



Euroopan komission

**TEKOÄLYÄ KÄSITTELEVÄ
KORKEAN TASON ASIANTUNTIJARYHMÄ**



**LUONNOS
LUOTETTAVAA TEKOÄLYÄ KOSKEVAT
EETTISET OHJEET
TIIVISTELMÄ**

Työasiakirja sidosryhmien kuulemista varten

Bryssel, 18. joulukuuta 2018

LUONNOS: LUOTETTAVAA TEKÖÄLYÄ KOSKEVAT EETTISET OHJEET



Tekoälyä käsittelevä korkean tason asiantuntijaryhmä
Luonnos: Luotettavaa tekoälyä koskevat eettiset ohjeet

Euroopan komissio
Viestinnän pääosasto

Yhteyshenkilö Nathalie Smuha - koordinaattori, AI HLEG
Sähköposti CNECT-HLG-AI@ec.europa.eu

Euroopan komissio
B-1049 Bryssel

Asiakirja julkaistiin 18. joulukuuta 2018 englanniksi.

AI HLEG on laatinut tämän työasiakirjan rajoittamatta jäsentensä henkilökohtaista kantaa tiettyihin kohtiin ja vaikuttamatta asiakirjan lopulliseen versioon. Tätä asiakirjaa kehitetään edelleen, ja sen lopullinen versio esitetään maaliskuussa 2019 eurooppalaisen tekoälyallianssin kautta järjestetyn sidosryhmien kuulemisen jälkeen.

Euroopan komissio ja sen nimissä toimivat henkilöt eivät vastaa siitä, miten tässä asiakirjassa olevia tietoja käytetään. Tämän työasiakirjan sisältö on yksinomaan tekoälyä käsittelevän korkean tason asiantuntijaryhmän (AI HLEG) vastuulla. Vaikka komission yksiköiden henkilöstö on edesauttanut näiden ohjeiden laatimista, tässä asiakirjassa esitetyt näkemykset edustavat AI HLEG:n kantaa eikä niitä voida missään olosuhteissa katsoa Euroopan komission viralliseksi kannanotoksi. Tässä asiakirjassa esitetään AI HLEG -ryhmän työn ensimmäiset alustavat tulokset. Sen lopullinen versio esitetään komissiolle maaliskuussa 2019. Lopullinen versio toisesta asiakirjasta, joka käsittelee tekoälypolitiikka- ja investointisuosituksia, esitetään vuoden 2019 puolivälissä.

Lisätietoa tekoälyä käsittelevästä korkean tason asiantuntijaryhmästä löytyy verkosta osoitteesta <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>.

Euroopan komission asiakirjojen uudelleenkäytöstä säädetään päätöksessä 2011/833/EU (EUVL L 330, 14.12.2011, s. 39). Sellaisten valokuvien tai materiaalin käyttöön, johon EU:lla ei ole tekijänoikeutta, on pyydettävä lupa suoraan tekijänoikeuden haltijalta.

TIIVISTELMÄ

Tämä työasiakirja on Euroopan komission tekoälyä käsittelevän korkean tason asiantuntijaryhmän (AI HLEG) luonnos tekoälyä koskeviksi eettisiksi ohjeiksi, joiden lopullisen version on määrä valmistua maaliskuussa 2019.

Tekoäly on yksi aikamme suurimmista muutosvoimista, ja sillä on väistämätön vaikutus yhteiskuntamme rakenteeseen. Sillä on erinomaiset edellytykset lisätä vaurautta ja kasvua, ja Euroopan ei tule jättää näitä mahdollisuuksia käyttämättä. Viime vuosikymmenen aikana on otettu merkittäviä edistysaskeleita saatavilla olevien valtaviin digitaalisten tietomäärien, tehokkaiden laskenta-arkkitehtuurien ja tekoälytekniikassa tapahtuneen edistyksen kuten koneoppimisen ansiosta. Tekoälyn mahdollistama merkittävä kehitys itseohjautuvissa ajoneuvoissa, terveydenhuollossa, koti- ja palveluroboteissa, koulutuksessa ja kyberturvallisuudessa parantaa elämänlaatuamme päivittäin. Tekoäly auttaa meitä vastaamaan moniin nyky maailman haasteisiin, jotka liittyvät globaaliin terveyteen ja hyvinvointiin, ilmastonmuutokseen, luotettaviin oikeusjärjestelmiin ja demokraattisiin järjestelmiin sekä muihin Yhdistyneiden kansakuntien kestävä kehityksen tavoitteisiin.

Vaikka tekoälystä onkin kiistatonta hyötyä yksilöille ja yhteiskunnalle, siihen liittyy myös tiettyjä riskejä, joita olisi hallittava asianmukaisesti. Koska tekoälyn hyödyt ovat kokonaisuudessaan sen riskejä suuremmat, meidän tulee pyrkiä **maksimoimaan tekoälyn hyödyt ja minimoimaan sen riskit**. Tekoälyn kehittämisessä on noudatettava **ihmiskeskeistä lähestymistapaa**: tekoälyn kehittäminen ja käyttö eivät ole itseisarvo vaan keinoja parantaa ihmisten hyvinvointia. Ensisijaisena tavoitteena on **tekoälyn luotettavuus**, sillä tekoälyn hyödyistä saa kaiken irti ainoastaan silloin, kun tekoälyteknologiaan voi luottaa.

Luotettavaan tekoälyyn liittyy **kaksi osa-aluetta**: (1) sen olisi kunnioitettava perusoikeuksia, sovellettavaa lainsäädäntöä sekä keskeisiä periaatteita ja arvoja, millä taataan sen ”**eettinen tarkoitusperä**”, ja (2) sen olisi oltava **teknisesti luotettava**, sillä vaikka tarkoitus olisi hyvä, teknologian kömpelyys voi aiheuttaa tahatonta haittaa.

Nämä ohjeet muodostavat **luotettavan tekoälyn puitteet**:

- **Luku I** käsittelee **tekoälyn eettisen tarkoitusperän varmistamista** ja esittää perusoikeudet, periaatteet ja arvot, joita tekoälyn tulisi noudattaa.
- **Luku II** antaa edellisen luvun periaatteiden perusteella **ohjeita luotettavan tekoälyn käytännön toteutukseen** ja käsittelee sekä eettistä tarkoitusperää että teknistä luotettavuutta. Tämä toteutetaan luettelemalla luotettavaa tekoälyä koskevat vaatimukset ja esittämällä yleiskatsaus teknisistä ja muista menetelmistä, joita voidaan käyttää nämä vaatimukset täyttävän tekoälyn toteuttamiseksi.
- **Luku III konkretisoi** edellisen luvun vaatimukset esittämällä kattavan mutta ei tyhjentävän luettelon kysymyksistä, jotka tulisi ottaa huomioon luotettavan tekoälyn arvioinnissa. Tätä luetteloa mukautetaan kunkin käyttötapauksen mukaan.

Toisin kuin muissa eettistä tekoälyä käsittelevissä asiakirjoissa, näissä ohjeissa siis ei pyritä esittämään jälleen uutta luetteloa tekoälyn keskeisistä arvoista ja periaatteista vaan pikemminkin antamaan ohjeita sen konkreettiseen käyttöönottoon tekoälyjärjestelmissä. Ohjeet konkretisoituvat luku luvulta: luku I (perusoikeudet, periaatteet ja arvot) on vielä teoreettinen mutta luku III (arviointiluettelo) jo lähellä käytäntöä.

Ohjeet on osoitettu kaikille **asianomaisille sidosryhmille, jotka osallistuvat tekoälyn kehitykseen, käyttöönottoon tai käyttöön**, mukaan lukien yritykset, organisaatiot, tutkijat, julkiset palvelut, toimielimet, yksityishenkilöt ja muut tahot. Näiden ohjeiden lopullisessa versiossa esitetään mekanismi,

jolla sidosryhmät voivat vapaaehtoisesti tukea niitä.

On tärkeää huomata, että näiden ohjeiden tarkoituksena ei ole korvata poliittista päätöksentekoa tai sääntelyä (joita käsitellään AI HLEG:n toisessa, politiikka- ja investointisuosituksia käsittelevässä asiakirjassa, jonka on määrä valmistua toukokuussa 2019) eikä olla esteenä niille. Näitä ohjeita ei tule nähdä kiveen hakattuina lopullisina totuuksina vaan niitä on päivitettävä säännöllisesti teknologian ja tietämyksemme kehittyessä. Tämän asiakirjan tavoitteena on olla lähtökohta keskustelulle ”**luotettavasta eurooppalaisesta tekoälystä**”.

Eettinen lähestymistapa tekoälyyn on vastuullisen kilpailukyvyn edellytys, sillä se lisää käyttäjien luottamusta ja helpottaa tekoälyn laajempaa käyttöönottoa. Euroopan on oltava maailman mittapuulla kilpailukykyinen toimija voidakseen levittää eettistä lähestymistapaansa. Näiden ohjeiden tavoitteena ei ole tukahduttaa tekoälyyn liittyviä innovaatioita Euroopassa vaan pyrkiä tekemään etiikasta eräänlainen tavaramerkki tekoälyn kehittämiselle, jotta tekoäly suojelisi ja hyödyttäisi sekä yksilöitä että yhteistä etua. Tämä mahdollistaa Euroopalle johtoaseman huipputasoisien, turvallisten ja eettisten tekoälyn alalla. Vasta kun luotettavuus on turvattu, eurooppalaiset pääsevät kunnolla hyötymään tekoälystä.

Euroopan ulkopuolella näiden ohjeiden tavoitteena on myös **edistää maailmanlaajuisia pohdintaa ja keskustelua** tekoälyn eettisistä puitteista.

SUUNTAVIIVAT

Kukin tämän asiakirjan luku tarjoaa luotettavan tekoälyn tueksi ohjeita, jotka on osoitettu kaikille tekoälyn kehitykseen, käyttöönottoon tai käyttöön osallistuville asianomaisille sidosryhmille. Lukujen sisältö lyhyesti alla:

Luku I: Ohjeet eettisen tarkoituksiperän varmistamiseen

- Varmista tekoälyn **ihmiskeskeisyys**: Tekoälyn kehittämisessä, käyttöönotossa ja käytössä tulisi pyrkiä kohti ”**eettistä tarkoituksiperää**”, joka perustuu perusoikeuksiin, yhteiskunnallisiin arvoihin sekä *hyvään pyrkimisen, vahinkojen välttämisen, ihmisen itsemääräämisoikeuden, oikeudenmukaisuuden ja selitettävyyden (explicability)* eettisiin periaatteisiin. Tämä on ratkaisevan tärkeää **luotettavaa tekoälyä** kohti pyrkiessä.
- Arvioi tekoälyn mahdollisia vaikutuksia ihmisiin ja yhteiseen etuun perusoikeuksien, eettisten periaatteiden ja arvojen perusteella. **Erityistä huomiota** tulisi kiinnittää tilanteisiin, joihin liittyy **haavoittuvia ryhmiä**, kuten lapsia, vammaisia tai vähemmistöjä, tai joissa vallitsee **vallan tai tiedon epätasapaino**, kuten työnantajien ja työntekijöiden tai yritysten ja kuluttajien välillä.
- Pidä mielessä, että vaikka tekoäly voi tuottaa merkittäviä etuja ihmisille ja yhteiskunnalle, sillä voi olla myös negatiivisia vaikutuksia. Keskeiset huolenaiheet on hyvä pitää mielessä.

Luku II: Ohjeet luotettavan tekoälyn käytännön toteutukseen

- Ota **luotettavaa tekoälyä koskevat vaatimukset** osaksi suunnittelua **mahdollisimman aikaisessa vaiheessa**: Vastuuvollisuus, datan hallinta, kaikkia palveleva suunnittelu, tekoälyn autonomian hallinta (ihmisen suorittama valvonta), syrjimättömyys, ihmisen itsemääräämisoikeuden kunnioittaminen, yksityisyyden kunnioittaminen, luotettavuus, turvallisuus ja läpinäkyvyys.
- Pohdi teknisiä ja muita menetelmiä, joilla nämä vaatimukset voidaan ottaa huomioon tekoälyjärjestelmässä. Vaatimukset on pidettävä mielessä myös koottaessa ryhmää, joka työstää itse järjestelmää, sen testiympäristöä ja mahdollisia sovelluksia.
- Tarjoa proaktiivisesti selkeää **tietoa sidosryhmille** (asiakkaille, työntekijöille jne.) tekoälyjärjestelmän valmiuksista ja rajoituksista. Näin sidosryhmien odotukset säilyvät realistisina.

Tekoälyjärjestelmän **jäljitettävyys** on tämän kannalta ensiarvoisen tärkeää.

- Tee luotettavasta tekoälystä **osa organisaation kulttuuria** ja kerro sidosryhmille, kuinka luotettavuus on toteutettu tekoälyjärjestelmien suunnittelussa ja käytössä. Luotettava tekoäly voidaan sisällyttää myös organisaatioiden eettisiin sääntöihin tai käytännesääntöihin.
- Varmista **sidosryhmien osallistuminen ja osallistaminen** tekoälyjärjestelmän suunnitteluun ja kehitykseen. Huomioi **monimuotoisuus**, kun muodostat järjestelmää kehittäviä, implementoivia ja testaavia ryhmiä.
- Pyri helpottamaan tekoälyjärjestelmien **tarkastettavuutta** erityisesti kriittisissä tilanteissa. Suunnittele järjestelmä mahdollisuuksien mukaan sellaiseksi, että sen päätökset on mahdollista jäljittää esim. siihen syötettyihin tietoihin tai koulutettuihin malleihin. Määritä myös tekoälyjärjestelmän **selitysmallit**.
- Varmista, että **vastuullisuuden hallinnalle** on tarkka prosessi.
- Järjestä **koulutusta** ja varmista, että esimiehet, kehittäjät, käyttäjät ja työntekijät tietävät luotettavasta tekoälystä ja saavat sitä koskevaa koulutusta.
- Huomioi, että eri tavoitteiden välillä voi olla perustavanlaatuisia jännitteitä (läpinäkyvyys voi mahdollistaa väärinkäytön; vääristymien tunnistaminen ja korjaaminen voi olla ristiriidassa yksityisyyden suojan kanssa). Kirjaa tavoitteiden välillä tehdyt valinnat ylös ja tiedota niistä.
- Edistä tutkimusta ja innovointia, jotka vievät luotettavan tekoälyn kehitystä kohti sille asetettuja tavoitteita.

Luku III: Ohjeet luotettavan tekoälyn arvioimiseen

- Laadi tekoälyn suunnittelu-, käyttöönotto- tai käyttövaiheessa **luettelo kysymyksistä, jotka tulee ottaa huomioon luotettavan tekoälyn arvioinnissa**, ja mukauta sitä järjestelmän käyttötilanteen mukaan.
- Pidä mielessä, että arviointiluettelo **ei ole koskaan tyhjentävä** ja että tekoälyn luotettavuutta ei takaa ruutujen rastittaminen. Luotettavuuden varmistaminen on jatkuva prosessi, jossa tunnistetaan vaatimuksia, arvioidaan ratkaisuja ja pyritään yhä parempiin tuloksiin tekoälyjärjestelmän koko elinkaaren ajan.

Nämä ohjeet ovat osa visiota ihmiskeskeisestä lähestymistavasta tekoälyyn. Tämä lähestymistapa mahdollistaa Euroopalle johtavan innovaatioaseman eettisen, turvallisen ja huipputasaisen tekoälyn kehityksessä. Näillä ohjeilla pyrimme edistämään **”luotettavan eurooppalaisen tekoälyn”** kehittämistä ja sitä kautta parantamaan eurooppalaisten hyvinvointia.