Europos Komisijos

AUKŠTO LYGIO EKSPERTŲ GRUPĖ DIRBTINIO INTELEKTO KLAUSIMAIS

PROJEKTAS
PATIKIMO DIRBTINIO INTELEKTO ETIKOS GAIRĖS

SANTRAUKA

Darbinis dokumentas suinteresuotųjų subjektų konsultacijoms

Briuselis, 2018 m. gruodžio 18 d.

PATIKIMO DIRBTINIO INTELEKTO

ETIKOS GAIRIŲ PROJEKTAS

Aukšto lygio ekspertų grupė dirbtinio intelekto klausimais
Patikimo dirbtinio intelekto etikos gairių projektas

Europos Komisija
Komunikacijos generalinis direktoratas

Asmuo ryšiams	Nathalie Smuha – Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais koordinatorė
E. paštas	CNECT-HLG-AI@ec.europa.eu

Europos Komisija
B-1049 Briuselis

Dokumentas paskelbtas 2018 m. gruodžio 18 d., Anglų kalba.

Šį darbinį dokumentą parengė Aukšto lygio ekspertų grupė dirbtinio intelekto klausimais, nedarydama poveikio individualiai jos narių pozicijai dėl konkrečių punktų ir galutinei dokumento versijai. Šis dokumentas bus toliau rengiamas ir jo galutinė versija bus pateikta 2019 m. kovo mėn. po Europos dirbtinio intelekto aljanse surengtų konsultacijų su suinteresuotaisiais subjektais.

Nei Europos Komisija, nei kiti Komisijos vardu veikiantys asmenys neatsako už tai, kaip gali būti naudojama toliau pateikta informacija. Už šio darbinio dokumento turinį atsakinga tik Aukšto lygio ekspertų grupė dirbtinio intelekto klausimais. Nors rengti gaires padėjo Komisijos tarnybų darbuotojai, šiame dokumente pateiktas požiūris atspindi Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais nuomonę ir jokiais atvejais negali būti laikomas oficialia Europos Komisijos pozicija. Tai Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais pirmojo darbo rezultato projektas. Galutinė šio dokumento versija bus pateikta Komisijai 2019 m. kovo mėn. Antrojo darbo rezultato – rekomendacijų dėl dirbtinio intelekto politikos ir investicijų – galutinė versija bus pateikta 2019 m. viduryje.

Daugiau informacijos apie Aukšto lygio ekspertų grupę dirbtinio intelekto klausimais galima rasti internete (https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence).

SANTRAUKA

Šiame darbiniame dokumente pateiktas Dirbtinio intelekto etikos gairių, kurias parengė Europos Komisijos Aukšto lygio grupė dirbtinio intelekto klausimais, projektas. Šio dokumento galutinė versija turi būti parengta 2019 m. kovo mėn.

Dirbtinis intelektas mūsų laikais yra viena labiausiai pokyčius skatinančių jėgų, galinti pakeisti visuomenės struktūrą. Tai puiki galimybė didinti gerovę ir skatinti ekonomikos augimą – būtent to ir turi siekti Europa. Pastarąjį dešimtmetį dėl didelių skaitmeninių duomenų kiekių prieinamumo, galingos skaičiavimo architektūros ir laimėjimų vystant dirbtinio intelekto metodus, pavyzdžiui, mašinų mokymąsi, padaryta didelė pažanga. Pagrindiniai dirbtiniu intelektu pagrįsti patobulinimai autonominių transporto priemonių, sveikatos priežiūros, namų ar paslaugų robotų, švietimo ar kibernetinio saugumo srityse padeda gerinti mūsų kasdienio gyvenimo kokybę. Be to, dirbtinis intelektas labai svarbus sprendžiant daugelį svarbių pasauliui kylančių uždavinių, tokių kaip visuotinė sveikata ir gerovė, klimato kaita, patikimos teisinės ir demokratinės sistemos, ir kitus Jungtinių Tautų darnaus vystymosi tiksluose nurodytus uždavinius.

Nors dirbtinis intelektas gali suteikti milžinišką naudą žmonėms ir visuomenei, jis taip pat kelia tam tikrą riziką, kurią reikėtų tinkamai suvaldyti. Kadangi apskritai dirbtinio intelekto nauda yra didesnė nei jo keliama rizika, privalome pasirinkti tokią kryptį, kuria einant būtų **kuo geriau išnaudojami dirbtinio intelekto pranašumai ir kuo labiau mažinama jo rizika**. Siekiant užtikrinti, kad einame teisinga kryptimi, **dirbtiniam intelektui turime taikyti į žmogų orientuotą požiūrį**, stengiantis nepamiršti, kad mūsų tikslas yra ne tiesiog kurti ir naudoti dirbtinį intelektą, o užtikrinti didesnę gerovę žmonėms. **Patikimas dirbtinis intelektas bus mūsų kelrodė žvaigždė**, nes tik pasitikėdami technologijomis žmonės galės patikimai ir visapusiškai pasinaudoti dirbtinio intelekto teikiama nauda.

Patikimam dirbtiniam intelektui būdingi **du aspektai:** 1) jis turėtų atitikti pagrindines teises, taikomas taisykles ir pagrindinius principus bei vertybes, užtikrinant **etinį tikslą**, ir 2) jis turėtų būti **techniškai patvarus** ir patikimas, nes net ir turint gerų ketinimų, bet per mažai technologinių žinių, žalą galima padaryti netyčia.

Todėl šiose gairėse nustatoma **patikimo dirbtinio intelekto sistema**.

- **I skyriuje** nagrinėjama, kaip **užtikrinti dirbtinio intelekto etinį tikslą**, nustatant pagrindines teises, principus ir vertybes, kuriuos jis turėtų atitikti.
- Remiantis šiais principais, **II skyriuje** pateikiamos patikimo dirbtinio intelekto **įgyvendinimo rekomendacijos**, kuriose sprendžiami tiek etinio tikslo, tiek techninio patvarumo klausimai. Tuo tikslu jose išvardijami patikimam dirbtiniam intelektui keliami reikalavimai ir apžvelgiami techniniai ir netechniniai metodai, kuriuos galima taikyti jam įgyvendinti.
- **III skyriuje aptariamas praktinis** reikalavimų taikymas – pateikiamas konkretus, tačiau nebaigtinis patikimo dirbtinio intelekto vertinimo kriterijų sąrašas. Tuomet šis sąrašas pritaikomas konkretiems naudojimo atvejams.

Taigi, priešingai nei kituose su etišku dirbtiniu intelektu susijusiuose dokumentuose, šiose gairėse nesiekiama pateikti dar vieno dirbtinio intelekto pagrindinių vertybių ir principų sąrašo, bet pateikiamos rekomendacijos, kaip juos konkrečiai įgyvendinti ir praktiškai pritaikyti dirbtinio intelekto sistemose. Tokios rekomendacijos pateikiamos trimis abstraktumo lygiais: nuo abstrakčiausio I skyriaus (pagrindinės teisės, principai ir vertybės) iki konkrečiausio III skyriaus (vertinimo kriterijų sąrašas).

Gairės skirtos visiems **susijusiems suinteresuotiesiems subjektams, kuriantiems, diegiantiems ar naudojantiems dirbtinį intelektą**, įskaitant įmones, organizacijas, tyrėjus, viešųjų paslaugų teikėjus, institucijas, fizinius asmenis ar kitus subjektus. Galutinėje šių gairių versijoje bus pateiktas mechanizmas, leisiantis suinteresuotiesiems subjektams savanoriškai jas patvirtinti.

Svarbu tai, kad šiomis gairėmis neketinama pakeisti jokios formos politikos formavimo ar reguliavimo veiksmų (šie klausimai bus aptarti antrajame Aukšto lygio ekspertų grupės dirbtinio intelekto klausimais darbo rezultate – rekomendacijose dėl politikos ir investicijų, kurios turi būti parengtos 2019 m. gegužės mėn.) ir jomis nesiekiama sutrukdyti jų imtis. Be to, šios gairės turėtų būti laikomos kintančiu dokumentu, kurį bėgant laikui reikia nuolat atnaujinti, kad jis visada būtų aktualus, nes technologijos ir mūsų žinios apie jas kinta. Todėl šis dokumentas turėtų būti laikomas atspirties tašku diskusijoms apie **patikimą Europoje sukurtą dirbtinį intelektą**.

Nors Europa savo etinį požiūrį į dirbtinį intelektą gali skleisti tik būdama konkurencinga pasauliniu mastu, **etinis požiūris į dirbtinį intelektą yra labai svarbus siekiant sudaryti sąlygas atsakingam konkurencingumui**, nes taip bus galima įgyti naudotojų pasitikėjimą ir sudaryti sąlygas platesniam dirbtinio intelekto naudojimui. Šių gairių tikslas – ne slopinti dirbtinio intelekto inovacijas Europoje, o pasitelkti etiką kaip įkvėpimo šaltinį kuriant unikalios rūšies dirbtinį intelektą, kuriuo būtų siekiama apsaugoti ir žmones, ir bendrą gerovę bei teikti jiems naudą. Taip Europa gali užimti pirmaujančias pozicijas pažangiojo, saugaus ir etiško dirbtinio intelekto srityje. Tik užtikrinus patikimumą Europos piliečiai galės visapusiškai pasinaudoti dirbtinio intelekto teikiama nauda.

Galiausiai šiomis gairėmis taip pat siekiama **skatinti apmąstyti ir aptarti** dirbtinio intelekto etinį pagrindą ne tik Europoje, bet ir **pasauliniu mastu**.

REKOMENDACIJOS

Kiekviename gairių skyriuje pateikiamos rekomendacijos, kaip užtikrinti patikimą dirbtinį intelektą, skirtos visiems susijusiems suinteresuotiesiems subjektams, kuriantiems, diegiantiems ar naudojantiems dirbtinį intelektą. Toliau pateikiama šių rekomendacijų santrauka.

I skyrius. Pagrindinės etinio tikslo užtikrinimo rekomendacijos

- Užtikrinkite, kad dirbtinis intelektas būtų **orientuotas į žmogų**. Dirbtinis intelektas turėtų būti kuriamas, diegiamas ir naudojamas siekiant **etinio tikslo**, įtvirtinto ir atspindėto pagrindinėse teisėse, socialinėse vertybėse ir etikos principuose (darymo gera, nekenkimo, žmonių savarankiškumo, teisingumo ir paaiškinamumo). Tai labai svarbu siekiant sukurti **patikimą dirbtinį intelektą**.
- Remkitės pagrindinėmis teisėmis, etikos principais ir vertybėmis, kad galėtumėte įvertinti galimą dirbtinio intelekto poveikį žmonėms ir bendrai gerovei ateityje. **Ypatingą dėmesį** skirkite situacijoms, susijusioms su labiau **pažeidžiamomis grupėmis**, pavyzdžiui, vaikais, neįgaliaisiais ar mažumomis, arba situacijoms, kurioms būdinga **galios ar informacijos asimetrija**, pavyzdžiui, tarp darbdavių ir darbuotojų arba įmonių ir vartotojų.
- Pripažinkite ir nepamirškite, kad nors dirbtinis intelektas teikia didžiulę naudą žmonėms ir visuomenei, jis taip pat gali turėti neigiamą poveikį. Neprarraskite budrumo didelį susirūpinimą keliančiose srityse.

II skyrius. Pagrindinės patikimo dirbtinio intelekto įgyvendinimo rekomendacijos

- **Nuo pirmųjų projektavimo etapų** įtraukite **patikimam** dirbtiniam intelektui **keliamus reikalavimus**: atskaitomybės, duomenų valdymo, visiems tinkamo projekto, dirbtinio intelekto savarankiškumo valdymo (žmogaus priežiūros), nediskriminavimo, pagarbos žmogaus savarankiškumui, pagarbos privatumui, patvarumo, saugumo, skaidrumo.
- Apsvarstykite, kokiais techniniais ir netechniniais metodais galima užtikrinti, kad šie reikalavimai būtų įgyvendinti dirbtinio intelekto sistemoje. Be to, nepamirškite šių reikalavimų burdami grupę, kuri dirbs su sistema, kurdami pačią sistemą, bandymų aplinką ir galimas sistemos taikomąsias programas.

- Aiškiai ir iniciatyviai teikite **informaciją suinteresuotiesiems subjektams** (klientams, darbuotojams ir t. t.) apie dirbtinio intelekto sistemos galimybes ir apribojimus, kad jie galėtų susidaryti realius lūkesčius. Šiuo atžvilgiu labai svarbu užtikrinti dirbtinio intelekto sistemos **atsekamumą**.
- Įtraukite patikimą dirbtinį intelektą į **organizacijos kultūrą** ir informuokite suinteresuotuosius subjektus, kaip patikimas dirbtinis intelektas įgyvendinamas projektuojant ir naudojant dirbtinio intelekto sistemas. Patikimas dirbtinis intelektas taip pat gali būti įtrauktas į organizacijos deontologijos įstatus ar elgesio kodeksą.
- Užtikrinkite, kad projektuojant ir plėtojant dirbtinio intelekto sistemą dalyvautų ir **būtų įtraukti suinteresuotieji subjektai**. Be to, sudarydami grupes, kurios kurs, įgyvendins ir išbandys produktą, užtikrinkite **įvairovę.**
- Stenkitės **supaprastinti galimybes atlikti** dirbtinio intelekto sistemų **auditą**, ypač esant kritinėms aplinkybėms ar situacijoms. Kiek įmanoma, projektuokite savo sistemą taip, kad būtų galima atsekti atskirus sprendimus dėl įvairių jūsų įvesčių, duomenų, iš anksto parengtų modelių ir t. t. Be to, nustatykite dirbtinio intelekto sistemos **paaiškinimo metodus**.
- Užtikrinkite konkretų **atskaitingumo valdymo** procesą.
- Numatykite **mokymo ir švietimo** kursus ir užtikrinkite, kad vadovai, kūrėjai, naudotojai ir darbuotojai turėtų žinių apie patikimą dirbtinį intelektą ir būtų parengti darbui su juo.
- Nepamirškite, kad siekiant skirtingų tikslų gali kilti esminių nesutarimų (dėl skaidrumo gali atsirasti netinkamo naudojimo galimybių; nuokrypių nustatymas ir ištaisymas gali būti nesuderinami su privatumo apsauga). Praneškite apie tokius kompromisus ir juos dokumentuokite.
- Siekdami prisidėti prie patikimam dirbtiniam intelektui keliamų reikalavimų įgyvendinimo, skatinkite mokslinius tyrimus ir inovacijas.

II skyrius. Pagrindinės patikimo dirbtinio intelekto vertinimo rekomendacijos

- Parenkite patikimo dirbtinio intelekto **vertinimo kriterijų sąrašą**, kuris būtų naudojamas kuriant, diegiant ar naudojant dirbtinį intelektą, ir pritaikykite jį konkrečiam sistemos naudojimo atvejui.
- Nepamirškite, kad vertinimo kriterijų sąrašas **niekada nebus baigtinis** ir kad patikimo dirbtinio intelekto užtikrinimo esmė – tai ne varnele pažymėti langeliai, o nuolatinis reikalavimų nustatymo, sprendimų vertinimo ir geresnių rezultatų užtikrinimo procesas per visą dirbtinio intelekto sistemos gyvavimo ciklą.

Šiomis rekomendacijomis prisidedama prie į žmogų orientuotu požiūriu grindžiamos dirbtinio intelekto vizijos, kuri leis Europai tapti pasaulyje pirmaujančia novatore etiško, saugaus ir pažangiojo dirbtinio intelekto srityje. Jomis siekiama padėti įgyvendinti **patikimą Europoje sukurtą dirbtinį intelektą** ir taip didinti Europos piliečių gerovę.