



Eiropas Komisijas

AUGSTA LĪMEŅA EKSPERTU GRUPA MĀKSLĪGĀ INTELEKTA JAUTĀJUMOS



PROJEKTS ĒTIKAS VADLĪNIJAS UZTICAMAM MĀKSLĪGAJAM INTELEKTAM KOPSAVILKUMS

Apspriešanās ar ieinteresētajām personām darba dokuments

ĒTIKAS VADLĪNIJU PROJEKTS

UZTICAMAM MĀKSLĪGAJAM INTELEKTAM



Augsta līmeņa ekspertu grupa mākslīgā intelekta jautājumos (ALEG MI)
Ētikas vadlīņu projekts uzticamam mākslīgajam intelektam

Eiropas Komisija
Komunikācijas ģenerāldirektorāts

Kontaktpersona *Nathalie Smuha* - ALEG MI koordinatore
E-pasts CNECT-HLG-AI@ec.europa.eu

Eiropas Komisija
B-1049 Brussels

Dokuments publiskots 2018. gada 18. decembrī, angļu valodā.

Šo darba dokumentu sagatavojusi Augsta līmeņa ekspertu grupa mākslīgā intelekta jautājumos (ALEG MI), neskarot tās locekļu individuālo nostāju par konkrētiem jautājumiem un neskarot dokumenta galīgo redakciju. Šis dokuments vēl tiks pilnveidots, un tā galīgā redakcija tiks prezentēta 2019. gada martā pēc apspriešanās ar ieinteresētajām personām Eiropas mākslīgā intelekta aliansē.

Ne Eiropas Komisija, ne jebkura persona, kas rīkojas Komisijas vārdā, nav atbildīga par turpmāk sniegtās informācijas izmantošanu. Par šā darba dokumenta saturu ir atbildīga tikai augsta līmeņa ekspertu grupa mākslīgā intelekta jautājumos (ALEG MI). Lai gan Komisijas dienestu darbinieki palīdzēja sagatavot šīs vadlīnijas, šajā dokumentā paustie uzskati atspoguļo ALEG MI viedokli, un to nekādā gadījumā nevar uzskatīt par Eiropas Komisijas oficiālu nostāju. Šis ir augsta līmeņa ekspertu grupas mākslīgā intelekta jautājumos pirmā paziņojuma projekts. Tā galīgā redakcija tiks iesniegta Komisijai 2019. gada martā. Otrā paziņojuma galīgā redakcija — mākslīgā intelekta politika un ieguldījumu rekomendācijas — tiks iesniegta 2019. gada vidū.

Plašāka informācija par augsta līmeņa ekspertu grupu mākslīgā intelekta jautājumos ir pieejama tiešsaistē (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>). Eiropas Komisijas dokumentu atkalizmantošanas politika ir noteikta ar Lēmumu 2011/833/ES (OV L 330, 14.12.2011., 39. lpp.). Lai varētu izmantot vai reproducēt fotoattēlus vai citus materiālus, uz kuriem ES nav autortiesību, atļauja jāprasa tieši autortiesību īpašniekam.

KOPSAVILKUMS

Šis darba dokuments ir mākslīgā intelekta ētikas vadlīniju projekts, ko izstrādājusi Eiropas Komisijas augsta līmeņa ekspertu grupa mākslīgā intelekta jautājumos (ALEG MI) un kura galīgā redakcija būs gatava 2019. gada martā.

Mākslīgais intelekts ir viens no mūsu laika pārveidojošākajiem spēkiem, un tas droši vien mainīs sabiedrības struktūru. Tas dod lielisku iespēju palielināt labklājību un izaugsmi, ko Eiropai būtu jācenšas sasniegt. Pēdējo desmit gadu laikā ir gūti lieli panākumi, jo bija pieejams ļoti liels digitālo datu daudzums, jaudīga skaitļošanas tehnika un tādu mākslīgā intelekta paņēmieni attīstība kā mašīnmācīšanās. Katru dienu mūsu dzīves kvalitāti uzlabo galvenās mākslīgā intelekta sekmētās attīstības iespējas attiecībā uz autonomajiem transportlīdzekļiem, veselības aprūpi, mājokļu/pakalpojumu robotiem, izglītību vai kiberdrošību. Turklāt mākslīgajam intelektam ir ļoti liela nozīme, lai risinātu daudzas no lielajām problēmām, ar kurām saskaras pasaule, piemēram, veselība un labklājība pasaulē, klimata pārmaiņas, uzticamas juridiskas un demokrātiskas sistēmas un citas, kuras iekļautas Apvienoto Nāciju Organizācijas ilgspējīgas attīstības mērķos.

Mākslīgais intelekts spēj radīt milzīgas priekšrocības iedzīvotājiem un sabiedrībai, tomēr tas rada arī zināmu risku, kas būtu pienācīgi jāpārvalda. Visumā mākslīgā intelekta ieguvumi ir lielāki par risku, tāpēc mums ir jāizvēlas rīcība, kas pēc iespējas **palielinātu mākslīgā intelekta sniegtās priekšrocības, vienlaikus samazinot risku**. Lai garantētu, ka darbojamies pareizajā virzienā, **ir vajadzīga uz cilvēku vērsta pieeja mākslīgajam intelektam**, un jācenšas vienmēr paturēt prātā, ka mākslīgā intelekta izstrāde un izmantošana nav līdzeklis pats par sevi, bet gan mērķis palielināt cilvēku labklājību. **Uzticams mākslīgais intelekts būs mūsu vadzvaigzne**, jo cilvēki varēs droši un pilnībā izmantot mākslīgā intelekta priekšrocības tikai tad, ja tie uzticēsies šai tehnoloģijai.

Uzticamam mākslīgajam intelektam ir **divas sastāvdaļas**: (1) tam ir jāievēro pamattiesības, piemērojamie noteikumi un pamatprincipi un vērtības, nodrošinot “**ētisku mērķi**”, un (2) tam jābūt **tehnikai robustam** un uzticamam, jo tehnoloģisko prasmju trūkums var radīt netīšu kaitējumu, pat ja nodomi ir labi.

Tādēļ šajās vadlīnijās ir noteikta **uzticama mākslīgā intelekta struktūra**:

- **I nodaļa** attiecas uz **MI ētiskā mērķa nodrošināšanu** un nosaka pamattiesības, principus un vērtības, kas tam būtu jāievēro.
- Pamatojoties uz šiem principiem, **II nodaļā** ir sniegti **norādījumi par uzticama MI īstenošanu** attiecībā uz ētisko mērķi un arī uz tehnisko robustumu. Šim nolūkam ir uzskaitītas prasības uzticamam MI un piedāvāts pārskats par tehniskām un netehniskām metodēm, kuras var izmantot tā īstenošanai.
- **III nodaļā** šīs prasības izkārtotas **operatīvā sistēmā**, sniedzot konkrētu, bet nepilnīgu novērtējumu sarakstu par uzticamu MI. Pēc tam šo sarakstu pielāgo konkrētiem izmantošanas gadījumiem.

Tādējādi atšķirībā no citiem dokumentiem, kas attiecas uz ētisku mākslīgo intelektu, šo vadlīniju mērķis nav sniegt vēl vienu sarakstu ar mākslīgā intelekta pamatvērtībām un pamatprincipiem, bet piedāvāt norādījumus par to konkrētu īstenošanu un pielietošanu mākslīgā intelekta sistēmās. Šādas norādes tiek sniegtas trijos abstrakcijas līmeņos — no visabstraktākajām I nodaļā (pamattiesības, principi un vērtības) līdz viskonkrētākajām III nodaļā (novērtējuma saraksts).

Vadlīnijas ir adresētas visām **attiecinīgajām ieinteresētajām personām, kas izstrādā, izvieta vai izmanto MI**, tostarp uzņēmumiem, organizācijām, pētniekiem, publiskajiem pakalpojumiem, iestādēm, privātpersonām vai citām struktūrām. Šo vadlīniju galīgajā redakcijā tiks paredzēts mehānisms, lai ieinteresētās personas varētu tās brīvprātīgi apstiprināt.

Svarīgi ir tas, ka vadlīnijas nav paredzētas kā politikas veidošanas vai tiesiska regulējuma aizvietotājs (par to tiks runāts ALEG MI otrajā ziņojumā “Politikas un investīciju ieteikumi”, kas jāiesniedz 2019. gada maijā), un to mērķis nav atturēt no MI ieviešanas. Turklāt vadlīnijas būtu jāuzskata par dinamisku dokumentu, kas laika gaitā ir regulāri jāatjaunina, lai nodrošinātu to nepārtrauktu atbilstību tehnoloģiju un mūsu zināšanu attīstībai. Tāpēc šim dokumentam vajadzētu būt par izejas punktu diskusijai **“Uzticams Eiropas mākslīgais intelekts”**.

Eiropa var izcelt tās ētisko pieeju mākslīgajam intelektam tikai tad, ja tas ir konkurētspējīgs pasaules mērogā, un **ētiska pieeja mākslīgā intelekta jomā ir būtiska, lai nodrošinātu atbildīgu konkurētspēju**, jo tā radīs lietotāju uzticēšanos un veicinās mākslīgā intelekta plašāku izmantošanu. Šīs vadlīnijas nav paredzētas, lai aplāpētu MI inovāciju Eiropā, bet to mērķis ir izmantot ētiku kā iedvesmas avotu, lai izveidotu unikālu mākslīgā intelekta zīmolu, kura mērķis ir aizsargāt un sniegt labumu gan privātpersonām, gan visai sabiedrībai. Tādējādi Eiropa varēs pozicionēt sevi kā progresīva, droša un ētiska mākslīgā intelekta līderi. Eiropas iedzīvotāji varēs pilnībā izmantot mākslīgā intelekta priekšrocības tikai tad, ja būs nodrošināta tā uzticamība.

Visbeidzot, arī ārpus Eiropas šo vadlīniju mērķis ir **veicināt pārdomas un diskusijas** par ētisku regulējumu mākslīgā intelekta jomā **pasaules mērogā**.

PĀRSKATS PAR VADLĪNIJĀM

Katrā vadlīniju nodaļā ir sniegti norādījumi par to, kā panākt uzticamu mākslīgo intelektu; tās adresētas visām attiecīgajām ieinteresētajām personām, kuras izstrādā, izvieto vai izmanto MI, un apkopotas turpmāk.

I nodaļa. Galvenie norādījumi par to, kā nodrošināt ētisko mērķi.

- Nodrošināt, ka mākslīgais intelekts ir **vērsts uz cilvēku**: mākslīgais intelekts jāizstrādā, jāizvieto un jālieto ar **“ētisku mērķi”**, kas balstās uz pamattiesībām, sabiedrības vērtībām un ētikas principiem un to atspoguļo, piemēram, *labdarība* (dari labu), *ļautuma nenodarišana* (nedari sliktu), *cilvēka autonomijas ievērošana*, *taisnīgums* un *izskaidrojamība*. Tas ir ļoti svarīgi, lai strādātu pie **uzticama MI radīšanas**.
- Lai perspektīvi novērtētu mākslīgā intelekta iespējamo ietekmi uz cilvēkiem un sabiedrību, pajauties uz pamattiesībām, ētikas principiem un vērtībām. **Īpašu uzmanību** pievērst situācijām, kurās iesaistītas **mazāk aizsargātas grupas**, piemēram, bērni, personas ar invaliditāti vai minoritātes, vai situācijām, kurās pastāv **varas vai informācijas asimetrija**, piemēram, starp darba devējiem un darba ņēmējiem vai uzņēmumiem un patērētājiem.
- Atzīt un apzināties, ka mākslīgais intelekts, vienlaikus sniedzot būtiskus ieguvumus privātpersonām un sabiedrībai, var arī izraisīt negatīvu ietekmi. Saglabāt modrību attiecībā uz problemātiskām jomām.

II nodaļa. Galvenie norādījumi uzticama MI īstenošanai.

- **Izstrādes posma pašā sākumā iekļaut MI uzticamības prasības**: pārskatatbildība, datu pārvaldība, projektēšana visu cilvēku labā, mākslīgā intelekta pārvaldība (cilvēka uzraudzība), nediskriminācija, cieņa pret cilvēka autonomiju, cieņa pret privātumu, robustums, drošība, pārredzamība.
- Apsvērt tehniskus un netehniskus paņēmienus, kas nodrošinātu šo prasību īstenošanu mākslīgā intelekta sistēmā. Turklāt ņemt vērā šīs prasības, veidojot komandu darbam pie sistēmas, pašu sistēmu, testēšanas vidi un sistēmas iespējamus lietojumus.
- Skaidri un proaktīvi sniegt **informāciju ieinteresētajām personām** (klientiem, darba ņēmējiem u. c.)

par mākslīgā intelekta sistēmas spējām un ierobežojumiem, ļaujot tiem veidot reālistiskas prognozes. Šajā ziņā svarīga ir mākslīgā intelekta sistēmas **izsekojamības** nodrošināšana.

- Padarīt uzticamu mākslīgo intelektu par **daļu no organizācijas kultūras** un sniegt informāciju ieinteresētajām personām par to, kā uzticams mākslīgais intelekts ir iekļauts mākslīgā intelekta sistēmu izstrādē un izmantošanā. Uzticamu mākslīgo intelektu var iekļaut arī organizāciju profesionālās ētikas hartās vai rīcības kodeksos.
- Nodrošināt **ieinteresēto personu līdzdalību un iekļaušanu** mākslīgā intelekta sistēmas projektēšanā un izstrādē. Turklāt nodrošināt **daudzveidību**, veidojot komandas, kuras izstrādās, īsteno un testēs galarezultātu.
- Censties **atvieglot mākslīgā intelekta sistēmu pārbaudāmību**, jo īpaši kritiskos apstākļos vai situācijās. Cik vien iespējams sistēmu izveidot tā, lai būtu iespējams izsekot atsevišķus lēmumus attiecībā uz dažādiem ievaddatiem, iepriekš izveidotiem modeļiem utt. Izstrādāt arī MI sistēmas **izskaidrošanas metodes**.
- Nodrošināt īpašu **pārskatatbildības pārvaldības** procesu.
- Paredzēt **apmācību un izglītību** un nodrošināt, ka vadītāji, izstrādātāji, lietotāji un darba devēji ir informēti un apmācīti par uzticamu MI.
- Apzināties, ka starp dažādiem mērķiem var pastāvēt būtiska spriedze (pārredzamība var dot iespēju ļaunprātīgai izmantošanai; konstatējot un labojot neobjektivitāti, var nonākt pretrunā ar privātuma aizsardzību). Komunicēt un dokumentēt šos kompromisus.
- Veicināt pētniecību un inovāciju, lai sekmētu uzticama mākslīgā intelekta prasību ievērošanu.

III nodaļa. Galvenie norādījumi uzticama MI novērtēšanai:

- Izstrādājot, izvietojot vai izmantojot mākslīgo intelektu, pieņemt uzticama mākslīgā intelekta **novērtēšanas sarakstu** un pielāgot to konkrētajam sistēmas izmantošanas gadījumam.
- Paturēt prātā, ka novērtējuma saraksts **nekad nebūs pilnīgs** un ka mākslīgā intelekta uzticamības nodrošināšana nav ķeksīšu ievilkšana, bet gan pastāvīgs process, lai noteiktu prasības, novērtētu risinājumus un nodrošinātu labākus rezultātus visā mākslīgā intelekta sistēmas darbības ciklā.

Šie norādījumi ir daļa no redzējuma, kas ietver uz cilvēku vērstu pieeju mākslīgajam intelektam, kurš ļaus Eiropai kļūt par pasaules mēroga vadošu novatoru ētiska, droša un progresīva MI jomā. Tie mēģina atvieglināt un padarīt iespējamu **“uzticamu Eiropas mākslīgo intelektu”**, kas uzlabos Eiropas iedzīvotāju labklājību.