



Bryssel 8.4.2019
COM(2019) 168 final

**KOMISSION TIEDONANTO EUROOPAN PARLAMENTILLE, NEUVOSTOLLE,
EUROOPAN TALOUS- JA SOSIAALIKOMITEALLE JA ALUEIDEN
KOMITEALLE**

Luottamuksen rakentaminen ihmiskeskeiseen tekoälyyn

1. JOHDANTO – EUROOPAN TEKOÄLYSTRATEGIA

Tekoäly voi muuttaa maailmaa paremmaksi: se voi parantaa terveydenhuoltoa, vähentää energiakulutusta, tehdä autoista turvallisempia ja antaa viljelijöille mahdollisuuden käyttää vettä ja luonnonvaroja tehokkaammin. Tekoälyn avulla voidaan ennustaa ympäristöön ja ilmastoon liittyviä muutoksia ja parantaa taloudellista riskinhallintaa. Se tarjoaa käyttöön välineitä, joiden avulla voidaan valmistaa tarpeisiimme sovitettuja tuotteita vähentäen jätteiden syntymistä. Tekoäly voi myös auttaa petosten ja kyberturvallisuusuhkien havaitsemisessa ja antaa lainvalvontaviranomaisille mahdollisuuden torjua rikollisuutta tehokkaammin.

Tekoäly voi hyödyttää koko yhteiskuntaa ja taloutta. Se on strateginen teknologia, jota kehitetään ja otetaan nopeassa tahdissa käyttöön kaikkialla maailmassa. Se kuitenkin tuo myös mukanaan uusia haasteita työn tulevaisuuden kannalta ja nostaa esiin oikeudellisia ja eettisiä kysymyksiä.

Näihin haasteisiin vastaamiseksi ja tekoälyn tarjoamien mahdollisuuksien hyödyntämiseksi mahdollisimman hyvin komissio julkaisi huhtikuussa 2018 erityisen eurooppalaisen strategian¹. Strategia asettaa ihmiset tekoälyn kehittämisen keskipisteeseen (**ihmiskeskeinen tekoäly**). Kyseessä on kolmiosainen lähestymistapa, jonka tavoitteena on tukea EU:n teknologista ja teollista kapasiteettia ja tekoälyn käyttöönottoa koko taloudessa, valmistautua sosioekonomisiin muutoksiin ja varmistaa asianmukaisen eettiset ja oikeudelliset puitteet.

Tekoälystrategian läpiviemiseksi **komissio on yhdessä jäsenvaltioiden kanssa laatinut koordinoitun tekoälysuunnitelman²**, jonka avulla voidaan luoda synergioita, koota dataa – monien tekoälysovellusten raaka-ainetta – ja lisätä yhteisiä investointeja. Komissio esitti suunnitelman joulukuussa 2018. Tarkoituksena on edistää yhteistyötä rajojen yli ja saada kaikki toimijat lisäämään julkisia ja yksityisiä investointeja **vähintään 20 miljardiin euroon** vuodessa seuraavan vuosikymmenen aikana³. Komissio on kaksinkertaistanut tekoälyyn kohdistetut investointinsa Horisontti 2020 -ohjelmassa ja aikoo investoida vuosittain miljardi euroa Euroopan horisontti -ohjelmasta ja Digitaalinen Eurooppa -ohjelmasta. Pyrkimyksenä on tukea muun muassa yhteisiä data-avaruuksia terveyden, liikenteen ja valmistuksen aloilla, laajamittaisia kokeilujärjestelyjä, kuten älykkäitä sairaaloita ja automatisoitujen ajoneuvojen infrastruktuureja, sekä strategista tutkimuslinjausta.

Tällaisen yhteisen strategisen tutkimus-, innovointi- ja käyttöönotto-ohjelman toteuttamiseksi komissio on tehostanut **vuoropuhelua kaikkien sidosryhmien kanssa**. Edustettuina ovat olleet niin teollisuus, tutkimuslaitokset kuin viranomaisetkin. Uusi Digitaalinen Eurooppa -ohjelma on myös avainasemassa, jotta tekoäly voidaan tuoda pienten ja keskiuurten yritysten saataville kaikissa jäsenvaltioissa digi-innovaatiokeskittymien, pitkälle vietyjen testaus- ja kokeilujärjestelyjen, data-avaruuksien ja koulutusohjelmien kautta.

Euroopan eettinen lähestymistapa tekoälyyn vahvistaa kansalaisten luottamusta digitaaliseen kehitykseen, mitä tukee Euroopan maine turvallisten ja korkealaatuisten tuotteiden valmistajana. Päämääränä on tältä pohjalta luoda kilpailuetu eurooppalaisille tekoäly-yrityksille. Tämän tiedonannon tarkoituksena on käynnistää kattava pilottivaihe, johon

¹ COM(2018) 237.

² COM(2018) 795.

³ Tämän tavoitteen saavuttamiseksi komissio on ehdottanut seuraavan ohjelmakauden 2021–2027 yhteydessä, että unioni myöntää Euroopan horisontti -ohjelmasta ja Digitaalinen Eurooppa -ohjelmasta vuosittain vähintään miljardin euron rahoituksen investointeina tekoälyyn.

sidosryhmät osallistuvat mahdollisimman laajalti, jotta voidaan testata tekoälyn kehittämiseen ja käyttöön sovellettavan eettisen ohjeiston käytännön täytäntöönpanoa.

2. LUOTTAMUKSEN RAKENTAMINEN IHMISKESKEISTÄ TEKÖÄLYÄ KOHTAAN

Eurooppalainen tekoälystrategia ja koordinoitu suunnitelma tekevät selväksi, että **luottamus on edellytys ihmiskeskeisen lähestymistavan varmistamiselle tekoälyä varten**: tekoäly ei ole itsetarkoitus vaan väline, joka palvelee ihmisiä ja jonka perimmäisenä tavoitteena on lisätä ihmisten hyvinvointia. Tämän saavuttamiseksi olisi varmistettava, että **tekoälyn luotettavuus on taattu**. Yhteiskuntiemme perustana olevat arvot on tarpeen niveltää täysin mukaan tekoälyn kehitykseen.

Unionin perustana olevat arvot ovat **ihmisarvon kunnioittaminen, vapaus, demokratia, tasa-arvo, oikeusvaltio ja ihmisoikeuksien kunnioittaminen**, vähemmistöihin kuuluvien oikeudet mukaan luettuina⁴. Nämä arvot ovat yhteisiä kaikkien jäsenvaltioiden yhteiskunnille, joissa vallitsevat moniarvoisuus, syrjimättömyys, suvaitsevaisuus, oikeudenmukaisuus, yhteisvastuu ja tasa-arvo. Lisäksi **EU:n perusoikeuskirjaan** on koottu yhdeksi tekstiksi ihmisten EU:ssa nauttimat yksilön oikeudet, kansalaisyhteiskunnan oikeudet sekä poliittiset, taloudelliset ja sosiaaliset oikeudet.

EU:lla on **vahva sääntelykehys**, joka asettaa maailmanlaajuisen standardin ihmiskeskeiselle tekoälylle. Yleinen tietosuojajärjestelmä varmistaa korkeatasoisen suojan henkilötiedoille ja edellyttää toimenpiteitä sisäänrakennetun ja oletusarvoisen tietosuojan varmistamiseksi⁵. Muiden kuin henkilötietojen vapaata liikkuvuutta koskevalla asetuksella poistetaan muiden kuin henkilötietojen vapaan liikkuvuuden esteet ja varmistetaan, että kaikkiin luokkiin kuuluvia tietoja voidaan käsitellä kaikkialla Euroopassa. Hiljattain hyväksytty kyberturvallisuusasetus auttaa vahvistamaan luottamusta verkkomaailmaan, ja ehdotetulla sähköisen viestinnän tietosuojajärjestelmällä⁶ pyritään tähän samaan tavoitteeseen.

Tekoäly tuo kuitenkin uusia haasteita, koska sen avulla koneet voivat ”oppia” ja tehdä ja panna täytäntöön päätöksiä ilman ihmisen osallistumista. Ennen pitkää tämäntyyppisestä toiminnasta tulee standardi monentyyppisissä tavaroissa ja palveluissa, olipa kyse älypuhelimista, automatisoiduista autoista, roboteista tai verkkosovelluksista. Algoritmien tekemät päätökset voivat kuitenkin olla seurausta tiedoista, jotka ovat epätäydellisiä ja sen vuoksi epäluotettavia. Mahdollista on myös se, että tietoihin on kajottu kyberhyökkäyksillä, tai tiedot voivat olla puolueellisia tai yksinkertaisesti virheellisiä. Teknologian kritiikin soveltaminen sitä mukaa kuin se kehittyy johtaisi näin ollen ongelmallisiin tuloksiin ja kansalaisten haluttomuuteen hyväksyä tai käyttää sitä.

Tekoälyteknologiaa olisi sen sijaan kehitettävä siten, että ihmiset asetetaan sen keskiöön varmistaen, että teknologia on kansalaisten luottamuksen arvoinen. Tämä merkitsee sitä, että

⁴ EU on myös liittynyt vammaisten henkilöiden oikeuksia koskevaan Yhdistyneiden kansakuntien yleissopimukseen.

⁵ Asetus (EU) 2016/679. Yleinen tietosuojajärjestelmä takaa henkilötietojen vapaan liikkuvuuden unionissa. Se sisältää myös määräyksiä sellaisesta päätöksenteosta, joka perustuu yksinomaan automatisoituihin prosesseihin, kuten profilointiin. Asianomaisilla on oikeus saada tietoa automaattisen päätöksenteon olemassaolosta ja merkitykselliset tiedot automatisoituihin päätöksentekoon liittyvästä logiikasta sekä siitä, mitä merkitystä ja mitä mahdollisia vaikutuksia käsittelystä on heille. Asianomaisilla on myös tällaisissa tapauksissa oikeus vaatia, että tiedot käsittelee luonnollinen henkilö, sekä oikeus esittää kantansa ja riitauttaa päätös.

⁶ COM(2017) 10.

tekoälysovelluksissa tulisi noudattaa paitsi lakia myös eettisiä periaatteita ja varmistettava, että niiden täytäntöönpanolla vältetään ei-toivotut vaikutukset. Tekoälyn kehittämisen kaikissa vaiheissa tulisi taata monimuotoisuus sukupuoleen, rotuun tai etniseen alkuperään, uskontoon tai vakaumukseen, vammaisuuteen ja ikään katsomatta. Tekoälysovelluksissa olisi huolehdittava kansalaisten vaikutusmahdollisuuksista ja kunnioitettava heidän perusoikeuksiaan. Niillä olisi pyrittävä parantamaan ihmisten kykyjä, ei korvaamaan niitä, ja myös vammaisten mahdollisuuksista käyttää niitä olisi huolehdittava.

Näistä syistä tarvitaan **eettisiä ohjeita**, jotka perustuvat nykyiseen sääntelykehykseen ja joita tekoälyn kehittäjien, toimittajien ja käyttäjien olisi noudatettava sisämarkkinoilla, jotta voidaan luoda eettisissä kysymyksissä tasapuoliset toimintaedellytykset kaikissa jäsenvaltioissa. Tämän vuoksi komissio on perustanut **tekoälyä käsittelevän korkean tason asiantuntijaryhmän**⁷, jossa on edustettuna laaja kirjo sidosryhmiä. Sen tehtävänä on laatia tekoälyä koskevat eettiset ohjeet ja valmistella joukko suosituksia laajempaa tekoälypolitiikkaa varten. Samaan aikaan perustettiin **eurooppalainen tekoälyallianssi**⁸, johon osallistuu useista sidosryhmistä yli 2700 jäsentä antamaan laajempaa palautetta ja aineksia tekoälyä käsittelevän korkean tason asiantuntijaryhmän työhön.

Tekoälyä käsittelevä korkean tason asiantuntijaryhmä julkaisi eettisten ohjeiden ensimmäisen luonnoksen joulukuussa 2018. **Sidosryhmäkuulemisen**⁹ ja **jäsenvaltioiden edustajien kanssa käytyjen kokousten**¹⁰ jälkeen asiantuntijaryhmä toimitti komissiolle tarkistetun asiakirjan maaliskuussa 2019. Sidosryhmät ovat tähän mennessä olleet yleisesti ottaen tyytyväisiä siihen, että suuntaviivat ovat luonteeltaan käytännöllisiä ja että tekoälyn kehittäjille, tavarantoimittajille ja käyttäjille annetaan konkreettisia ohjeita siitä, miten luotettavuus voidaan varmistaa.

2.1. Korkean tason asiantuntijaryhmän laatimat luotettavaa tekoälyä koskevat ohjeet

Tekoälyä käsittelevän korkean tason asiantuntijaryhmän laatimat ohjeet, joihin tässä tiedonannossa viitataan¹¹, perustuvat erityisesti luonnontieteiden ja uusien teknologioiden etiikkaa käsittelevän eurooppalaisen työryhmän ja Euroopan unionin perusoikeusviraston tekemään työhön.

Suuntaviivoissa esitetään, että luotettavan tekoälyn edellytyksenä on kolme osatekijää: 1) sen olisi kunnioitettava lainsäädäntöä, 2) sen olisi täytettävä eettiset periaatteet ja 3) sen olisi oltava teknisesti luotettava.

Näiden kolmen osatekijän ja jaksossa 2 esitettyjen eurooppalaisten arvojen pohjalta ohjeissa määritetään seitsemän keskeistä vaatimusta sille, että tekoälysovelluksia voidaan pitää luotettavina. Ohjeisiin sisältyy myös luettelo kysymyksistä, jotka tulisi ottaa huomioon arvioitaessa, täyttyvätkö nämä vaatimukset.

⁷ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

⁸ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

⁹ Sidosryhmäkuulemisessa saatiin kommentteja 511 organisaatiolta, yhdistykseltä, yritykseltä, tutkimuslaitokselta, yksityishenkilöltä ja muilta tahoilta. Tiivistelmä saadusta palautteesta on osoitteessa https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf

¹⁰ Jäsenvaltiot suhtautuivat myönteisesti asiantuntijaryhmän työhön 18. helmikuuta 2019 annetuissa neuvoston päätelmissä, jolloin ne muun muassa panivat merkille tarkoituksen julkistaa eettiset ohjeet ja tukivat komission pyrkimyksiä saada EU:n eettinen lähestymistapa käyttöön maailmanlaajuisesti: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/en/pdf>

¹¹ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

Seitsemän keskeistä vaatimusta ovat seuraavat:

- Ihmisen toimijuus ja ihmisen suorittama valvonta
- Tekninen luotettavuus ja turvallisuus
- Yksityisyyden suoja ja datan hallinta
- Läpinäkyvyys
- Monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus
- Yhteiskunnallinen ja ekologinen hyvinvointi
- Vastuuvellisuus

Vaikka näitä vaatimuksia on tarkoitus soveltaa kaikkiin tekoälyjärjestelmiin eri ympäristöissä ja toimialoilla, se kulloinkin konteksti, jossa niitä sovelletaan, olisi otettava huomioon niiden konkreettista ja oikeasuhteista täytäntöönpanoa varten käyttäen vaikutuslähtöistä lähestymistapaa. Esimerkiksi tekoälysovellus, joka ehdottaa luettavaksi tarkoitukseen sopimatonta kirjaa, on paljon vaarattomampi kuin virheellinen syöpädiagnoosi, minkä vuoksi tällainen sovellus voitaisiin jättää vähemmälle valvonnalle.

Korkean tason asiantuntijaryhmän laatimat ohjeet eivät ole sitovia, eivätkä ne sinällään luo uusia oikeudellisia velvoitteita. Monet jo käytössä olevat (ja usein käyttö- tai alakohtaiset) unionin oikeuden säännökset tietenkin heijastavat jo yhtä tai useampaa näistä keskeisistä vaatimuksista, esimerkkeinä säännöt, jotka koskevat turvallisuutta, henkilötietojen suojaa, yksityisyyden suojaa tai ympäristönsuojelua.

Komissio on tyytyväinen tekoälyä käsittelevän korkean tason asiantuntijaryhmän työhön ja pitää sitä arvokkaana panoksena päätöksentekoon.

2.2. Luotettavan tekoälyn keskeiset vaatimukset

Komissio tukee seuraavassa esitettyjä keskeisiä vaatimuksia, jotka perustuvat eurooppalaisiin arvoihin. Komissio kannustaa sidosryhmiä soveltamaan vaatimuksia ja testaamaan luetteloa kysymyksistä, jotka olisi otettava huomioon arvioitaessa, täyttyvätkö vaatimukset käytännössä. Tavoitteena on luoda oikea luottamuksen ilmapiiri, jotta tekoälyn kehittämisestä ja käytöstä voidaan huolehtia onnistuneesti. Komissio ottaa sidosryhmiltä mielellään vastaan palautetta, jonka pohjalta voidaan arvioida, onko luetteloon sisältyviä kysymyksiä tarpeen mukauttaa edelleen.

I. Ihmisen toimijuus ja ihmisen suorittama valvonta

Tekoälyjärjestelmien pitäisi auttaa ihmisiä tekemään parempia ja paremmin tietoon perustuvia valintoja omien tavoitteidensa mukaisesti. Niiden tulisi toimia kukoistavan ja tasa-arvoisen yhteiskunnan mahdollistajina tukemalla ihmisen toimijuutta ja edistämällä **perusoikeuksia**, eikä niillä saisi vähentää, rajoittaa tai ohjata väärin ihmisen itsenäistä päätäntävaltaa. **Käyttäjän yleisen hyvinvoinnin** tulisi olla järjestelmän toimivuuden kannalta keskiössä.

Ihmisen suorittama valvonta auttaa varmistamaan, ettei tekoälyjärjestelmä heikennä ihmisen itsenäistä päätäntävaltaa tai aiheuta muita kielteisiä vaikutuksia. Kulloisestakin tekoälypohjaisesta järjestelmästä ja sen sovellusalasta riippuen olisi varmistettava

asianmukaisesti mitoitettut **valvontatoimenpiteet**, mukaan lukien tekoälypohjaisten järjestelmien mukautuvuus, tarkkuus ja selitettävyyden (explainability)¹². **Valvonta** voidaan toteuttaa eri hallintamekanismeilla, esimerkiksi ihmisen osallistumista edellyttävällä periaatteella, ihmisen osallistumisen mahdollistavalla periaatteella tai periaatteella, jonka mukaan ihminen määrää.¹³ On varmistettava, että viranomaiset pystyvät käyttämään valvontavaltuuksiaan toimeksiantonsa mukaisesti. Muutoin samanlaisten edellytysten vallitessa pätee, että mitä vähemmän ihminen voi valvoa tekoälyjärjestelmää, sitä laajempaa testausta ja tiukempaa hallintaa tarvitaan.

II. Tekninen luotettavuus ja turvallisuus

Luotettava tekoäly edellyttää, että algoritmit ovat turvallisia, luotettavia ja riittävän robusteja virheellisyyksien tai epäjohdonmukaisuuksien varalta kaikissa tekoälyjärjestelmän elinkaaren vaiheissa ja että algoritmit selviytyvät virheellisistä tuloksista. Tekoälyjärjestelmien on oltava **luotettavia** ja **resilienssiltään** riittävän turvallisia, jotta ne voivat kestää sekä avoimia hyökkäyksiä että hienovaraisempia yrityksiä manipuloida dataa tai itse algoritmeja, ja niissä on varmistettava **varasuunnitelma** ongelmatapausten varalta. Niiden päätösten on oltava **tarkkoja** tai ainakin tarkkuustasoaan vastaavia, ja niiden tulosten tulisi olla **toistettavissa**.

Lisäksi tekoälyjärjestelmiin olisi integroitava sisäänrakennetun turvan ja turvallisuuden mekanismit, jotta voidaan varmistaa, että ne ovat **todistettavasti turvallisia** kaikissa vaiheissa ottaen huomioon kaikkien asianomaisten fyysinen ja henkinen turvallisuus. Tähän sisältyy järjestelmän toiminnassa mahdollisesti ilmenevien tahattomien seurausten tai virheiden minimointi ja mahdollisuuksien mukaan peruutettavuus. Olisi otettava käyttöön prosesseja, joilla selvennetään ja arvioidaan tekoälyjärjestelmien käyttöön liittyviä mahdollisia riskejä eri sovellusaloilla.

III. Yksityisyyden suoja ja datan hallinta

Yksityisyyden suoja ja **tietosuoja** on taattava tekoälyjärjestelmän elinkaaren **kaikissa vaiheissa**. On mahdollista, että tekoälyjärjestelmät voivat tehdä ihmisen käyttäytymisen digitaalisista tallenteista päätelmiä yksilöiden mieltymyksistä, iästä ja sukupuolesta ja myös heidän sukupuolisesta suuntautumistaan ja uskonnollisista tai poliittisista näkemyksistään. Jotta yksilöt voisivat luottaa tietojen käsittelyyn, on varmistettava, että heillä on täydet mahdollisuudet valvoa heitä koskevia tietoja ja ettei tällaisia tietoja käytetä heitä vahingoittavalla tai syrjivällä tavalla.

Yksityisyyden ja henkilötietojen suojaamisen lisäksi korkealaatuisten tekoälyjärjestelmien on täytettävä myös muita datahallinnan vaatimuksia. Käytettävien tietueiden laadulla on erittäin tärkeä merkitys tekoälyjärjestelmien suorituskyvyn kannalta. Kun tietoja kerätään,

¹² Yleinen tietosuoja-asetus antaa yksilöille oikeuden olla joutumatta sellaisen päätöksen kohteeksi, joka perustuu yksinomaan automaattiseen käsittelyyn, kun tämä aiheuttaa oikeusvaikutuksia käyttäjille tai vaikuttaa heihin vastaavalla tavalla merkittävästi (yleisen tietosuoja-asetuksen 22 artikla).

¹³ Näistä periaatteista *human-in-the-loop* (HITL) edellyttää ihmisen osallistumista jokaiseen järjestelmän päätöksentekoketjuun, mikä ei monissa tapauksissa ole mahdollista tai edes toivottavaa. *Human-on-the-loop* (HOTL) edellyttää ihmisen mahdollisuutta osallistua järjestelmän suunnittelusyklin aikana ja seurata järjestelmän toimintaa. *Human-in-command* (HIC) edellyttää mahdollisuutta valvoa tekoälyjärjestelmän yleistä toimintaa (ml. sen laajemmat taloudelliset, yhteiskunnalliset, oikeudelliset ja eettiset vaikutukset) ja kykyä päättää, milloin ja miten järjestelmää käytetään tietyssä tilanteessa. Tähän voi sisältyä päätös olla käyttämättä tiettyä tekoälyjärjestelmää tietyssä tilanteessa, määrittää ihmisen harkintavallan tasot järjestelmän käytön aikana tai varmistaa, että järjestelmän tekemä päätös voidaan kumota.

ne voivat heijastaa yhteiskunnallisesti rakentuneita vääristymiä tai sisältää epätarkkuuksia, virheitä ja puutteellisuuksia. Tähän on puututtava ennen tekoälyjärjestelmän kouluttamista tietyn tietokokonaisuuden kanssa. Lisäksi on varmistettava tietojen **eheys**. Käytettävät prosessit ja tietokokonaisuudet on testattava ja dokumentoitava kussakin vaiheessa, kuten suunnittelussa, kouluttamisessa, testauksessa ja käyttöönotossa. Tätä olisi sovellettava myös tekoälyjärjestelmiin, joita ei ole kehitetty yrityksen sisällä vaan jotka on hankittu muualta. Lisäksi tietojen **saatavuutta** on säänneltävä ja valvottava asianmukaisesti.

IV. Läpinäkyvyys

Tekoälyjärjestelmien **jäljitettävyys** olisi varmistettava. On tärkeää kirjata ja dokumentoida sekä järjestelmien tekemät päätökset että päätöksiin johtanut koko prosessi (ml. kuvaus tietojen keräämisestä ja käytetyistä tunnisteista sekä kuvaus käytetystä algoritmista). Tähän liittyen tulisi mahdollisuuksien mukaan huolehtia algoritmeihin perustuvan päätöksentekoprosessin **selitettävyydestä** tavalla, joka on mukautettu asianomaisten henkilöiden tarpeisiin. Meneillään olevaa tutkimusta selitettävyydsmekanismien kehittämiseksi olisi jatkettava. Lisäksi tulisi olla saatavilla selvityksiä siitä, missä määrin tekoälyjärjestelmä vaikuttaa organisaation päätöksentekoprosessiin ja muovaa sitä, sekä järjestelmän suunnitteluun liittyvistä valinnoista ja sen käyttöönoton perusteista (näin varmistetaan paitsi tietojen ja järjestelmän myös liiketoimintamallin läpinäkyvyys).

On myös tärkeää **tiedottaa** riittävästi tekoälyjärjestelmän kapasiteetista ja rajoituksista asianomaisille eri sidosryhmille kulloiseenkin tapaukseen soveltuvalla tavalla. Lisäksi tekoälyjärjestelmät olisi voitava tunnistaa sellaisiksi, jotta käyttäjät tietävät, milloin he ovat vuorovaikutuksessa tekoälyjärjestelmän kanssa ja ketkä ovat järjestelmän vastuuhenkilöt.

V. Monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus

Tekoälyjärjestelmien (sekä kouluttamista että toimintaa varten) käyttämät tietokokonaisuudet voivat kärsiä tahattomista ajallisista vääristymistä, epätäydellisyyksistä ja huonoista tietohallintomalleista. Tällaisten vääristymien jatkuminen voi johtaa (epä)suoraan syrjintään. Haittaa voi aiheutua myös tarkoituksellisesta (kuluttaja)suuntauksien hyödyntämisestä tai vilpillisestä kilpailusta. Vääristymät voivat vaikuttaa myös tapaan, jolla tekoälyjärjestelmiä kehitetään (esim. tapaan, jolla algoritmin ohjelmakoodi on kirjoitettu). Näihin ongelmiin olisi puututtava järjestelmän kehittämisen alusta lähtien.

Näihin ongelmiin voidaan puuttua myös perustamalla **monimuotoisia suunnitteluryhmiä** ja ottamalla käyttöön mekanismeja, joilla varmistetaan erityisesti kansalaisten **osallistuminen** tekoälyn kehittämiseen. On suositeltavaa kuulla sidosryhmiä, joihin järjestelmä saattaa vaikuttaa suoraan tai välillisesti sen elinkaaren aikana. Tekoälyjärjestelmissä olisi otettava huomioon kaikki inhimilliset kyvyt, taidot ja vaatimukset ja varmistettava esteettömyys omaksumalla kaikkia palvelevan suunnittelun lähestymistapa, jolla pyritään huolehtimaan yhtäläisistä mahdollisuuksista myös vammaisille.

VI. Yhteiskunnallinen ja ekologinen hyvinvointi

Jotta tekoäly olisi luotettava, sen vaikutus **ympäristöön ja muihin tunteviin olentoihin** olisi otettava huomioon. Ihannetapauksessa kaikki ihmiset, myös tulevat sukupolvet, voivat hyötyä luonnon monimuotoisuudesta ja elinkelpoisesta ympäristöstä. Tekoälyjärjestelmien kestävyyttä ja **ekologista vastuuta** tulisi näin ollen edistää. Sama koskee tekoälyratkaisuja, jotka koskettavat maailmanlaajuisia kysymyksiä, kuten YK:n kestävä kehityksen tavoitteita.

Tekoälyjärjestelmien vaikutuksia olisi tarkasteltava paitsi yksilön näkökulmasta myös **koko yhteiskunnan** näkökulmasta. Tekoälyjärjestelmien käyttöä olisi harkittava huolellisesti varsinkin demokraattiseen prosessiin liittyvissä tilanteissa, kuten julkisessa mielipiteenmuodostuksessa, poliittisessa päätöksenteossa ja vaalikontekstissa. Lisäksi olisi otettava huomioon tekoälyn **sosiaaliset vaikutukset**. Tekoälyjärjestelmiä voidaan käyttää sosiaalisten taitojen parantamiseen, mutta ne voivat yhtä lailla vaikuttaa myös niiden heikentymiseen.

VII. Vastuuvollisuus

Käyttöön tulisi ottaa mekanismeja, joilla varmistetaan tekoälyjärjestelmiin ja niiden tuloksiin liittyvä vastuu ja vastuuvollisuus ennen järjestelmien toteuttamista ja sen jälkeen. Tekoälyjärjestelmien **tarkastettavuus** on tässä yhteydessä keskeistä, sillä sisäisten ja ulkoisten tarkastajien suorittama tekoälyjärjestelmien arviointi sekä tällaisten arviointiselvitysten saatavuus tukevat merkittävästi teknologian luotettavuutta. Ulkoisen tarkastuksen mahdollisuus tulisi varmistaa erityisesti perusoikeuksiin vaikuttavissa sovelluksissa, mukaan lukien turvallisuuden kannalta kriittiset sovellukset.

Tekoälyjärjestelmien mahdolliset kielteiset vaikutukset olisi tunnistettava, arvioitava, dokumentoitava ja minimoitava. Vaikutusarviointien käyttö helpottaa tätä prosessia. Näiden arviointien tulisi olla suhteessa tekoälyjärjestelmien riskien laajuuteen. Eri vaatimusten väliset **kompromissit** – joita ei useinkaan voida välttää – tulisi tehdä rationaalisella ja järjestelmällisellä tavalla ja ne tulisi ottaa huomioon. Perusteettomien kielteisten vaikutusten varalta tulisi huolehtia siitä, että käytössä on helpokäyttöisiä mekanismeja **asianmukaista muutoksenhakua** varten.

2.3. Seuraavat vaiheet: pilottivaihe, johon sidosryhmät osallistuvat mahdollisimman laajalti

Pääseminen yksimielisyyteen näistä tekoälyjärjestelmien keskeisistä vaatimuksista on ensimmäinen tärkeä merkkipaalu eettistä tekoälyä koskevien ohjeiden käyttöönotolle. Seuraavaksi komissio varmistaa, että näitä ohjeita voidaan testata ja soveltaa käytännössä.

Tätä varten komissio käynnistää nyt kohdennetun pilottivaiheen, jonka tarkoituksena on saada jäsenneilyä palautetta sidosryhmiltä. Tässä keskitytään erityisesti korkean tason asiantuntijaryhmän laatimaan luetteloon kysymyksistä, jotka tulisi ottaa huomioon arvioitaessa kunkin keskeisen vaatimuksen täyttymistä.

Tässä työssä on kaksi osaa: i) ohjeiden pilotointivaihe tekoälyä kehittäville tai käyttäville sidosryhmille, julkinen hallinto mukaan lukien, ja ii) jatkuva kuulemis- ja tiedotusprosessi jäsenvaltioissa ja eri sidosryhmiin päin, mukaan lukien teollisuus- ja palvelusalat:

- i) Kesäkuusta 2019 alkaen kaikkia sidosryhmiä ja yksityishenkilöitä pyydetään testaamaan luettelo huomioon otettavista kysymyksistä ja antamaan palautetta sen parantamiseksi. Lisäksi tekoälyä käsittelevä korkean tason asiantuntijaryhmä

käynnistää yksityisen ja julkisen sektorin sidosryhmien kanssa perusteellisen tarkastelun kerätäkseen tarkempaa palautetta siitä, miten ohjeet voidaan panna täytäntöön useilla eri sovellusaloilla. Kaikki ohjeiden toimivuudesta ja toteutettavuudesta kerätty palaute arvioidaan vuoden 2019 loppuun mennessä.

- ii) Samaan aikaan komissio järjestää myös muita tiedotustoimia, jotka antavat tekoälyä käsittelevän korkean tason asiantuntijaryhmän edustajille mahdollisuuden esitellä suuntaviivat sidosryhmille jäsenvaltioissa, mukaan lukien teollisuus- ja palvelualat. Samalla sidosryhmille tarjoutuu lisämahdollisuus kommentoida ohjeita ja antaa niihin panoksensa.

Komissio ottaa huomioon verkottuneeseen ja automatisoituun ajamiseen liittyvää etiikkaa käsittelevän asiantuntijaryhmän¹⁴ työn ja tekee yhteistyötä EU:n rahoittamien tekoälyyn liittyvien tutkimushankkeiden ja julkisen ja yksityisen sektorin kumppanuuksien kanssa keskeisten vaatimusten täytäntöönpanoon liittyen¹⁵. Komissio esimerkiksi tukee koordinoitusti yhdessä jäsenvaltioiden kanssa yhteisen lääketieteellisen kuvatietokannan kehittämistä siten, että aluksi keskitytään yleisimpiin syöpätyyppeihin, jotta algoritmit voidaan kouluttaa oireiden diagnosointiin suurella tarkkuudella. Vastaavasti komission ja jäsenvaltioiden yhteistyö mahdollistaa yhä suuremman määrän rajat ylittäviä testauskäytäviä verkkoon liitettyjen ja automatisoitujen ajoneuvojen testaamista varten. Ohjeita olisi sovellettava ja testattava näissä hankkeissa, ja tulokset otetaan huomioon arviointiprosessissa.

Pilottivaiheeseen ja sidosryhmäkuulemiseen osallistuvat myös eurooppalainen tekoälyallianssi ja tekoälyn tilausalusta AI4EU. Tammikuussa 2019 käynnistetty AI4EU-hanke¹⁶ kokoaa yhteen algoritmeja, välineitä, tietokokonaisuuksia ja palveluja, joiden avulla organisaatiot, erityisesti pienet ja keskisuuret yritykset, voivat toteuttaa tekoälyratkaisuja. Eurooppalainen tekoälyallianssi jatkaa yhdessä AI4EU-aloitteen kanssa tekoälyekosysteemin mobilisointia Euroopassa, myös tekoälyä koskevien eettisten ohjeiden pilotoimiseksi ja ihmiskeskeisen tekoälyn periaatteiden edistämiseksi.

Vuoden 2020 alussa tekoälyä käsittelevä korkean tason asiantuntijaryhmä **tarkistaa ja päivittää ohjeita** pilottivaiheen aikana saadun palautteen arvioinnin perusteella. Kyseisen tarkastelun ja saatujen kokemusten perusteella **komissio arvioi saadut tulokset ja ehdottaa mahdollisia seuraavia vaiheita.**

Eettinen tekoäly on kaikkia hyödyttävä hanke. On sinänsä olennaista varmistaa perusarvojen ja -oikeuksien kunnioittaminen, mutta se myös helpottaa hyväksyttävyyttä suuren yleisön keskuudessa ja lisää eurooppalaisten tekoäly-yritysten kilpailuetua luomalla tavaramerkin ihmiskeskeiselle, luotettavalle tekoälylle – brändin, josta eettiset ja turvalliset tuotteet tunnetaan. Tässä rakennetaan yleisemmin eurooppalaisten yritysten vahvalle maineelle turvallisten ja korkealaatuisten tuotteiden tarjoajina. Pilottivaihe auttaa varmistamaan, että tekoälytuotteet lunastavat tämän lupauksen.

2.4. Kohti kansainvälisiä eettisiä ohjeita

Tekoälyn etiikkaa koskevat kansainväliset keskustelut ovat tiivistyneet sen jälkeen, kun asia nostettiin tärkeälle sijalle G7-asialistalla Japanin puheenjohtajakaudella vuonna 2016. Koska

¹⁴ Ks. verkkoon liitettyä ja automatisoitua liikkuvuutta koskeva komission tiedonanto COM(2018) 283.

¹⁵ Euroopan puolustusrahaston puitteissa komissio aikoo kehittää myös erityisiä eettisiä ohjeita, jotka koskevat tekoälyalan hanke-ehdotusten arviointia puolustustarkoituksiin.

¹⁶ <https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019>

tekoälyn kehittäminen on kansainvälisesti verkottunutta datan liikkuvuudessa, algoritmeihin kehittämiseen ja tutkimusinvestoinneissa, **komissio aikoo jatkaa pyrkimyksiään saada unionin lähestymistapa käyttöön maailmanlaajuisesti ja luoda konsensusta ihmiskeskeiselle tekoälylle**¹⁷.

Tekoälyä käsittelevän korkean tason asiantuntijaryhmän työ ja erityisesti luettelo vaatimuksista ja kuulemisprosessi sidosryhmien kanssa antavat komissiolle arvokkaan lisäpanoksen kansainvälisiin keskusteluihin. Euroopan unionilla voi olla johtava rooli kehitettäessä kansainvälisiä tekoälyn suuntaviivoja ja mahdollisuuksien mukaan niihin liittyvää arviointimekanismia.

Tämän vuoksi komissio aikoo:

Lujittaa yhteistyötä samanmielisten kumppaneiden kanssa

- selvittämällä, missä määrin lähentymistä voidaan saada aikaan suhteessa kolmansien maiden (esim. Japanin, Kanadan ja Singaporen) eettisten ohjeiden luonnoksiin ja valmistelemalla tämän samanmielisten maiden ryhmän kanssa laajempaa keskustelua tukeutuen toimiin, joilla pannaan täytäntöön kumppanuusväline kolmansien maiden kanssa tehtävää yhteistyötä varten¹⁸; ja
- selvittämällä, miten EU:n ulkopuolisten maiden yritykset ja kansainväliset järjestöt voivat osallistua eettisten ohjeiden pilottivaiheeseen testauksen ja validoinnin kautta.

Jatkaa aktiivista roolia kansainvälisissä keskusteluissa ja aloitteissa

- osallistumalla monenvälisiin foorumeihin, kuten G7- ja G20-kokouksiin;
- käymällä vuoropuhelua EU:n ulkopuolisten maiden kanssa ja järjestämällä kahden- ja monenvälisiä tapaamisia yhteisymmärryksen kehittämiseksi ihmiskeskeisestä tekoälystä;
- osallistumalla asiaa koskeviin kansainvälisten standardointijärjestöjen standardointitoimiin tämän vision edistämiseksi; ja
- tehostamalla julkiseen politiikkaan liittyvien tulosten ja näkemysten keräämistä ja jakamista yhteistyössä asiaankuuluvien kansainvälisten järjestöjen kanssa.

3. PÄÄTELMÄT

EU:n perustana on joukko perusarvoja, joiden varaan se luonut vahvan ja tasapainoisen sääntelykehiksen. Tähän nykyiseen sääntelyyn pohjautuen tarvitaan eettisiä ohjeita tekoälyn

¹⁷ Unionin ulkoasioiden ja turvallisuuspolitiikan korkea edustaja muun muassa koordinoi komission tuella ja YK:ssa, maailmanlaajuisessa teknologiapaneelissa (Global Tech Panel) ja muissa monenkeskisissä foorumeissa käytyjen keskustelujen pohjalta ehdotuksia asiaan liittyvien monimutkaisten turvallisuushaasteiden ratkaisemiseksi.

¹⁸ Euroopan parlamentin ja neuvoston asetus (EU) N:o 234/2014, annettu 11 päivänä maaliskuuta 2014, kumppanuusvälineen perustamisesta kolmansien maiden kanssa tehtävää yhteistyötä varten (EUVL L 77, 15.3.2014, s. 77). Esimerkiksi suunniteltu kansainvälinen allianssi ihmiskeskeistä tekoälyä varten (*An international alliance for a human-centric approach to artificial intelligence*) helpottaa yhteisiä aloitteita samanmielisten kumppanien kanssa, jotta voidaan edistää eettisiä ohjeita ja sopia yhteisistä periaatteista ja toimintaa koskevista päätelmistä. Sen avulla EU ja samanmieliset maat voivat keskustella toimintaa koskevasta päätelmästä asiantuntijaryhmän ehdottamien tekoälyä koskevien eettisten ohjeiden pohjalta tavoitteenaan sopia yhteisestä lähestymistavasta. Lisäksi tarkoituksena on seurata tekoälyteknologian käyttöönottoa maailmanlaajuisesti. Allianssihankkeessa suunnitellaan myös julkisen diplomatian toimia kansainvälisiin tapahtumiin, kuten G7- ja G20-kokousten, sekä Taloudellisen yhteistyön ja kehityksen järjestön (OECD) yhteydessä.

kehittämistä ja käyttöä varten, mihin ovat syynä teknologian uutuus ja sen mukanaan tuomat erityishaasteet. Tekoälyä voidaan pitää luotettavana vain sillä ehdolla, että sitä kehitetään ja käytetään tavalla, joka kunnioittaa laajalti tunnustettuja eettisiä arvoja.

Tätä tavoitetta katsoen komissio on tyytyväinen tekoälyä käsittelevän korkean tason asiantuntijaryhmän työpanokseen. Asiantuntijaryhmä on määrittänyt keskeiset vaatimukset sille, että tekoälysovelluksia voidaan pitää luotettavina. Tältä pohjalta komissio käynnistää nyt kohdennetun pilottivaiheen, jotta voidaan varmistaa, että tuloksena olevat tekoälyn kehittämistä ja käyttöä koskevat eettiset ohjeet voidaan panna täytäntöön käytännössä. Komissio pyrkii myös saamaan aikaan laajan yhteiskunnallisen konsensuksen ihmiskeskeisestä tekoälystä, myös kaikkien sidosryhmien ja kansainvälisten kumppaniemme kanssa.

Tekoälyn eettinen ulottuvuus ei ole ylellisyyttä tai oheisominaisuus: sen on oltava erottamaton osa tekoälyn kehittämistä. Pyrkimällä kohti luottamukseen perustuvaa ihmiskeskeistä tekoälyä taataan eurooppalaisen yhteiskunnan keskeisten arvojen kunnioittaminen ja luodaan Euroopalle ja sen teollisuudelle erottuva tavaramerkki johtavana tekoälyn kehittäjänä, johon voidaan luottaa kaikkialla maailmassa.

Jotta tekoälyä voitaisiin kehittää Euroopassa eettisellä tavalla sen laajemmassa yhteydessä, komissio noudattaa kokonaisvaltaista lähestymistapaa erityisesti seuraavin toimintalinjoin, jotka on määrä panna täytäntöön vuoden 2019 kolmannella neljänneksellä:

- Komissio aikoo käynnistää joukon **tekoälytutkimuksen huippuosaamiskeskusten verkostoja** Horisontti 2020 -ohjelman kautta. Se aikoo valita enimmillään neljä verkostoa keskittyen merkittäviin tieteellisiin tai teknologisiin haasteisiin, kuten selitettävyyteen (explainability) ja pitkälle vietyyn ihmisen ja koneen vuorovaikutukseen, jotka ovat luotettavan tekoälyn keskeisiä osatekijöitä.
- Komissio aloittaa **digi-innovaatiokeskittymien verkostojen**¹⁹ perustamisen keskittyen tekoälyyn valmistusteollisuudessa ja massadataan.
- Komissio aloittaa yhdessä jäsenvaltioiden ja sidosryhmien kanssa valmistelevat keskustelut kehittääkseen ja ottaakseen käyttöön **mallin datan yhteiskäyttöä ja yhteisten data-avaruuksien parasta hyödyntämistä varten** keskittyen erityisesti liikenteeseen, terveydenhuoltoon ja teolliseen valmistukseen.²⁰

Lisäksi komissio laatii parhaillaan kertomusta tekoälyn haasteista turvallisuus- ja vastuukehysten kannalta sekä ohjeasiakirjaa tuotevastuudirektiivin täytäntöönpanosta²¹. Samaan aikaan Euroopan suurteholaskennan yhteisyritys (EuroHPC)²² kehittää supertietokoneiden uutta sukupolvea, koska laskentakapasiteetti on olennainen tekijä tiedonkäsittelyn ja tekoälyn kouluttamisen kannalta ja koska Euroopan on hallittava koko digitaalinen arvoketju. Jäsenvaltioiden ja teollisuuden kanssa parhaillaan käynnissä oleva mikroelektroniikkakomponentteja ja -järjestelmiä koskeva kumppanuus (ECSEL)²³ ja

¹⁹ <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs>

²⁰ Tarvittavat resurssit otetaan käyttöön Horisontti 2020 -ohjelmasta (josta tekoälyyn on osoitettu lähes 1,5 miljardia euroa vuosina 2018–2020) ja sen suunnitellusta seuraajasta ”Euroopan horisontti”, Verkkojen Eurooppa -välineen digitaalisesta osasta ja erityisesti tulevasta Digitaalinen Eurooppa -ohjelmasta. Hankkeissa hyödynnetään myös yksityiseltä sektorilta ja jäsenvaltioiden ohjelmista saatavia resursseja.

²¹ Ks. tekoälyä Euroopassa koskeva komission tiedonanto COM(2018) 237.

²² <https://eurohpc-ju.europa.eu>

²³ www.ecsel.eu

eurooppalainen prosessorialoite²⁴ edistävät pienitehoisen prosessoriteknologian kehittämistä luotettavaa ja turvallista suurteholaskentaa ja reunalaskentaa (edge computing) varten.

Tekoälyä koskevien eettisten ohjeiden tavoin kaikki nämä aloitteet perustuvat kaikkien asianomaisten sidosryhmien, jäsenvaltioiden, teollisuuden, yhteiskunnallisten toimijoiden ja kansalaisten **tiiviseen yhteistyöhön**. Kaiken kaikkiaan Euroopan lähestymistapa tekoälyyn osoittaa, miten taloudellinen kilpailukyky ja yhteiskunnallinen luottamus lähtevät samoista perusarvoista ja vahvistavat toisiaan.

²⁴ www.european-processor-initiative.eu