Kas yra Dirbtinio intelekto aktas?

2023-10-06

Dirbtinio intelekto (DI) aktas yra Europos Komisijos pasiūlymas reglamentuoti Europoje sukurtą arba naudojamą dirbtinį intelektą. Šiuo aktu yra siekiama užtikrinti, kad DI atitiktų žmogaus teises, įskaitant žmonių saugumo, privatumo, nediskriminavimo, skaidrumo, žmogaus atliekamos priežiūros ir socialinės bei aplinkos gerovės užtikrinimą.

Be to, DI akte taip pat yra skiriama daug dėmesio vartotojų apsaugai ir „rinkos reguliavimui“. Pasiūliusi DI aktą, Europos Komisija pabrėžė, jog aiškios ir iš anksto žinomos taisyklės paskatins naujovių diegimą. Komisija manė, kad jei ES galės suteikti teisinio aiškumo, įmonės bus labiau linkusios skirti savo laiką ir pinigus produktų kūrimui. Taip joms daugiau nebereikės nerimauti, ar ateityje nekils teisinių problemų dėl jų gaminių.

Šis teisės aktas yra svarstomas jau kurį laiką. Birželio mėnesį Europos Parlamentas pareiškė savo derybinę poziciją. Remiantis jo pozicija, yra siūloma gerinti žmogaus teisių apsaugą, pavyzdžiui, įpareigojant žmonėms su negalia pritaikyti didelės rizikos DI sistemas, uždraudžiant viešose vietose naudoti veido atpažinimą realiuoju laiku ir uždraudžiant naudoti emocijų pripažinimą itin delikačiose situacijose (pavyzdžiui, kai šiomis sistemomis naudojasi darbdaviai, mokyklos, policija ir migracijos institucijos).

Šiuo metu dvi Europos Sąjungos teisėkūros institucijos (Europos Parlamentas ir Europos Sąjungos Taryba) derasi dėl galutinio teisės akto teksto. Todėl svarbu to nepamiršti, nes kai kurie punktai, aprašyti tolesnėje straipsnio dalyje, **gali keistis**.

Bet kodėl tai yra svarbu? Šiame straipsnyje apžvelgsime galimą jo poveikį žmonių su negalia teisėms.

Naršymo meniu:

Poveikis žmonių su negalia teisėms. Sistemų pritaikymas

Vienas iš pagrindinių Europos negalios forumo reikalavimų buvo užtikrinti, kad žmonėms su negalia būtų pritaikytos ne tik priemonės ir sistemos, kuriose yra naudojamas DI, bet ir priemonės, skirtos kurti tokias DI sistemas. Naudotojams turint prieigą prie šių priemonių, jų kūrėjams su negalia būtų žymiai paprasčiau bendradarbiauti, o aktyvistams vertinti jų keliamą riziką.

Europos Parlamentas į savo pozicijos dokumentą įtraukė mūsų pasiūlytą **naują reikalavimą dėl DI sistemų prieinamumo**, kad didelės rizikos DI sistemas būtų privaloma pritaikyti žmonėms su negalia.

Šiuo metu siekiame, kad šie privalomi reikalavimai taip pat galiotų vidutinės ir nedidelės rizikos DI sistemoms.

Poveikis žmonių su negalia teisėms. Rizikos

DI sistemos, priklausomai nuo jų keliamos rizikos pagrindinėms žmogaus teisėms lygmens, Akte yra skirstomos į keturias kategorijas. Panagrinėkime siūlomame DI akte aprašytas skirtingas kategorijas ir su jomis susijusius apribojimus.

Nepriimtina rizika

Kai kurie DI naudojimo būdai yra laikomi tokiais pavojingais, kad jų naudojimas privalo būti uždraustas. Šiuo atveju nepakanka įvesti griežtų taisyklių, kuriomis būtų ribojamas tokių technologijų naudojimas. Ši rizika apima piliečių vertinimą, kuriuo yra pažeidžiamos pagrindinės žmogaus teisės ir žmogaus orumas.

Pavyzdžiui, DI grįsta piliečių vertinimo sistema, nustatanti, ar esate „geras pilietis“. Ji gali analizuoti, kokios rūšies maistą valgote arba ar parduotuvėje perkate daug alkoholinių gėrimų, kad nustatytų jūsų „patikimumą“ arba „sveikatos būklę“. Tokia sistema, remdamasi jūsų propaguojamu gyvenimo būdu, rekomenduotų jums nustatyti didesnes draudimo įmokas arba jūsų potencialiam darbdaviui patartų, ar jus verta samdyti, ar ne.

ES Komisija pasiūlė šiuo teisės aktu taip pat riboti teisėsaugos pareigūnų naudojamą veido atpažinimą viešose vietose realiuoju laiku, jei tai nėra „būtina“. Tai yra vienas iš labiausiai ginčytinų punktų.

Didelė rizika

Viena iš DI akto savybių yra didelės rizikos priskyrimas tam tikroms sistemoms. Šios sistemos yra naudojamos, pavyzdžiui, priimant sprendimą, kas bus kviečiamas į pokalbį dėl darbo, kas bus priimtas į tam tikras universitetinių studijų programas, ar bus patvirtinta paraiška dėl valstybės pašalpos skyrimo arba ar turėtų būti suteikta banko paskola.

Svarbu prisiminti, kad DI sistemų skirstymas į skirtingas rizikos kategorijas priklauso nuo politinių sprendimų, o politikai tokio skirstymo atžvilgiu turi skirtingas ideologines nuomones.

Šiuo aktu yra siūloma **griežtai reguliuoti** aukštos rizikos sistemų naudojimą.

Be to, didelės rizikos DI sistemų teikėjai, prieš išleisdami tokią sistemą į rinką, turėtų pateikti jos techninius dokumentus. Idealiausiu atveju, bus uždrausta dauguma naudojimo būdų – mes ypač reikalavome uždrausti veido atpažinimo technologijas, kurias naudoja policija arba kurios yra pasitelkiamos darbo pokalbių metu. Tačiau, jei jas bus leidžiama naudoti, labai svarbu užtikrinti, kad policijos naudojamos veido atpažinimo programinės įrangos naudojimo vadove, naudotojai būtų aiškiai informuojami apie tokios technologijos apribojimus. Veido atpažinimo programinė įranga turi didesnę tikimybę suklysti, bandydama nustatyti tam tikras marginalizuotas grupes, pavyzdžiui, žmones su negalia, turinčia įtakos jų fizinei išvaizdai, ir žmones, turinčius tamsesnę odos spalvą. Norint sumažinti riziką, kad policija suims ne tą asmenį, labai svarbu užtikrinti, kad vadove policijos pareigūnai būtų įspėjami, kad

technologija marginalizuotų grupių atžvilgiu veikia nepatikimai. Taigi policija turi būti ypač atsargi ir dažniau informaciją tikrinti rankiniu būdu, o ne pasikliaudama veido atpažinimo sistema, galinčia žmogų su negalia nurodyti kaip galimai atitinkantį nurodytus kriterijus.

Didelės rizikos DI sprendimai taip pat turėtų būti kuriami, jų veikimą stebint žmonėms, siekiant nustatyti sistemoje kylančias problemas. To reikia, siekiant užtikrinti, kad sistemos veiktų tiksliai, patikimai ir saugiai.

Vidutinė rizika

Vidutinės rizikos sistemos – tai sistemos, kurios, politikų nuomone, nėra itin pavojingos. Pavyzdžiui, šios sistemos bando analizuoti emocijas arba nustatyti, kokiai demografinei grupei priklausote.

Siūlomu reglamentu vidutinės rizikos DI sistemoms yra nustatomi **minimalūs reikalavimai**.

Vidutinės rizikos DI sistemos taip pat gali apimti emocijų atpažinimo arba biometrinio kategorizavimo sistemas, kuriomis nėra nustatoma asmens tapatybė (kai žmonės yra skirstomi į kategorijas pagal jų lytinę tapatybę, akių spalvą, amžių, etninę kilmę ar kitas savybes), arba „sintetinių vaizdo klastočių“ DI sistemas, kuriomis yra kuriamas arba klastojamas turinys. Paslaugų teikėjai privalo informuoti naudotojus, kai šios sistemos yra naudojamos atsakymams arba paslaugoms teikti, išskyrus atvejus, kai jomis naudojasi teisėsaugos institucijos.

Nedidelė rizika

Nedidelės rizikos DI sistemos – tai sistemos, kurios, politikų nuomone, nekelia grėsmės žmogaus teisėms. Šios sistemos vis dar turi atitikti kitus teisės aktus, pavyzdžiui, Bendrąjį duomenų apsaugos reglamentą, tačiau DI aktu jiems nebus nustatomos papildomos taisyklės.

Siūlomame DI akte taip pat nenustatomi **jokie reikalavimai** tokioms nedidelės rizikos DI sistemoms, kaip brukalų filtrai. Tačiau šiame akte yra siūloma sukurti vidutinės ir nedidelės rizikos DI sistemų teikėjų elgesio kodeksus, kad jie savanoriškai laikytųsi taisyklių, kurios būtų panašios į aukštos rizikos DI sistemoms galiojančias privalomas taisykles.

Nauja „pagrindinių modelių“ kategorija

Europos Parlamentas pristatė naują rizikos kategoriją („Pagrindiniai modeliai“), kuri iki šiol nebuvo įtraukta į pradinį Europos Komisijos pasiūlymą.

Pagrindiniai modeliai yra DI sprendimai, kuriuos kiekvienas žmogus gali naudoti ir integruoti į savo programėlę. Tarkim, matematikos mokytoja nori sukurti programėlę, kuri mokiniams padėtų atlikti namų darbus. Mokytoja programėlę gali išmokyti matematikos, tačiau programėlė vis tiek turės gebėti suprasti ir vartoti žmonių kalbą, kad galėtų kalbėtis su mokiniais. Čia pasitelkiami tokie pagrindiniai modeliai, kaip GPT4 arba Llama, kuriuose jau yra vartojama žmonių kalba, todėl juos galima naudoti kaip pagrindinę programėlės pokalbių

funkciją. Tuomet mokytoja galės pagrindinį savo dėmesį skirti savo kompetencijos sričiai – matematikai.

Remiantis pradiniu Komisijos pasiūlymu, tokie pokalbių robotai kaip ChatGPT būtų priskiriami vidutinės rizikos kategorijai. Pavyzdžiui, pirkėjas, įsigydamas lėktuvo bilietą pokalbio su klientų aptarnavimo skyriaus atstovu metu, turėtų teisę žinoti, kad jis bendrauja su pokalbių robotu, o ne žmogumi.

Rinkoje pasirodžius ChatGPT, pasikeitė Europos Parlamento turėtas požiūris, todėl buvo imtasi griežtesnių kontrolės priemonių.

ChatGPT modeliui „staiga užvaldžius pasaulį“, Europos Parlamento nariai suvokė jo potencialą ir galimą pavojų, todėl panoro šias technologijas priskirti didelės rizikos kategorijai. Tačiau po ilgų derybų – ir intensyvaus šiuos modelius kuriančių įmonių lobizmo – dabar tikėtina, jog pagrindiniai modeliai bus priskirti vidutinės rizikos kategorijai, kuriai bus taikomi papildomi reikalavimai.

Taigi nors žmogaus teisių gynimo ir privatumo apsaugos organizacijos vis dar reikalauja pagrindinius modelius priskirti didelės rizikos kategorijai, taip greičiausiai nenutiks.

Kodėl pagrindiniai modeliai turėtų būti įtraukti į didelės rizikos sąrašą?

Dirbtinį intelektą galima išmokyti spręsti užduotis įvairiais būdais – viso to rezultatas yra vadinamas baziniu modeliu. Šis modelis yra mokomas, jam pateikiant įvairius duomenis, kuriais remdamasis jis geba spręsti įvairiausias užduotis. Pavyzdžiui, kelionių agentūra gali naudoti pagrindinį modelį, kad sukurtų pokalbių robotą, gebantį dirbti pagal jam pateiktus konkrečius agentūros dokumentus, kad suprastų ir gebėtų atsakyti į gautas klientų užklausas.

Tačiau kiti DI pagrindiniai modeliai gali turėti trūkumų, pavyzdžiui, būti šališki ir informaciją išsigalvoti. Šiems trūkumams vis pasireiškiant bet kuriame produkte arba paslaugoje, sukurtoje naudojant tokį pagrindinį modelį, kelionių agentūrai gali tekti už tai teisiškai atsakyti. Pavyzdžiui, jei klientus aptarnaujantis pokalbių robotas pradėtų juos diskriminuoti dėl pagrindiniam modeliui būdingo šališkumo, kelionių agentūra gali patirti teisines pasekmes.

Pavojus dar tik priešaky

2023 m. rugsėjo mėnesį daugiau kaip 100 pilietinės visuomenės organizacijų paragino derybų dėl DI akto dalyvius išspręsti didžiulius trūkumus, kuriais būtų galima pasinaudoti, siekiant priskirti DI sistemas „klaidingai kategorijai“. Derybų dalyviai svarsto, dėl kokių išimčių galėtų atsirasti pavojingų spragų, susijusių su taisyklių taikymu didelės rizikos kategorijai priskiriamam DI. Pavyzdžiui, įmonė kurianti DI, kuris bus naudojamas tokiose itin rizikingose situacijose, kaip įdarbinimas arba policijos darbas, galėtų save atleisti nuo taisyklių laikymosi.

Praktiškai tai reikštų, kad DI kuriančiai įmonei būtų leista dirbtinį intelektą vertinti savarankiškai be jokios priežiūros institucijos patvirtinimo. Tokiu būdu, įmonė pati nuspręstų, ar jos sistema kelia pavojų žmogaus teisėms, ar ne. Organizacijos baiminosi, kad taip būtų labai sumenkinta ambicingų taisyklių, kurių tikslas yra apsaugoti žmones nuo didelės rizikos

DI daromos žalos, svarba. Todėl organizacijos paragino derybų dalyvius grįžti prie pradinio Europos Komisijos pasiūlymo.