TEKOÄLYRAPORTTI

Eduskunnan Tulevaisuusvaliokunta

Sisällys

Tekoälystä vastuullista 3

Mistä on kyse? 3

Tekoälyn eettiset kysymykset ovat moniulotteisia 3

Eettiset koodistot ja ohjeet viitoittavat suuntaa 4

Tekoäly muutosvoimana 5

Tekoäly haastaa luottamusta 5

Lisää näkemystä kokonaisuuden hallintaan 5

Lainsäädännön rinnalle pehmeää ohjausta 6

Suunnittelijan vastuu on suuri 7

Viestit päättäjille ja kansalaisille 7

Tekoälystä vastuullista

Jaana Leikas, Teknologian tutkimuskeskus VTT Oy

Mistä on kyse?

Algoritmisen päätöksenteon mahdollisuuksien ja eettisten haasteiden ymmärtäminen ja tekoälyn vastuullinen hyödyntäminen ovat aikamme avainkysymyksiä. **Taloudellisten hyötyjen ja lisääntyvän tehokkuuden lisäksi tekoälystä on povattu merkittävää tekijää hyvän (tekoäly-)yhteiskunnan ja YK:n sosiaalisten kehitystavoitteiden edistämisessä.** Julkisessa puheessa tekoälyn avulla pyritään luomaan paitsi vaurautta ja kilpailukykyä, myös oikeudenmukaista ja kestävää yhteiskuntaa, sekä poliittista ja sosiaalista tasa-arvoa. Parhaassa tapauksessa tekoäly on hyvä työkalu suuren datamäärän ja kompleksisten suhteiden jäsentämisessä. Se voi olla loistava apuri kansalaisten elämän ja organisaatioiden toiminnan suuntaamisessa. Itse asiassa näiden tekoälyyn kohdistettujen odotusten vuoksi sen alikäyttöä ja potentiaalin hyödyntämättä jättämistä voidaan pitää eettisesti arveluttavana ja jopa riskinä yhteiskunnalle.

Kolikolla on myös kääntöpuolensa. **Tekoälyn suunnitteluun, kehittämiseen ja käyttöönottoon liittyy ristiriitaisia kysymyksiä.** Riskinä voidaan nähdä tekoälyn väärinkäyttö ja sen hyödyntäminen tarkoituksiin, jotka eivät tue ihmiskunnan kukoistusta pitkällä aikavälillä. Tekoälyjärjestelmien kehittäminen voidaan nähdä myös poliittisena kysymyksenä, joka liittyy vallan ja legitiimiyden kysymyksiin. Huolenaiheena on, että tekoälyjärjestelmien soveltaminen voi entisestään lisätä vallan ja vaurauden epäsymmetriaa, ja siten syventää olemassa olevaa ja jopa luoda uudenlaista eriarvoisuutta.

Tekoälyn eettiset kysymykset ovat moniulotteisia

Tekoälyn eettiset haasteet (pois lukien superälykästä tekoälyä koskevat kysymykset) voidaan karkeasti jakaa **koneoppimisesta ja algoritmeista juontuviin haasteisiin** (esimerkiksi datan vääristyneisyys, syrjintä ja datan turvallisuus) ja **yhteiskunnallisiin haasteisiin** (esimerkiksi muutokset ihmisen autonomiassa sekä ihmisen ja koneen välisessä vuorovaikutuksessa, järjestelmien sosiaalinen hyväksyntä, tasa-arvokysymykset ja yhteiskunnalliset valtarakenteet sekä työn muutokseen ja ekologiseen kestävyyteen liittyvät kysymykset).[1]

Kaikkia tekoälyn negatiivisia vaikutuksia ei luonnollisesti osata arvioida ennalta. Riskien arviointia helpottanee tarkastelun aikajanan asettaminen nykyisyyteen, lähitulevaisuuteen ja pitkälle aikavälille: Millaisia ovat ne tekoälyn riskit, jotka voivat olla nähtävissä jo tänään? Millaisia riskejä voimme kohdata tulevaisuudessa tekoälyn käytössä

[1] Esim. Stahl, Bernt Carsten (2021). Artificial Intelligence for a Better Future. Springer briefs in research and innovation governance.

nykyisten teknologioiden konteksteissa? Mitä riskejä on kohdattavissa pitkällä aikavälillä, kun teknologia kehittyy? Näiden kysymysten operationalisointi ja istuttaminen tekoälyn konteksteihin vaatii analyyttisyyttä ja uusia työkaluja.

Ihmiskeskeisyyttä pidetään keskeisenä tekijänä tekoälyn kehittämisessä. Useat Euroopan komission strategiat ja ohjeet korostavat konseptia keskeisenä tavoitteena: ihmisten tulee voida luottaa siihen, että tekoälyä käytetään turvallisesti ja lain mukaisesti, ja että kansalaisten perusoikeuksia kunnioitetaan. Asiakirjoissa viitataan kuitenkin ihmiskeskeisyyteen vain pinnallisesti, rajoittuen lähinnä ihmisoikeuksien toteutumiseen. Sama näkyy teknologian suunnitteluohjeissa: vaikka ihmiskeskeisyyden soveltamisella on pitkät perinteet ihmisen ja teknologian välisen vuorovaikutuksen suunnittelussa, suunnittelun kentältä puuttuu edelleen etiikkaa ja sosiaalista kestävyyttä tarkastelevia teknologian suunnittelumenetelmiä.

Eettiset koodistot ja ohjeet viitoittavat suuntaa

Tilannetta helpottamaan on luotu erilaisia eettisiä koodistoja, mm. turvaamaan teknologian käytön kohteena olevien henkilöiden asema ja oikeudet. Useiden valtioiden, suuryritysten ja kansainvälisten järjestöjen tekoälyn etiikkaa linjaavissa dokumenteissa eettisinä periaatteina mainitaan läpinäkyvyys, oikeudenmukaisuus, vahingonteon välttäminen, vastuullisuus ja yksityisyys. Euroopan unionissa vaatimuksina ovat lisäksi tekoälytoiminnan selitettävyys ja ihmislähtöisyys. Eurooppalaisten maiden poliittisissa dokumenteissa vaaditaan lisää luottamusta, vastuullisuutta, vastuuvelvollisuutta, läpinäkyvyyttä ja turvallisuutta tekoälyn kehitykseen ja hyödyntämiseen.

Myös tutkimuskenttä on kehittynyt nopeasti, ja **lukuisia erilaisia ohjeistuksia ja käytäntöjä sosiaalisesti hyväksyttävän tekoälyn suunnitteluun on kehitteillä**. Tutkijat ovat niin ikään esittäneet näkemyksiä eettisten periaatteiden juurruttamiseksi tekoälyn ohjaus- ja koordinaatiomekanismeihin, joista osa on soveltaen kirjattu osaksi EU:n tekoälyn hallinnan toimenpidesuosituksia [2] sekä OECD:n periaatteisiin[3].

Tekoälyä ohjaavia eettisiä suosituksia, ohjeistuksia ja periaatteita on kritisoitu siitä, että niiden käytännön vaikutus on vähäinen. **Ohjeiden noudattaminen on suunnittelijoille ja kehittäjille hankalaa, sillä ohjeet ovat tulkinnanvaraisia, eivätkä tarjoile yhtä ainoaa tekoälyn eettisyyttä määrittävää periaatetta tai riittävää mittaria**. Ne eivät myöskään näytä juurtuvan politiikkatoimenpiteisiin. Lähinnä ne mahdollistavat ns. etiikkapesun yrityksissä.

Eettisiä periaatteita voidaan käyttää lähinnä kompassina etsittäessä tapoja ymmärtää tekoälyn etiikkaa. Niiden nimeäminen helpottaa asioiden punnintaa, mutta ne eivät aina vie tekoälyajattelua kovin pitkälle. Suunnittelijalle voi tulla eteen tilanteita, joissa periaatteet ovat ristiriidassa keskenään. Yksi syy tähän on se, että **teknologian etiikkaa koskevat kysymykset ovat aina tilannesidonnaisia. Teknologian vaikutukset eivät koske pelkästään suoria käyttötilanteita, vaan ovat käyttötilanteen kautta**

---

[2] AI HLEG (2019). Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, European Commission, April 2019.
[3] OECD (2017). Recommendation of the Council on Open Government. Paris: OECD.

sidoksissa erilaisiin sidosryhmiin, joilla saattaa olla ristiriitaisia tavoitteita. Eri vaakakupeissa voivat olla myös yhteiskunnan ja yksilön etu. Samoin yksilön oikeus päätöksentekoon voi sotia jonkun toisen etua vastaan. Siksi etiikan tarkastelu vaatii laajempaa keskustelua tekijöistä, tavoitteesta ja käytön kontekstista sekä eettisen dilemman analysointia yhdessä oleellisten sidosryhmien, suunnittelijoiden ja eettisten asiantuntijoiden kanssa. Tämä auttaa ymmärtämään, mitkä eettiset periaatteet ja arvot määrittävät kyseisen tekoälysovelluksen rajat.

Tekoäly muutosvoimana

Tekoäly haastaa luottamusta

Kun julkishallinnossa kehitetään tekoälyratkaisuja, otetaan samalla vahvasti kantaa siihen, millaista tulevaisuuden yhteiskuntaa halutaan rakentaa. Mikään muu teknologia ei ole ihmiskunnan historiassa vastaavalla tavalla horjuttanut kansalaisten luottamusta samalla kun se on avannut huikeita kehitysnäkymiä ihmisen hyvinvoinnin parantamisesta. Keskustelut heijastavat kasvavaa huolta ihmisen hallinnan menettämisestä tekoälykehityksessä: tavasta, jolla päätöksentekoalgoritmit voivat heikentää julkisen päätöksenteon legitiimiyttä ja yhteiskunnallista solidaarisuutta, lisätä maailmanlaajuista vallan ja vaurauden epätasapainoa ja syventää eriarvoisuutta.

Tekoälyjärjestelmien kehittämistä voidaankin pitää poliittisena, valtaa ja legitiimiyttä käsittelevänä kysymyksenä. Kyse on sekä kansalaisten ja hallitusten välisestä, että maallikoiden ja asiantuntijoiden välisestä luottamuksesta oleellisena yhteiskunnallisena haasteena. Luottamus on se kulmakivi, jonka varaan tekoälysovellusten kehittäminen ja käyttö yhteiskunnassa tulisi rakentaa. **Se rakentuu kuulemisesta ja osallisuudesta, mukaan ottamisesta, vaikuttamisesta ja demokratiasta. Ilman luottamusta julkiseen päätöksentekoon, tahtoon ja harkintaan ei tekoälyn potentiaalia voida täysin hyödyntää.** Täten tekoäly haastaa paitsi ihmiskunnan hyvän toteutumisen ja ihmisten kukoistamisen, myös **tavat, joiden avulla sen kehitystä koordinoidaan ja hallinnoidaan.** Siksi on hyvä kysyä, miten nykyisiä hallintorakenteita ja -välineitä voisi kehittää edistämään luottamuksen rakentumista ja tukemaan yhteisen ymmärryksen luomista tekoälyn haasteista ja ratkaisuista yhteiskunnassa.

Lisää näkemystä kokonaisuuden hallintaan

Kuten edellä on mainittu, tekoälyn hyödyntämisessä painottuvat sosiaalisiin konteksteihin ja ihmisten käyttäytymiseen liittyvät ongelmat. Euroopan komission asettama High-level Expert Group on Artificial Intelligence (AI HLEG) toteaa, että koska tekoäly voi parhaimmillaan olla ihmiskunnan kukoistusta tukeva teknologia, juuri yhteiskunnallisesta näkökulmasta tulisi ymmärtää, miten tekoälyä voidaan hyödyntää, jotta sillä edistetään ihmisen hyvää ja parannetaan ihmisten elämänlaatua. Siksi erityisesti julkishallinnon alalla tarvitaan tutkimusta, joka tunnistaa **tekoälykehityksen systeemisen luonteen tuo vaikutusten ennakoivan arvioinnin** mukaan kehitykseen.

Tekoälyn käytön vaikutuksia yhteiskunnassa, julkisessa hallinnossa ja etenkin kunnissa, on toistaiseksi selvitetty vain vähän, osittain siitä syystä, että tekoälyn hyödyntäminen on vasta suunnittelun tasolla, eikä sitä ole laajamittaisesti vielä otettu käyttöön. Osittain tästä syystä nykyinen tekoälyn sääntely ja ohjaus on epäjohdonmukaista ja hajanaista. Sääntelyviitekehys peilautuu ratkaisujen kehittämiseen; **hyödyntämisen esteenä ovat muun muassa liiallinen varovaisuus, osaamisen puute ja epätietoisuus tavoista hyödyntää tekoälyä eettisesti ja vastuullisesti**. Monimutkaisia ongelmia pyritään ratkaisemaan tekoälyteknologioiden avulla. Haluttujen vaikutusten aikaansaamista kuitenkin hidastaa se, ettei kokonaisuutta hallita ja toiminnan reunaehdoista vallitsee epätietoisuus. **Nämä puutteet hidastavat uusien teknologioiden käyttöönottoa ja hyödyntämissä sekä julkishallinnossa että kaupallisessa innovaatiossa. Kuntasektorilla koetaan tekoälyn käyttöönottoa hidastaviksi tekijöiksi yhtäältä sääntelyn ja toisaalta ymmärryksen, osaamisen ja koulutuksen puute sekä datan puutteellinen laatu.** Myös vähäiset resurssit kunnissa sekä tekoälykeskustelun nostamat eettiset huolenaiheet tunnistetaan hidastavan käyttöönottoa. Kunnissa ei juurikaan ole järjestetty koulutusta tekoälyjärjestelmien hyödyntämiseen, ja osaamistarpeen ja resurssien suuntaamisen välillä koetaan olevan ristiriitaa. Edelläkävijäorganisaatioiksi mielletään ne harvat, jotka ovat panostaneet koulutukseen.[4]

Lainsäädännön rinnalle pehmeää ohjausta

Lainsäädännön ja ns. ”kovan” sääntelyn rinnalle tarvitaan erilaisia käytännön ajattelutyökaluja ja käsitteellisiä kehikkoja, jotka auttavat yritys- ja julkisia toimijoita ajattelemaan moniulotteisia vastuullisuuteen ja etiikkaan liittyviä kysymyksiä. Positiivisen eettisen vaikutuksen maksimoimiseksi **tarvitaan eettistä johtamista ja innovatiivisia työkaluja ja kehitysalustoja tulevaisuuden tarpeiden ja tekoälyn vaikutusten ymmärtämiseen.**

Vastuullinen tekoälyn kehittäminen edellyttää myös kansalaisyhteiskunnan integroimista aidosti tekoälykehitykseen. Niin viime aikoina esitetty tekoälyä käsittelevä kirjallisuus kuin esimerkiksi OECD (the Council on Open Government) patistavat panostamaan avoimuuteen ja sidosryhmien osallistamiseen tekoälykeskustelussa. **Sidosryhmien osallistamisen päätöksentekoon nähdään lisäävän hallinnon vastuuvelvollisuutta, laajentavan kansalaisten vaikutusmahdollisuuksia ja vaikutusvaltaa päätöksenteossa, kehittävän kansalaisvalmiuksia, parantavan päätöksenteon tietopohjaa, vähentävän täytäntöönpanokustannuksia ja hyödyntävän laajempia verkostoja innovaatiotoiminnassa ja poliittisessa päätöksenteossa.** [5]

OECD:n luottamusarvioinnin mukaan[6] Suomessa on korkea luottamus julkiseen hallintoon, mutta kokemus vaikutusmahdollisuuksista ja päätöksenteon ymmärtämisestä on suhteellisen matala. Avoimuutta sekä osallistavampaa valmistelua ja päätöksentekoa korostava toimintatapa edellyttää oikeanlaista organisaatioiden eettistä agendaa, sen varmistamista, että tietämys ja eettinen visio tekoälystä on oikealla

[4] ETAIROS-tutkimushanke
[5] OECD (2020). Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave.
[6] OECD (2021). Drivers of Trust in Public Institutions in Finland.

tasolla. Ennen kaikkea julkiselta toimijalta edellytetään uskottavuutta, hyvin perusteltuja, realistisia ja toteuttamiskelpoisia näkemyksiä.

Suunnittelijan vastuu on suuri

Tekoälyn eettisen suunnittelun ja suunnitteluosaamisen voidaan helposti nähdä erilaisten ”teknisten” osa-alueiden huomioimisena, ja keskittyä arvioimaan lähinnä järjestelmän **läpinäkyvyyttä** (transparency) ja **vastuuvelvollisuuden** toteutumista (accountability). Tietoisuuden herättäminen kansalaisten keskuudessa on avainkysymys eettisesti kestävän tekoäly-yhteiskunnan kehittämisessä, ja **selitettävyyttä** (explainability) voidaan pitää kykynä tunnistaa ne avaintekijät, jotka vaikuttavat algoritmin päätöksentekoprosessiin. Oikeudellisessa keskustelussa korostuu näiden ”teknisten” osa-alueiden ja erityisesti vastuuvelvollisuuden huomioiminen nostamalla esiin teknisten menetelmien, teollisuusstandardien, sertifikaattien, auditointimenetelmien ja vaikutusten arvioinnin merkitystä. **On kuitenkin tärkeää huomata, että kysymystä algoritmisen päätöksenteon toivottavuudesta yhteiskunnassa ei tulisi käsitellä yksin vastuuvelvollisuuden kautta. Kysymys algoritmisten päätöksentekojärjestelmien laajemmasta vastuullisuudesta (responsibility) on yhä tärkeämpi tekoälyn etiikassa.**

Vastuullinen, tarkastelukulma perustuu toisiaan täydentäviin näkökohtiin: **Ennakointiin**, jossa mietitään kehitys- ja innovaatiotoiminnan sekä tavoiteltuja että tahattomia seurauksia; **oman tekemisen reflektointiin,** jossa pohditaan niitä kehitystyön taustalla olevia olettamuksia ja sitoumuksia, jotka voisivat vaikuttaa kehitystyöhön; **sitouttamiseen**, jossa otetaan relevantit sidosryhmät mukaan avoimeen vuoropuheluun tekoälykehityksen halutuista ja ei-halutuista seurauksista ja taataan haavoittuvien ja heikoimmassa asemassa olevien ihmisten tasa-arvoinen osallistuminen; sekä v**astaanottavuuteen ja reagointiin**, jossa muokataan työtä oikeaan suuntaa saadun palautteen, näkemysten ja arvojen mukaisesti. Näiden prosessien avulla on mahdollista ymmärtää, miten suunnittelija tai kehitystiimi pystyy yhtäältä ennakoimaan eettisiä ristiriitoja ja toisaalta tarvittaessa muuttamaan työnsä etenemistä kehitysprojektin aikana. [7]

Viestit päättäjille ja kansalaisille

Tarvitaan yhteinen koordinoitu käsitys vastuullisesta tekoälystä. Koska tekoälyn soveltaminen ei ole ainoastaan ns. tekninen kysymys vaan laajemmin yhteiskunnallinen oikeudenmukaisuus ja valtakysymys, se nostaa keskeisenä yhteiskunnallisena haasteena esille luottamuksen kansalaisten ja julkisen hallinnon sekä maallikoiden ja asiantuntijoiden välillä. Hallinnon aloilla on oltava **yhteinen ja koordinoitu käsitys tekoälyn toteuttamisesta ja sääntelystä**. Tämä edellyttää koordinointia eri hallintotasojen välillä, sekä sosiaalisia ja teknisiä rakenteita, jotka takaavat avoimuuden, eri näkökulmien huomioon ottamisen, toiminnan ennustettavuuden ja korjattavuuden. Tekoälyyn kohdistuvaa luottamusta voidaan edistää myös kuntien itsesääntelyn avulla.

[7] FCAI Ethical Advisory Board. Ethics Exercise Tool

Tarvitaan julkisen, yksityisen ja kolmannen sektorin yhteistyön tehostamista. Tekoälyjärjestelmien kehittäminen ja soveltaminen edellyttävät tietoisuuden lisäämistä tekoälykehityksen systeemisestä luonteesta, toimijoiden ja yhteiskuntien keskinäisistä riippuvuuksista ja jännitteistä maailmanlaajuisesti. Vastuuta reiluudesta ja oikeudenmukaisuudesta ei voida asettaa vain yksilöiden tai yksittäisten organisaatioiden hartioille. Tekoälyn eettinen käyttö vaatii kollektiivista yhteistyötä ja jatkuvaa keskustelua. Käytännössä tämä voi tarkoittaa esimerkiksi tarvetta luoda **uudenlaisia sosiaalisia sopimuksia**, jotka huomioivat algoritmisen päätöksenteon vaikutukset tulevaisuudessa.

Tarvitaan kansalaisten osallisuutta. Kansalaisten roolin korostaminen aktiivisina kehittäjinä edellyttää viestinnän, vastavuoroisuuden, osallisuuden ja avoimuuden lisäämistä yhteiskunnan, kansalaisten ja erilaisten data- ja palveluekosysteemien välillä. Julkishallinnon tehtävänä on herättää, mahdollistaa ja ylläpitää kansalaiskeskustelua. Tämä voi toteutua esimerkiksi digitaalisten **kansalaisalustojen** organisoinnin avulla, joissa kansalaisilla on mahdollisuus ymmärtää ja analysoida mitä algoritmit todella tekevät ja millaisia vaikutuksia tällä tekemisellä on yhteiskuntaan. **Sidosryhmäpaneelien** avulla voidaan osallistaa sukupuolen, kulttuuritaustan tai osaamisen suhteen monimuotoisia ryhmiä keskusteluun. Kansainväliset kokemukset vuosien ajalta näistä technology assessment -paneeleista ovat olleet positiivisia.

Tarvitaan lisää empiiristä ja selittävää tutkimusta. Tekoälyn käyttötapausten systemaattinen tarkastelu eettisestä näkökulmasta on silmiä avaavaa ja tuo eteemme eettisiä ulottuvuuksia, joita muuten olisi mahdotonta hahmottaa. Tätä kontekstuaalista ymmärrystä tukemaan tarvitaan lisää empiiristä ja selittävää tutkimusta tekoälyn käytöstä, läpinäkyvyyttä tutkimustietojen jakamisessa, sekä tietoa parhaista käytännöistä käytön riskien hallitsemiseksi. Erityisesti hallintoalalla tarvitaan tutkimusta, joka tunnistaa eri yhteisöjen kulttuuriset herkkyydet ja tuo nämä mukaan tekoälykehitykseen.

Tarvitaan koulutukseen panostamista. Suunnittelun ja suunnitteluosaamisen kehittäminen on oleellista. Lisäksi tarvitaan tekoälyn etiikan ymmärtämisen lisäämistä **kaikilla kansalaistasoilla** esimerkiksi monialaiseen koulutukseen perustuvan, yleisen kansalaistaidon avulla, joka kannustaa itsesääntöleviin toimiin.

Tarvitaan fokuksen muuttamista. Sorrumme tällä hetkellä helposti etsimään tekoälyn eettisiä **ongelmia**, vaikka pitäisi kysyä, miten tekoälyä voidaan maksimaalisesti hyödyntää, jotta sillä edistetään ihmisen hyvää ja parannetaan ihmisten elämänlaatua. Tekoälyratkaisujen ennakointia tulisi siksi parantaa ja **kehittää arvoherkkää ja moninäkökulmaista teknologian hyödyntämisen yhteiskunnallista arviointia**. Tässäkin sidosryhmien osallistaminen mahdollisimman laajasti keskusteluun tekoälyn toivotuista vaikutuksista on olennainen tekijä merkityksellisen lopputuloksen aikaan saamiseksi.