



Euroopa Komisjoni

KÕRGETASEMELINE TEHISINTELLEKTI EKSPERDIRÜHM



KAVAND EETIKASUUNISED USALDUSVÄÄRSE TEHISINTELLEKTI ARENDAmiseks KOMMENTEERITUD KOKKUVÕTE

Töödokument sidusrühmadega konsulteerimiseks

KAVANDATAVAD EETIKASUUNISED

USALDUSVÄÄRSE TEHISINTELLEKTI ARENDAMISEKS



Kõrgetasemeline tehisintellekti eksperdirühm
Kavandatavad eetikasuunised usaldusväärse tehisintellekti arendamiseks

Euroopa Komisjon
Teabevahetuse peadirektoraat

Kontaktisik Nathalie Smuha – kõrgetasemelise tehisintellekti eksperdirühma koordinaator
E-post CNECT-HLG-AI@ec.europa.eu

Euroopa Komisjon
B-1049 Brüssel

Dokument avaldati 18. detsembril 2018 inglise keeles.

Käesoleva töödokumendi on koostanud kõrgetasemeline tehisintellekti eksperdirühm ning see ei piira eksperdirühma liikmete individuaalseid seisukohti konkreetsetes küsimustes ega dokumendi lõplikku teksti. Töö dokumendi kallal jätkub ja selle lõplik versioon esitatakse 2019. aasta märtsis pärast seda, kui sidusrühmadega on Euroopa tehisintellekti liidu kaudu konsulteeritud.

Euroopa Komisjon ega ükski komisjoni nimel tegutsev isik ei vastuta käesolevas dokumendis sisalduva teabe kasutamise eest. Selle töödokumendi sisu eest vastutab ainsana kõrgetasemeline tehisintellekti eksperdirühm. Ehkki suuniseid on aidanud ette valmistada ka komisjoni talituste töötajad, kajastavad käesolevas dokumendis väljendatud seisukohad kõrgetasemelise tehisintellekti eksperdirühma arvamust ja neid ei saa mingil juhul pidada Euroopa Komisjoni ametlikuks seisukohaks. Tegemist on kõrgetasemelise tehisintellekti eksperdirühma esimese dokumendi kavandiga. Selle lõplik tekst esitatakse komisjonile 2019. aasta märtsis. Teise dokumendi – tehisintellekti poliitika ja investeerimissoovitused – lõplik tekst esitatakse 2019. aasta keskel.

Täiendavat teavet kõrgetasemelise tehisintellekti eksperdirühma kohta leiab internetist (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

Euroopa Komisjoni dokumentide taaskasutamist reguleeritakse komisjoni otsusega 2011/833/EL (ELT L 330, 14.12.2011, lk 39). ELi autoriõigusega hõlmamata fotode või muu materjali kasutamiseks ja taasesitamiseks tuleb küsida luba otse autoriõiguse valdaja käest.

KOMMENTEERITUD KOKKUVÕTE

Käesolev töödokument kujutab endast tehisintellekti arendamise eetikasuuniste kavandit, mille on koostanud Euroopa Komisjoni kõrgetasemeline tehisintellekti eksperdirühm ja mille lõplik tekst avaldatakse 2019. aasta märtsis.

Tehisintellekt on meie aja üks olulisemaid kujundavaid jõude, mis meie ühiskonna struktuuri kindlasti muudab. See pakub suurepäraselt võimalust suurendada heaolu ja majanduskasvu ning Euroopa peab püüdma sellest võimalusest kinni haarata. Viimase kümne aasta jooksul on tehtud suuri edusamme tänu tohutule hulgale digitaalsetele andmetele, võimsatele andmetöötlusarhitektuuridele ja tehisintellekti tehnika arengule, näiteks masinõppele. Meie elu kvaliteeti aitavad iga päev parandada olulised tehisintellektil põhinevad saavutused autonoomsete sõidukite, tervishoiu, kodu-/teenindusrobotite, hariduse või küberturvalisuse valdkonnas. Lisaks sellele aitab tehisintellekt lahendada olulisi ülesandeid näiteks sellistes valdkondades nagu ülemaailmne heaolu ja tervishoid, kliimamuutused, usaldusväärsed õiguslikud ja demokraatlikud süsteemid ning tegeleda muude ÜRO kestliku arengu eesmärkide raames väljendatud mureküsimumestega.

Tehisintellekt võib tuua üksikisikutele ja ühiskonnale väga palju kasu, aga sellega võivad ühtlasi kaasned teatavad riskid, mida tuleks nõuetekohaselt juhtida. Võttes arvesse, et üldiselt on tehisintellektist saadav kasu suurem kui sellega kaasnevad riskid, peame püüdma leida tee, mis aitab **tehisintellekti eeliseid maksimeerida ja selle riske minimeerida**. Õigel teel püsimiseks **on tarvis kujundada välja tehisintellekti inimkeskne käsitus**, mis aitaks meil meeles pidada, et tehisintellekti arendamist ja kasutamist ei tohiks pidada omaette eesmärgiks, vaid vahendiks, mille otstarve on suurendada inimeste heaolu. **Meie juhtmotiiv on tehisintellekti usaldusväärsus**, sest inimesed saavad tehisintellekti eeliseid enesekindlalt ja täies ulatuses kasutada vaid siis, kui nad seda tehnoloogiat usaldavad.

Usaldusväärsel tehisintellektil on **kaks aspekti**: 1) see peaks järgima põhiõigusi, kohaldatavaid õigusnorme, aluspõhimõtteid ja väärtusi, et täita **eetilist eesmärki**, ning 2) see peaks olema **tehniliselt töökindel** ja usaldusväärne, sest isegi kui kavatsused on head, võib tehnoloogiline oskamatus tekitada tahtmatut kahju.

Käesolevates suunistes on seetõttu sätestatud **raamistik usaldusväärse tehisintellekti arendamiseks**.

- **I peatükis** käsitletakse **tehisintellekti eetilist eesmärki** ning selles kirjeldatakse põhiõigusi, põhimõtteid ja väärtusi, mida tehisintellekt peaks järgima.
- Nendest põhimõtetest lähtudes on **II peatükis** esitatud **juhised** usaldusväärse tehisintellekti **loomiseks**, pidades silmas nii selle eetilist eesmärki kui ka tehnilist töökindlust. Selleks on loetletud usaldusväärsele tehisintellektile esitatavad nõuded ning antud ülevaade tehnilistest ja muudest meetoditest, mida saab tehisintellekti rakendamiseks kasutada.
- **III peatükis** viiakse need nõuded **kasutuslikule kujule** – selleks on seal esitatud usaldusväärse tehisintellekti konkreetne, kuid mitteammendav hindamisloend. Seda loendit kohandatakse seejärel konkreetsetele kasutusjuhtudele.

Erinevalt muudest dokumentidest, milles käsitletakse eetilist tehisintellekti, ei ole käesolevate suuniste eesmärk pakkuda veel üht tehisintellekti alusväärtuste ja -põhimõtete loetelu, vaid anda hoopis juhiseid nende konkreetse rakendamise ja juurutamise kohta tehisintellekti süsteemides. Suunised on koostatud kolmel üldistustasemel, alates kõige üldisemast I peatükist (põhiõigused, põhimõtted ja väärtused) ja lõpetades kõige konkreetsema III peatükiga (hindamisloend).

Käesolevad suunised on mõeldud kõigile **sidusrühmadele, kes arendavad, juurutavad või kasutavad tehisintellekti**, sealhulgas äriühingutele, organisatsioonidele, teadlastele, avalikele teenistustele, institutsioonidele, üksikisikutele või muudele üksustele. Dokumendi lõplikus tekstis tutvustatakse

mehhanismi, mis võimaldab sidusrühmadel suunised vabatahtlikult heaks kiita.

On oluline märkida, et suunised ei asenda mingil moel poliitikat ega õigusnorme (neid käsitletakse kõrgetasemelise tehisintellekti eksperdirühma teises dokumendis – tehisintellekti poliitika ja investeerimissoovitused –, mis esitatakse 2019. aasta mais) ega takista nende kehtestamist. Lisaks tuleks suuniseid käsitada dokumendina, mida tuleb aja jooksul korrapäraselt muuta ja täiendada, et see suudaks tehnoloogia ja teadmiste pideva arenguga kaasas käia. Seepärast peaks käesolev dokument olema lähteks arutelule teemal „**Euroopas loodav usaldusväärne tehisintellekt**“.

Euroopa saab oma eetilist tehisintellekti käsitust levitada ainult siis, kui ta on üleilmsel tasandil konkurentsivõimeline, ning samas on **eetiline tehisintellekti käsitus on vastutustundliku konkurentsivõime alus**, sest see tekitab kasutajates usaldust ja lihtsustab tehisintellekti laiemat kasutuselevõttu. Käesolevate suunistega ei püüta lämmatada tehisintellekti innovatsiooni Euroopas, vaid kasutada eetikat kui inspiratsiooni, et töötada välja ainulaadne tehisintellekt, mis kaitseks ja oleks kasulik nii üksikisikute kui ka ühise heaolu seisukohast. See võimaldab Euroopal hõivata liidripositsiooni uuendusliku, turvalise ja eetilise tehisintellekti valdkonnas. Euroopa kodanikud saavad tehisintellekti eelistest täiel määral kasu lõigata vaid siis, kui me suudame tagada selle usaldusväärset.

Lõpetuseks on käesolevate suuniste eesmärk **edendada mõtisklusi ja arutelu** tehisintellekti eetilise raamistiku üle mitte ainult Euroopas, vaid **kogu maailmas**.

TEGEVUSJUHISED

Suuniste igas peatükis antakse juhiseid selle kohta, kuidas arendada välja usaldusväärne tehisintellekt. Need juhised on mõeldud kõigile sidusrühmadele, kes arendavad, juurutavad või kasutavad tehisintellekti. Peatükkide sisu on kokku võetud allpool.

I peatükk. Eetilise eesmärgi tagamise põhijuhised

- Tehisintellekt peab olema **inimkeskne**: selle arendamise, juurutamise ja kasutamise **eesmärgid peavad olema eetilised** ja rajanema põhiõigustel, ühiskondlikel väärtustel ja sellistel eetilistel põhimõtetel nagu *heategemine, kahju tegemisest hoidumine, inimese autonoomia, õiglus ja selgitatavus*. See on **usaldusväärse tehisintellekti** arendamiseks äärmiselt tähtis.
- Tehisintellekt peab tuginema põhiõigustele, eetilistele põhimõtetele ja väärtustele, et oleks võimalik prognoosida tehisintellekti võimalikku mõju inimestele ja ühisele heaolule. Eriti suurt tähelepanu tuleb pöörata olukordadele, mis puudutavad **haavatavaid rühmi**, näiteks lapsi, puuetega inimesi ja vähemusi, või olukordadele, kus esineb **võimu või teabe asümmeetria**, näiteks suhetes tööandjate ja töötajate või ettevõtjate ja tarbijate vahel.
- Teadvustada tuleb asjaolu, et ehkki tehisintellekt toob üksikisikutele ja ühiskonnale suurt kasu, võib sellel olla ka negatiivne mõju. Valvas tuleb olla kriitilise tähtsusega valdkondades.

II peatükk. Põhijuhised usaldusväärse tehisintellekti loomiseks

- **Usaldusväärse tehisintellekti arendamise nõudeid** tuleb arvesse võtta **juba kõige varasemas projekteerimisetapis**. Need nõuded on aruandekohustus, andmehaldus, universaalsaisain, tehisintellekti autonoomia reguleerimine (inimjärelevalve), mittediskrimineerimine, isikute autonoomia austamine, privaatsuse põhimõtte austamine, kindlus, ohutus ja läbipaistvus.
- Nende nõuete rakendamiseks tehisintellekti süsteemis tuleb kaaluda nii tehnilisi kui ka muid meetodeid. Lisaks tuleb neid nõudeid arvesse võtta süsteemi arendamiseks töötava meeskonna moodustamisel, süsteemi enda arendamisel, testimiskeskonna sisseseadmisel ja süsteemi potentsiaalsete rakenduste väljatöötamisel.

- **Sidusrühmadele** (kliendid, töötajad jne) tuleb anda selgelt ja ettevaatavalt **teavet** tehisintellekti süsteemi suutlikkuse ja selle piirangute kohta, et nad saaksid oma ootusi tegelikkusega kohendada. Sellega seoses on eriti tähtis tehisintellekti süsteemi **jälitatus**.
- Usaldusväärne tehisintellekt tuleb muuta **organisatsiooni kultuuri osaks** ning sidusrühmi tuleb teavitada sellest, kuidas on usaldusväärse tehisintellekti käsitus lõimitud tehisintellekti süsteemide väljatöötamise ja kasutamisse. Usaldusväärse tehisintellekti käsituse võib lisada ka organisatsiooni eetikakoodeksisse või käitumisjuhendisse.
- Tehisintellekti süsteemi väljatöötamise ja arendamise tuleb **kaasata sidusrühmad**. Lisaks sellele tuleb toote arendus-, rakendus- ja testimismeeskondade moodustamisel tagada **mitmekesisus**.
- Tehisintellekti süsteemid tuleb püüda välja töötada nii, et need oleksid **hõlpsasti kontrollitavad**, eriti kriitilistes olukordades. Süsteem tuleb kavandada selliselt, et oleks võimalik tagantjärele kindlaks teha iga otsuse aluseks olevad sisendid, andmed, eeltreenitud mudelid jne. Lisaks tuleb kindlaks määrata tehisintellekti süsteemi toimimise **selgitamise meetodid**.
- Juurutada tuleb **aruandluskohustuse täitmise** eriprotsess.
- Pakkuda tuleb **koolitust ja väljaõpet** ning tagada tuleb, et juhid, arendajad, kasutajad ja tööandjad oleksid usaldusväärsest tehisintellektist teadlikud ja sellealase koolituse saanud.
- Meeles tuleb pidada asjaolu, et eri eesmärkide vahel võivad tekkida põhimõttelised konfliktid (läbipaistvus võib avada ukse kuritarvitamisele, kallutatuse tuvastamine ja korrigeerimine võib sattuda vastuollu privaatsuse kaitse põhimõttega). Tehtud kompromissid tuleb dokumenteerida ja neist tuleb teatada.
- Soodustada tuleb teadusuuringuid ja innovatsiooni, et usaldusväärsele tehisintellektile esitatavaid nõudeid veelgi paremini täita.

III peatükk. Põhijuhised usaldusväärse tehisintellekti hindamiseks

- Kui tegeletakse usaldusväärse tehisintellekti arendamise, juurutamise või kasutamisega, tuleb kasutusele võtta **hindamisloend** ning kohendada seda vastavalt konkreetse süsteemi otstarbele.
- Meeles tuleb pidada, et hindamisloend **ei ole kunagi ammendav** ning et usaldusväärset tehisintellekti ei saa tagada lahtritesse ristikeste tegemise, vaid nõuete kindlaksmääramise, lahenduste hindamise ja saadavate tulemuste parandamise pideva protsessiga kogu tehisintellekti süsteemi elutsükli jooksul.

Käesolevad juhised on osa nägemusest, mis kätkeb inimesekeskse tehisintellekti käsitust ja Euroopa võimalust saada maailmas juhtivaks innovaatoriks eetilise, turvalise ja uuendusliku tehisintellekti valdkonnas. Juhiste eesmärk on hõlbustada ja võimaldada **Euroopas loodava usaldusväärse tehisintellekti** arendamist ja suurendada sellega Euroopa kodanike heaolu.