėjantį etikos gairių paskelbimą ir nurodoma, kad Taryba remia Komisijos pastangas ES taikomą etišką požiūrį perkelti į pasaulio areną – <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/lt/pdf>.

¹¹ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

- privatumas ir duomenų valdymas;
- skaidrumas;
- įvairovė, nediskriminavimas ir teisingumas;
- visuomenės ir aplinkos gerovė;
- atskaitomybė.

Nors šie reikalavimai skirti visoms įvairiose aplinkose ir sektoriuose naudojamoms dirbtinio intelekto sistemoms, siekiant užtikrinti, kad konkretus jų įgyvendinimas būtų proporcingas, turėtų būti atsižvelgiama į jų taikymo kontekstą ir galimą poveikį. Pavyzdžiui, pasiūlymas skaityti netinkamą knygą yra daug mažiau pavojingas nei klaidingos vėžio diagnozės nustatymas, todėl knygas siūlančiai dirbtinio intelekto prietaikai galėtų būti taikoma ne tokia griežta priežiūra.

Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais parengtos gairės yra neprivalomos ir jomis nenustatoma jokių naujų teisinių įpareigojimų. Kita vertus, daugelis esamų (dažnai su konkrečia naudojimo paskirtimi ar sritimi susietų) Sąjungos teisės nuostatų jau yra susijusios su vienu ar daugiau šių pagrindinių reikalavimų (pavyzdžiai – saugos, asmens duomenų apsaugos, privatumo ar aplinkos apsaugos taisyklės).

Komisija palankiai vertina Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais darbą ir laiko jį vertingu indėliu, padedančiu jai formuoti politiką.

2.2. Pagrindiniai patikimam dirbtiniam intelektui keliami reikalavimai

Komisija pritaria toliau išvardytiems pagrindiniams patikimam dirbtiniam intelektui keliamiems reikalavimams, pagrįstiems Europos vertybėmis. Ji ragina suinteresuotuosius subjektus taikyti šiuos reikalavimus ir išbandyti vertinimo klausimų sąrašą, padedantį juos įgyvendinti praktiškai, kad būtų sukurta pasitikėjimo atmosfera, reikalinga dirbtiniam intelektui sėkmingai plėtoti ir naudoti. Komisija laukia suinteresuotųjų subjektų atsiliepimų, kad galėtų nustatyti, ar šį gairėse pateiktą vertinimo klausimų sąrašą dar reikia tikslinti.

I. Žmogiškasis veiksnys ir žmogaus atliekama priežiūra

Dirbtinio intelekto sistemos turėtų padėti asmenims priimti geresnius, labiau informacija pagrįstus sprendimus, atitinkančius jų tikslus. Jos turėtų stiprinti žmogiškąjį veiksnį bei **pagrindines teises** ir taip sudaryti sąlygas klestinčiai ir teisingai visuomenei, o ne mažinti, apriboti arba netinkama linkme nukreipti žmogaus autonomiją. Svarbiausias sistemos funkcijų tikslas turėtų būti bendra **naudotojo gerovė**.

Žmogaus atliekama priežiūra padeda užtikrinti, kad dirbtinio intelekto sistema nepakenktų žmogaus autonomijai ir nesukeltų kitų neigiamų padarinių. Priklausomai nuo konkrečios dirbtiniu intelektu pagrįstos sistemos ir jos naudojimo srities, turėtų būti taikomos atitinkamo lygio **kontrolės priemonės**, be kita ko, susijusios su dirbtiniu intelektu pagrįstų sistemų pritaikomumu, tikslumu ir paaiškinamumu¹². **Priežiūra** gali būti

¹² Bendrajame duomenų apsaugos reglamente asmenims suteikiama teisė reikalauti, kad jiems nebūtų taikomas tik automatizuotu duomenų tvarkymu grindžiamas sprendimas, kai dėl jo naudotojams kyla teisinių pasekmių arba panašiu būdu daromas didelis poveikis (Bendrojo duomenų apsaugos reglamento 22 straipsnis).

atliekama taikant valdymo mechanizmus, kuriais užtikrinama, kad žmogus dalyvautų procese (angl. *human-in-the-loop*), jį prižiūrėtų (angl. *human-on-the-loop*) arba jį valdytų (angl. *human-in-command*)¹³. Turi būti užtikrinta, kad valdžios institucijos galėtų naudotis turimais priežiūros įgaliojimais pagal savo kompetenciją. Kai kitos sąlygos nesikeičia, kuo mažesnės yra žmogaus galimybės prižiūrėti dirbtinio intelekto sistemą, tuo nuodugniau ji turi būti išbandoma ir tuo griežčiau valdoma.

II. Techninis patvarumas ir saugumas

Kad dirbtinis intelektas būtų patikimas, jo algoritmai turi būti pakankamai saugūs, patikimi ir patvarūs, kad visais dirbtinio intelekto sistemos gyvavimo ciklo etapais susidorotų su klaidomis ar neatitikimais ir kad tinkamai reaguotų į klaidingus rezultatus. Dirbtinio intelekto sistemos turi būti pakankamai **patikimos** ir saugios, kad būtų **atsparios** tiek atviriems išpuoliams, tiek subtilesniems bandymams manipuluoti duomenimis ar pačiais algoritmais. Be to, turi būti numatytas probleminiams atvejams skirtas **atsarginis planas**. Šių sistemų sprendimai turi būti **tikslūs** ar bent atitikti sistemų tikslumo lygį, o jų gautus rezultatus turėtų būti įmanoma **atkurti**.

Be to, į projektuojamas dirbtinio intelekto sistemas turėtų būti integruoti saugos ir saugumo užtikrinimo mechanizmai, **kad jos būtų užtikrintai saugios** visais etapais, itin daug dėmesio skiriant visų susijusių subjektų fiziniam ir psichiniam saugumui. Tai apima nenumatytų sistemos veikimo padarinių ar klaidų mažinimą ir, jei įmanoma, galimybę atlikti grįžtamuosius veiksmus. Turėtų būti nustatyti galimos rizikos, susijusios su dirbtinio intelekto sistemų naudojimu įvairiose srityse, išaiškinimo ir vertinimo procesai.

III. Privatumas ir duomenų valdymas

Visais dirbtinio intelekto sistemos gyvavimo ciklo **etapais** turi būti užtikrintas privatumas ir **duomenų apsauga**. Skaitmeniniai žmogaus elgsenos duomenys gali suteikti dirbtinio intelekto sistemoms galimybę daryti išvadą ne tik apie asmens prioritetus, amžių ir lytį, bet ir apie jo seksualinę orientaciją ir religines ar politines pažiūras. Kad asmenys galėtų pasitikėti duomenų tvarkymu, reikia užtikrinti, kad jie galėtų visiškai kontroliuoti savo duomenis ir kad su jais susiję duomenys nebūtų panaudoti siekiant jiems pakenkti ar juos diskriminuoti.

Turi būti užtikrinta ne tik privatumo ir asmens duomenų apsauga, bet ir aukšta dirbtinio intelekto sistemų kokybė. Didžiausią įtaką dirbtinio intelekto sistemų veiksmingumui turi naudojamų duomenų rinkinių kokybė. Surinkti duomenys gali atspindėti socialiai susiformavusias šališkas tendencijas, juose gali būti netikslumų, klaidų ir apsirikimų. Šį klausimą reikia išspręsti prieš naudojant konkrečius duomenų rinkinius dirbtinio intelekto sistemai mokytis. Be to, turi būti užtikrintas duomenų **vientisumas**. Taikomi procesai ir naudojami duomenų rinkiniai turi būti tikrinami ir dokumentuojami kiekvienu, pavyzdžiui, planavimo, mokymo, bandymo ir diegimo, etapu. Tai taip pat turėtų būti

¹³ Jei žmogus dalyvauja procese (angl. *human-in-the-loop*), nė vienas sprendimas sistemoje nepriimamas be jo įsikišimo, o tai daugeliu atvejų nei įmanoma, nei pageidautina. Jei žmogus prižiūri procesą (angl. *human-on-the-loop*), jis dalyvauja kuriant sistemą ir stebi jos veikimą. Jei žmogus valdo procesą (angl. *human-in-command*), jis prižiūri bendrą dirbtinio intelekto sistemos veikimą (be kita ko, stebi platesnį jos ekonominį, visuomeninį, teisinį ir etinį poveikį) ir sprendžia, kaip ir kada naudoti sistemą konkrečiomis aplinkybėmis. Pavyzdžiui, jis gali nuspręsti nenaudoti sistemos tam tikromis aplinkybėmis, taip pat nustatyti veiksmų laisvės, kurią naudodamasis sistema turi žmogus, lygius arba užtikrinti, kad būtų galima panaikinti sistemos priimtą sprendimą.

taikoma ne savo sukurtoms, o išorėje įsigytoms dirbtinio intelekto sistemoms. Galiausiai turi būti tinkamai valdoma ir kontroliuojama **prieiga** prie duomenų.

IV. Skaidrumas

Turėtų būti užtikrintas dirbtinio intelekto sistemų **atsekamumas**. Svarbu registruoti ir dokumentuoti tiek sistemų priimtus sprendimus, tiek visą tų sprendimų priėmimo procesą (be kita ko, aprašyti duomenų rinkimo ir žymėjimo procesą ir naudojamą algoritmą). Algoritminį sprendimų priėmimo procesą taip pat turėtų būti galima, kiek įmanoma, **paiškinti** atitinkamiems asmenims suprantama kalba. Turėtų būti tęsiami šiuo metu vykdomi moksliniai tyrimai, kuriais siekiama sukurti paaiškinamumo mechanizmus. Be to, turėtų būti paaiškintas mastas, kuriuo dirbtinio intelekto sistema veikia ir formuoja organizacinį sprendimų priėmimo procesą, taip pat sistemos projektavimo sprendimų pasirinkimas ir sistemos diegimo loginis pagrindas (taigi užtikrintas ne tik duomenų ir sistemos, bet ir verslo modelio skaidrumas).

Galiosiausiai svarbu atsižvelgiant į konkretų naudojimo atvejį tinkamu būdu deramai **informuoti** skirtingus suinteresuotuosius subjektus apie dirbtinio intelekto sistemos galimybes ir apribojimus. Be to, dirbtinio intelekto sistemos turėtų būti atpažįstamos, kad naudotojai suvoktų, jog sąveikauja su dirbtinio intelekto sistema, ir žinotų, kas už ją atsakingi.

V. Įvairovė, nediskriminavimas ir teisingumas

Dirbtinio intelekto sistemų naudojami duomenų rinkiniai (kuriais grindžiamas jų mokymasis ar veikimas) per neapdairumą gali atspindėti istoriškai susiformavusias šališkas tendencijas, būti neišsamūs arba netinkamai valdomi. Tokių šališkų tendencijų išlaikymas gali lemti (ne)tiesioginę diskriminaciją. Žalos gali padaryti ir sąmoningas (vartotojų) polinkių išnaudojimas arba nesąžininga konkurencija. Tendencingas gali būti ir dirbtinio intelekto sistemų kūrimo būdas (pavyzdžiui, tai, kaip rašomas algoritmo programavimo kodas). Tokie susirūpinimą keliantys klausimai turėtų būti sprendžiami vos pradėjus kurti dirbtinio intelekto sistemą.

Spręsti šiuos klausimus gali padėti ir **mišrių projektavimo grupių** sudarymas bei mechanizmų, kuriais būtų užtikrinamas įvairių subjektų, visų pirma piliečių, **dalyvavimas** plėtojant dirbtinį intelektą, sukūrimas. Patartina konsultuotis su suinteresuotaisiais subjektais, kuriems dirbtinio intelekto sistema kuriuo nors savo gyvavimo ciklo metu gali daryti tiesioginį arba netiesioginį poveikį. Kuriant dirbtinio intelekto sistemas turėtų būti atsižvelgiama į visą žmonių gebėjimų, įgūdžių ir poreikių spektrą ir taikant universalumo principą užtikrinama, kad jos būtų prieinamos visiems, siekiant sudaryti vienodas sąlygas jomis naudotis ir neįgaliesiems.

VI. Visuomenės ir aplinkos gerovė

Kad dirbtinis intelektas būtų patikimas, turėtų būti atsižvelgta į jo poveikį **aplinkai ir kitiems jausliams gyvams**. Idealiu atveju visiems žmonėms, įskaitant ateities kartas, turėtų būti užtikrinta gyventi tinkama aplinka, kurioje išsaugota biologinė įvairovė. Todėl turėtų būti skatinama užtikrinti, kad dirbtinio intelekto sistemos būtų tvarios ir atitiktų **ekologinės atsakomybės** principą. Tas pats taikytina dirbtinio intelekto sprendimams, susijusiems su pasaulinės svarbos sritimis, pavyzdžiui, JT darnaus vystymosi tikslais.

Be to, reikėtų apgalvoti ne tik, kaip dirbtinio intelekto sistemos paveiks atskirus asmenis, bet ir kokį poveikį jos padarys **visai visuomenei**. Galimybė naudoti dirbtinio intelekto sistemas turėtų būti nuodugniai apsvairstoma, ypač aplinkybėmis, susijusiomis su demokratiniu procesu, be kita ko, nuomonės formavimu, politinių sprendimų priėmimu ar rinkimais. Turėtų būti atsižvelgta ir į dirbtinio intelekto **socialinį poveikį**. Nors dirbtinio intelekto sistemos gali būti naudojamos socialiniams įgūdžiams gerinti, jos taip pat gali prisidėti prie jų prastėjimo.

VII. Atskaitomybė

Turėtų būti nustatyti mechanizmai, kuriais būtų užtikrinta atsakomybė už dirbtinio intelekto sistemas bei jų rezultatus ir su jais susijusi atskaitomybė tiek iki, tiek po jų diegimo. Šiuo atžvilgiu labai svarbi **galimybė atlikti** dirbtinio intelekto sistemų **audita**, nes vidaus ir išorės auditorių atliekamas dirbtinio intelekto sistemų vertinimas ir tokio vertinimo ataskaitų pateikimas labai padeda užtikrinti technologijos patikimumą. Visų pirma turėtų būti užtikrinta galimybė atlikti pagrindinėms teisėms poveikį darančių prietaikų, įskaitant saugumo požiūriu itin svarbias prietaikas, išorės audita.

Turėtų būti nustatomas, vertinamas, dokumentuojamas ir kuo labiau mažinamas **galimas neigiamas** dirbtinio intelekto sistemų **poveikis**. Šį procesą palengvina poveikio vertinimų atlikimas. Šie vertinimai turėtų būti proporcingi dirbtinio intelekto sistemų keliamos rizikos dydžiui. Dažnai neišvengiami skirtingų reikalavimų derinimo **kompromisai** turėtų būti daromi racionaliai ir metodiškai ir turėtų būti paaiškinami. Galiausiai turėtų būti numatyti mechanizmai, kuriais būtų užtikrintos **tinkamos galimybės ginti savo teises**, kai padaromas neteisingas neigiamas poveikis.

2.3. Tolesni veiksmai: nuodugnaus bandymo etapas, kuriame turėtų dalyvauti kuo įvairesni suinteresuotieji subjektai

Pasiektas susitarimas dėl šių pagrindinių dirbtinio intelekto sistemoms keliamų reikalavimų yra pirmas svarbus su gairėmis dėl etiško dirbtinio intelekto susijęs žingsnis. Dabar Komisija užtikrins, kad šias gaires būtų galima išbandyti ir įgyvendinti praktiškai.

Šiuo tikslu Komisija pradės tikslinį bandymo etapą, kurio tikslas – gauti struktūruotų suinteresuotųjų subjektų atsiliepimų. Ypač daug dėmesio bus skiriama aukšto lygio ekspertų grupės parengtam su visais pagrindiniais reikalavimais susietam vertinimo klausimų sąrašui.

Šį darbą sudarys dvi veiksmų sritys: i) gairių bandymo etapas, kuriame dalyvaus dirbtinį intelektą plėtojantys ar naudojantys suinteresuotieji subjektai, be kita ko, viešojo administravimo institucijos, ir ii) tęstinis konsultacijų su suinteresuotaisiais subjektais ir informuotumo didinimo procesas, apimantis visas valstybes nares ir įvairias suinteresuotųjų subjektų grupes, be kita ko, pramonės ir paslaugų sektorius:

- (i) nuo 2019 m. birželio mėn. visi suinteresuotieji subjektai ir individualūs asmenys bus kviečiami išbandyti vertinimo klausimų sąrašą ir pateikti atsiliepimų dėl jo tobulinimo. Be to, Aukšto lygio ekspertų grupė dirbtinio intelekto klausimais kartu su privačiojo ir viešojo sektorių suinteresuotaisiais subjektais atliks nuodugnią peržiūrą, kurios tikslas – gauti išsamesnių atsiliepimų apie tai, kaip būtų galima įgyvendinti gaires įvairiose taikymo srityse. Visi atsiliepimai dėl gairių tinkamumo ir įgyvendinamumo bus įvertinti iki 2019 m. pabaigos;

- (ii) tuo pat metu Komisija organizuos papildomą informavimo veiklą – joje dalyvaujantys Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais atstovai galės pristatyti gaires atitinkamiems suinteresuotiesiems valstybių narių, be kita ko, pramonės ir paslaugų sektorių, subjektams, o šie subjektai turės papildomą galimybę pateikti pastabų dėl dirbtinio intelekto gairių ir įnešti į jas savo indėlį.

Komisija atsižvelgs į Susietojo ir automatizuoto vairavimo etikos ekspertų grupės¹⁴ darbą ir pagrindinių reikalavimų įgyvendinimo klausimais bendradarbiaus su ES finansuojamų dirbtinio intelekto mokslinių tyrimų projektų vykdytojais ir atitinkamomis viešojo ir privačiojo sektorių partnerystės organizacijomis¹⁵. Pavyzdžiui, Komisija, derinama veiksnius su valstybėmis narėmis, remis bendros medicininių vaizdų – iš pradžių susijusių su dažniausių rūšių vėžiu – duomenų bazės kūrimą, kad būtų galima išmokyti algoritmus labai tiksliai diagnozuoti simptomus. Be to, dėl Komisijos ir valstybių narių bendradarbiavimo vis daugėja tarpvalstybinių koridorių, skirtų susietosioms ir automatizuotoms transporto priemonėms bandyti. Įgyvendinant šiuos projektus turėtų būti taikomos ir išbandomos parengtos gairės – to rezultatai bus įtraukti į vertinimo procesą.

Bandymo etapu ir konsultuojantis su suinteresuotaisiais subjektais bus naudingas Europos dirbtinio intelekto aljanso ir AI4EU, užsakomojo dirbtinio intelekto platformos, indėlis. 2019 m. sausio mėn. pradėto AI4EU projekto¹⁶ tikslas – sutelkti algoritmus, priemones, duomenų rinkinius ir paslaugas siekiant padėti organizacijoms, visų pirma mažosioms ir vidutinėms įmonėms, diegti dirbtinio intelekto sprendimus. Europos dirbtinio intelekto aljansas kartu su AI4EU toliau visoje Europoje kurs dirbtinio intelekto ekosistemą, kad, be kita ko, būtų galima išbandyti dirbtinio intelekto etikos gaires ir skatinti pagarbą į žmogų orientuotam dirbtiniam intelektui.

2020 m. pradžioje Aukšto lygio ekspertų grupė dirbtinio intelekto klausimais, įvertinusi bandymo etapu gautus atsiliepimus, **peržiūrės ir atnaujins gaires**. Remdamasi atlikta peržiūra ir įgyta patirtimi, **Komisija įvertins rezultatus ir pasiūlys tolesnių veiksmų**.

Etiškas dirbtinis intelektas naudingas visiems. Pagarbos pagrindinėms vertybėms ir teisėms užtikrinimas svarbus ne tik savaime – jis taip pat padeda pelnyti visuomenės palankumą ir didinti Europos dirbtinio intelekto įmonių konkurencinį pranašumą, nes sudaro sąlygas kurti į žmones orientuotą patikimą dirbtiniu intelektu pagrįstą etišką ir saugių produktų prekės ženklą. Apskritai jį kurti padeda ir Europos įmonių įgytas su saugių ir patikimų aukštos kokybės produktų tiekimu susijęs geras vardas. Bandomasis etapas padės užtikrinti, kad dirbtinio intelekto produktai pateisintų šias viltis.

2.4. Siekis parengti tarptautines dirbtinio intelekto etikos gaires

Tarptautinės diskusijos dėl dirbtinio intelekto etikos suintensyvėjo tada, kai 2016 m. G 7 pirmininkavusi Japonija įtraukė šį klausimą į darbotvarkės prioritetus. Atsižvelgdama į tai, kad duomenų judėjimo, algoritmų kūrimo ir investicijų į mokslinius tyrimus atžvilgiais dirbtinio intelekto plėtojimas yra tarptautinio pobūdžio, **Komisija toliau stengsis perkelti**

¹⁴ Žr. Komisijos komunikatą dėl susietojo ir automatizuoto judumo, COM(2018) 283.

¹⁵ Įgyvendindama Europos gynybos fondo veiklą Komisija taip pat parengs konkrečias etikos rekomendacijas dėl gynybai skirto dirbtinio intelekto projektų pasiūlymų vertinimo.

¹⁶ <https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019>

Sąjungos požiūrį į pasaulio areną ir pasiekti sutarimą dėl žmogų orientuoto dirbtinio intelekto¹⁷.

Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais atliktas darbas, konkrečiau – parengtas reikalavimų sąrašas ir bendravimas su suinteresuotaisiais subjektais, suteikia Komisijai vertingos papildomos informacijos, kuria galima prisidėti prie pasaulinių diskusijų. Europos Sąjunga gali atlikti lyderės vaidmenį rengiant tarptautines gaires dėl dirbtinio intelekto ir, jei įmanoma, kuriant susijusį vertinimo mechanizmą.

Todėl Komisija:

stiprins bendradarbiavimą su bendraminčiais partneriais:

- nagrinės, koku mastu būtų galima suvienodinti gaires su trečiųjų šalių (pavyzdžiui, Japonijos, Kanados, Singapūro) etikos gairių projektais, ir telkdama dėmesį į šią bendraminčių šalių grupę rengsis platesnio masto diskusijoms, kurioms pagrindą padės veiksmai, kuriais įgyvendinama bendradarbiavimo su trečiosiomis šalimis partnerystės priemonė¹⁸;
- nagrinės, kaip ne ES šalių įmonės ir tarptautinės organizacijos galėtų bandymo ir patvirtinimo veiksmais prisidėti prie gairių bandymo etapo;

toliau aktyviai dalyvaus tarptautinėse diskusijose ir iniciatyvose:

- prisidės prie daugiašalių forumų, pavyzdžiui, G 7 ir G 20, veiklos;
- palaikys dialogą su ne ES šalimis ir rengs dvišalius ir daugiašalius susitikimus, kad būtų pasiektas sutarimas dėl žmogų orientuoto dirbtinio intelekto;
- siekdama propaguoti šią viziją, prisidės prie atitinkamos tarptautinių standartų rengimo organizacijų vykdomos standartizacijos veiklos;
- bendradarbiaudama su atitinkamomis tarptautinėmis organizacijomis, intensyviau rinks ir skleis su viešąja politika susijusias įžvalgas.

3. IŠVADOS

ES grindžiama tam tikromis pagrindinėmis vertybėmis ir yra sukūrusi tvirtą ir subalansuotą šiuo pamatu pagrįstą reglamentavimo sistemą. Atsižvelgiant į dirbtinio intelekto technologijos naujumą ir konkrečius jos keliamus uždavinius, būtina, remiantis šia esama reglamentavimo sistema, parengti dirbtinio intelekto plėtojimo ir naudojimo etikos gaires. Dirbtinis intelektas gali būti laikomas patikimu tik jei yra plėtojamas ir naudojamas laikantis plačiai pripažintamų etinių vertybių.

¹⁷ Sąjungos vyriausioji įgaliotinė užsienio reikalams ir saugumo politikai, padedama Komisijos, atsižvelgs į konsultacijas Jungtinėse Tautose, Pasaulinėje technologijų grupėje ir kituose daugiašaliuose forumuose ir visų pirma koordinuos pasiūlymus, kaip spręsti su tuo susijusius sudėtingus saugumo uždavinius.

¹⁸ 2014 m. kovo 11 d. Europos Parlamento ir Tarybos reglamentas (ES) Nr. 234/2014, kuriuo nustatoma bendradarbiavimo su trečiosiomis šalimis partnerystės priemonė (OL L 77, 2014 3 15, p. 77). Pavyzdžiui, planuojamas projektas „Dirbtinio intelekto orientavimo į žmogų tarptautinis aljansas“ sudarys palankesnes sąlygas vykdyti bendras iniciatyvas su bendraminčiais partneriais, siekiant skatinti taikyti etikos gaires, patvirtinti bendrus principus ir padaryti praktines išvadas. Jis suteiks ES ir bendramintėms šalims galimybę aptarti su ekspertų grupės pasiūlytomis etikos gairėmis dėl dirbtinio intelekto susijusias praktines išvadas, kad būtų galima suformuoti bendrą požiūrį. Be to, jis sudarys sąlygas stebėti dirbtinio intelekto technologijos diegimą pasaulyje. Galiausiai įgyvendinant projektą planuojama organizuoti viešosios diplomatijos veiklą, susijusią su tarptautiniais, pavyzdžiui, G 7, G 20 ir Ekonominio bendradarbiavimo ir plėtros organizacijos, renginiais.

Komisija palankiai vertina šiuo tikslu Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais atliktą darbą. Dabar Komisija pradės su pagrindiniais reikalavimais, kuriuos turi atitikti dirbtinis intelektas, kad būtų laikomas patikimu, susijusį tikslinį bandymo etapą siekdama užtikrinti, kad parengtas dirbtinio intelekto plėtojimo ir naudojimo etikos gairės būtų įmanoma įgyvendinti praktiškai. Komisija taip pat stengsis pasiekti platų visuomenės sutarimą dėl į žmogų orientuoto dirbtinio intelekto, be kita ko, sieks sutarimo su visais susijusiais suinteresuotaisiais subjektais ir mūsų tarptautiniais partneriais.

Dirbtinio intelekto etinis aspektas nėra nei prabangos klausimas, nei papildomas elementas – jis turi būti neatsiejama dirbtinio intelekto plėtojimo dalis. Siekdami kurti į žmogų orientuotą ir pasitikėjimu grindžiamą dirbtinį intelektą užtikriname, kad būtų gerbiamos mūsų pagrindinės visuomenės vertybės, ir formuojame išskirtinį Europos ir jos pramonės kaip lyderių, kuriančių pažangų dirbtinį intelektą, kuriuo galima pasitikėti visame pasaulyje, įvaizdį.

Siekdama užtikrinti, kad etinio dirbtinio intelekto plėtojimo principas Europoje būtų taikomas platesniu mastu, Komisija laikosi visapusiško požiūrio, kuris visų pirma apima toliau nurodytus iki 2019 m. trečiojo ketvirčio įgyvendintinus veiksmus:

- pagal programą „Horizontas 2020“ Komisija pradės keletą **dirbtinio intelekto mokslinių tyrimų kompetencijos centrų tinklų** veiklą. Ji atrinks ne daugiau kaip keturis tinklus, daugiausia dėmesio skiriančius pagrindiniams moksliniams ar technologiniams uždaviniams, tokiems kaip paaiškinamumas ir pažangi žmogaus bei mašinos sąveika, kurie yra pagrindiniai patikimo dirbtinio intelekto elementai;
- Komisija pradės kurti **skaitmeninių inovacijų centrų¹⁹ tinklus**, daugiausia dėmesio skiriančius dirbtiniam intelektui gamybos srityje ir didiesiems duomenims;
- kartu su valstybėmis narėmis ir suinteresuotaisiais subjektais Komisija pradės parengiamąsias diskusijas siekdama sukurti ir įgyvendinti **keitimosi duomenimis ir kuo geresnio naudojimosi bendromis duomenų erdvėmis modelį**, daugiausia skirtą transporto, sveikatos ir pramoninės gamybos sritims²⁰.

Be to, Komisija rengia ataskaitą dėl sunkumų, kuriuos dirbtinis intelektas kelia saugos ir atsakomybės sistemoms, ir rekomendacinį dokumentą dėl Atsakomybės už gaminius direktyvos įgyvendinimo²¹. Tuo pat metu Europos našiosios kompiuterijos bendroji įmonė („EuroHPC“)²² sukurs naujos kartos superkompiuterius, nes kompiuteriniai pajėgumai yra labai svarbūs norint apdoroti duomenis ir mokyti dirbtinį intelektą ir Europa turi valdyti visą skaitmeninę vertės grandinę. Šiuo metu įgyvendinama mikroelektroninių komponentų ir sistemų srities partnerystė su valstybėmis narėmis ir pramone (ECSEL)²³ ir Europos procesorių iniciatyva²⁴ padės sukurti mažos galios procesorių technologiją, reikalingą patikimoms ir saugioms našiosios ir tinklo paribio kompiuterijos sistemoms.

Kaip ir su dirbtinio intelekto etikos gairėmis susijęs darbas, visos šios iniciatyvos grindžiamos **glaudžiu visų suinteresuotųjų subjektų**, valstybių narių, pramonės atstovų, visuomenės

¹⁹ <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs>

²⁰ Reikiami ištekliai bus sutelkti iš programos „Horizontas 2020“ (pagal ją dirbtiniam intelektui 2018–2020 m. skirta beveik 1,5 mlrd. EUR) ir ją pakeisti turinčios programos „Europos horizontas“, taip pat skaitmeninėms paslaugoms skirtos Europos infrastruktūros tinklų priemonės dalies ir ypač būsimos Skaitmeninės Europos programos lėšų. Projektams taip pat bus naudojami privačiojo sektoriaus teikiami ir pagal valstybių narių programas skiriami ištekliai.

²¹ Žr. Komisijos komunikatą „Dirbtinis intelektas Europai“, COM(2018) 237.

²² <https://eurohpc-ju.europa.eu>

²³ www.ecsel.eu

²⁴ www.european-processor-initiative.eu

veikėjų ir piliečių **bendradarbiavimu**. Apskritai Europos požiūris į dirbtinį intelektą rodo, kad ekonominis konkurencingumas ir visuomenės pasitikėjimas turi būti grindžiami tomis pačiomis pagrindinėmis vertybėmis ir stiprinti vienas kitą.