

Sisu Suominen

**EUROOPAN UNIONIN TEKÖÄLYSÄÄDÖKSEN
VAIKUTUS ALGORITMIEN MUSTIIN
LAATIKOIHIN**

Kandidaatintyö
Informaatioteknologian ja viestinnän tiedekunta
Tarkastaja: Mikko Nurminen
Lokakuu 2025

TIIVISTELMÄ

Sisu Suominen: Euroopan unionin tekoälysäädöksen vaikutus algoritmien mustiin laatikoihin
Kandidaatintyö
Tampereen yliopisto
Tietotekniikka
Lokakuu 2025

Euroopan unionin tekoälysäädös on vuonna 2024 asetettu laki, jonka tarkoituksena on tekoälyn vastuullinen kehitys ja käyttö Euroopan unionissa. Sädös sitoo kaikkia toimijoita, jotka käyttävät tai tuovat markkinoille tekoälyjärjestelmiä Euroopan unionin alueella. Tekoälyteknologian nopean kehityksen ja yhä useampien sovelluskohteiden myötä on kyseinen teknologia vaatinut myös uutta lainsäädäntöä. Tekoälysäädös onkin yksi maailman ensimmäisistä tekoälyä kattavasti käsittelevistä laeista.

Tämä työ on kirjallisuuskatsaus EU:n tekoälysäädökseen, jossa tarkoituksena on arvioida säädöksen toimivuutta automatisoidun päätöksenteon näkökulmasta. Automatisoidulla päätöksenteolla tarkoitetaan esimerkiksi pankin tekoälylle ulkoistamia lainapäätöksiä tai tekoälyn käyttöä oikeusjärjestelmässä. Tekoälyn päätöksentekologiikka on kuitenkin usein vaikea jäljittää ja hahmottaa; algoritmi on niin sanottu musta laatikko.

Työssä perustellaan automatisoidun päätöksenteon läpinäkyvyyden merkittävyys ja esitellään tekoälysäädöksen keskeiset toimintaperiaatteet, jonka jälkeen työssä tuodaan esille kirjallisuuskatsauksen puitteissa huomattuja ongelmakohtia säädöksen toimivuudessa. Kirjallisuuskatsauksen tuloksena tekoälysäädös jättää merkittävästi tulkinnanvaraisuutta, eikä se määrittele selkeitä teknisiä vaatimuksia, joihin tekoälyjärjestelmien kehittäjien tulisi sitoutua. Työssä käsitellään myös yritysten mahdollisuuksia vaikuttaa siihen, miten säädöstä lopulta tulkitaan käytännössä.

Avainsanat: EU AI Act, tekoälysäädös, tekoäly, automatisoitu päätöksenteko

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

TEKOÄLYN KÄYTTÖ OPINNÄYTTEESSÄ

Opinnäytteessä on käytetty tekoälysovelluksia:

- Ei
- Kyllä

Ilmoitukseni mukaan olen käyttänyt opinnäytteessäni tutkielmaproessin aikana seuraavia tekoälysovelluksia: ChatGPT-5.

Käyttötarkoitus: Otsikoiden muotoilu, aiheen rajausta ja suomennokset.

Osiot, joissa tekoälyä on käytetty: Otsikoiden muotoilu olemassaolevan tekstin pohjalta, sekä joidenkin englanninkielisten lähteiden osien ja termien suomennos oman tekstin tuottamisen avuksi. Tekoälyn tuottamaa tekstiä ei ole sellaisenaan kopioitu työhön otsikoiden ulkopuolella. Tekoälyä on käytetty työn alkuvaiheessa aiheen rajauksen ideointiin generoimalla joitakin esimerkkejä erilaisista mahdollisista tutkimuskysymyksistä. Lopullinen tutkimuskysymys ei ole tekoälyn luoma.

Olen tietoinen siitä, että olen täysin vastuussa koko opinnäytteeni sisällöstä, mukaan lukien osat, joissa on hyödynnetty tekoälyä, ja hyväksyn vastuun mahdollisista eettisten ohjeiden rikkomuksista.

SISÄLLYSLUETTELO

| | | |
|-----|---|----|
| 1. | Johdanto | 1 |
| 2. | Tekoälyn selitettävyyden merkitys ja sen haasteet. | 3 |
| 2.1 | Ehdotettuja keinoja tekoälyn toiminnan selkeyttämiseen | 4 |
| 3. | Yleiskatsaus Euroopan unionin tekoälysäädökseen | 5 |
| 3.1 | Säädöksen taustaa | 5 |
| 3.2 | Säädöksen tavoitteet ja toimintaperiaatteet | 5 |
| 3.3 | Tekoälysäädöksen tukena toimiva lainsäädäntö | 7 |
| 4. | Säädökseen toimivuuteen liittyvät haasteet | 8 |
| 4.1 | Säädöksen lakitekkinen lähestymistapa ja standardeihin nojaaminen | 8 |
| 4.2 | Säädöksen monitulkintaisuus | 9 |
| 4.3 | Riskitasojen määrittelystä vastaava taho | 9 |
| 4.4 | Yhteenveto tekoälysäädöksen tavoitteista ja haasteista, sekä tulevaisuuden kehityssuunnat | 11 |
| 5. | Keskustelu ja pohdinta | 13 |
| 6. | Yhteenveto | 15 |
| | Lähteet | 17 |

1. JOHDANTO

Tekoälyä sovelletaan yhä useammassa käyttökohteessa aina asiakaspalvelusta terveydenhuoltoon ja taloushallintoon. Vuonna 2024 jopa yli 40 prosenttia suurista yrityksistä Euroopan unionin alueella kertoi käyttävänsä ainakin yhtä tekoälyteknologiaa. Yli viidesosa suurista yrityksistä hyödyntää tekoälyä työtehtävien automatisoinnissa ja päätöksenteossa. [3] Tekoälyn yleistyminen herättää kuitenkin kysymyksiä siitä, kuinka luotettavia algoritmien tekemät päätökset ovat ja mihin ne pohjautuvat.

Koneoppimisalgoritmien toiminta muuttuu jatkuvasti yhä monimutkaisemmaksi ja tekoälyn opetusdatasta muodostaman mallin hallitsemiseen on vähän keinoja. Sen sijaan että monimutkaisia malleja yritettäisiin ymmärtää, on vallitsevana tapana ollut antaa tekoälyn toimia itsenäisesti, kunhan algoritmi vaikuttaa toimivan oikein. Tämä lähestymistapa on kuitenkin muutoksessa tekoälyn levitessä yhä useammalle alalle. [7] Useat asiantuntijat ovat alkaneet ajamaan eteenpäin ajatusta selitettävästä tekoälystä (*explainable AI* tai lyhyemmin *XAI*): mallista, jonka päätöksenteon logiikka voidaan jäljittää. Tekoälymallien toiminnan selkeyttäminen ei kuitenkaan ole yksinkertaista, kuten tässäkin työssä myöhemmin todetaan. Lisäksi tulee pohtia, onko yrityksillä oikeus pimittää algoritmien toiminta markkinaedun takaamiseksi.

Tekoälyyn liittyvät huolet ovat vaivanneet asiantuntijoita ja päätöksentekijöitä. Tämä johdi lopulta siihen, että vuonna 2021 Euroopan komissio teki aloitteen tekoälyä koskevasta lainsäädännöstä. Tämä on osa Euroopan unionin strategiaa, jolla pyritään edistämään digitaalisten teknologioiden, kuten tekoälyn, turvallista käyttöä. [14] Kolmen vuoden mittaisen valmisteluprosessin jälkeen Euroopan unionin lainsäädännöstä vastaava laitos, eli Euroopan parlamentti, hyväksyi asetuksen 2024/1689. Tämä on niin kutsuttu EU:n tekoälysäädös (*EU Artificial Intelligence Act, EU AI Act*), joka on osa koko Euroopan unionia koskevaa lainsäädäntöä ja samalla yksi maailman ensimmäisistä tekoälylakeista [14]. Tämä tekoälysäädös pyrkii vähentämään tekoälyn käyttöön liittyviä riskejä erityisesti kansalaisten turvallisuuteen, terveyteen ja yksityisyyteen liittyvissä sovelluskohteissa. Säädöksen tavoitteet ovat kuitenkin laajat:

"Tämän asetuksen tarkoituksena on parantaa sisämarkkinoiden toimintaa vahvistamalla yhtenäinen oikeudellinen kehys erityisesti unionin arvojen mukaiselle tekoälyjärjestelmien kehittämiselle, markkinoille saattamiselle, käyttöönnotolle ja käytölle unionissa, edistää ih-

miskeskeisen ja luotettavan tekoälyn käyttöönottoa varmistaen samalla – – terveyden, turvallisuuden ja perusoikeuksien, mukaan lukien demokratia, oikeusvaltio ja ympäristönsuojelu, suojelun korkea taso, suojautua tekoälyjärjestelmien haitallisilta vaikutuksilta unionissa sekä tukea innovointia." [1]

Säädökseen liittyy kuitenkin yhä epävarmuutta, eikä sen toimintaa olla vielä todistettu käytännössä. Tekoälysäädös astui osittain voimaan kesällä 2024, josta alkoi myös kahden vuoden siirtymävaihe [14]. Säädös on siis kokonaisuudessaan sitova vasta kesällä 2026.

Tämän työn tavoitteena on tuoda esille tekoälyn käytön huolia päätöksenteossa, sekä vastata kysymykseen: Kuinka Euroopan unionin tekoälysäädös vastaa tekoälyn päätöksenteon selitettävyyteen liittyviin haasteisiin, sekä minkälaiset puitteet se tarjoaa korkean riskin tekoälyjärjestelmien vastuulliselle käytölle? Keinoja algoritmien valaisemiseen käsitellään erityisesti luvussa 2. Luku 3 puolestaan esittelee tekoälysäädöksen toimintaperiaatteita ja keinoja edistää säädöksen tavoitetta teknologian läpinäkyvyyden kannalta. Tämän jälkeen luvussa 4 tuodaan esille ongelmakohtia, jotka näyttävät ettei tekoälysäädös pysty vielä luotettavasti kattamaan edellä esitettyjä tavoitteita.

Tämä työ on toteutettu kirjallisuuskatsauksena keskittyen 2020 jälkeen käytyyn akateemiseen keskusteluun aiheesta. Lähteitä on etsitty eri internetin tietokannoista, jotka keskittyvät tieteellisiin julkaisuihin. Se, ovatko lähteet vertaisarvioituja, on tarkistettu Tampereen yliopiston Andor-tietokannasta.

Alkuvaiheessa kun aihetta lähdettiin määrittelemään, lähteitä ja aihetta kartoitettiin Google Scholarin kautta. Hakusanoina oli aluksi "algorithmic bias" ja "black box algorithm". Koska aihetta oli tarkoitus alusta asti tarkastella enemmän eettisestä ja yhteiskunnallisesta näkökulmasta, puhtaasti tekniset julkaisut jätettiin pois.

Aiherajauksen tarkentuessa hakuja tehtiin ensin Google Scholariin hakusanoilla "EU AI Act". Löydetyistä lähteistä valittiin aihetta kokonaisvaltaisesti käsittelevät julkaisut, eli esimerkiksi yksittäisiä tapauskohtaisia tarkasteluja ei tässä kohtaa valittu mukaan. Ensimmäisistä lähteistä valittiin mukaan myös termi "XAI (eXplainable AI), joka liittyy vahvasti aihepiiriin. Tällä termillä tehtiin uusia hakuja SpringerLink -tietokantaan tarkoituksena etsiä julkaisuja, joissa aiheena oli eettinen näkökulma, AI:n regulointi ja EU AI Act. Hakuja SpringerLinkin kautta tehtiin myös yhdistämällä XAI + EU AI Act, sekä rajaamalla XAI haut Philosophy of Artificial Intelligence -kategoriaan.

Lähteistä on rajattu pois kirjallisuuskatsaukset, kyselytutkimukset sekä julkaisut, joissa aiheetta on käsitelty yksinomaan teknisestä näkökulmasta. Lähteitä on valikoitu tutkimuksen alussa otsikon, julkaisufoorumien sekä tiivistelmän perusteella. Tutkimuksen edistytessä lähteisiin perehdyttiin kokonaisvaltaisemmin, jolloin joitakin lähteitä karsiutui pois muun muassa sen vuoksi, että ne käsitelivät säädöstä aiheen ohi menevästä näkökulmasta.

2. TEKOÄLYN SELITETTÄVYYDEN MERKITYS JA SEN HAASTEET

Uudet teknologiat ovat aina herättäneet pelkoa ja vastalauseita. Tekoälyn yleistymiseen ja laajaan käyttöön liittyy kuitenkin paljon aiheellista kritiikkiä. Tekoälyn hyödyntäminen on levinnyt yhä useammalle alalle, kuten logistiikkaan [3], rajojen valvontaan [18] ja lain-säädännön laatimiseen [16]. Riskialttiit käyttökohteet tekevät algoritmin toiminnan oikeel-lisuudesta, datan eheydestä, sekä tekoälyn käytön läpinäkyvyydestä ensisijaisen tärkeitä [16].

Akateemisessa keskustelussa tekoälyalgoritmien ymmärrettävyys on noussut esille yhä uudelleen. Tekoälylle ulkoistettua päätöksentekoa kritisoidaan erityisesti sen läpinäkyvyy-den puutteesta [7]. Mikäli tekoälyn toimintaan ei voi luottaa pikaruokaravintolan autokais-talla [6], voidaanko sen uskoa pystyvän tekemään päätöksiä oikeusjärjestelmässä? Rea-lisoituneet virhetapaukset ovat osoittaneet, että huolet tekoälyn käytöstä ovat aiheellisia: erittäin saastunut ilma on todettu turvalliseksi hengittää, ja ehdonalainen on evätty vir-heellisesti tekoälyn toimesta [17].

Tekoälyn mustia laatikoita ei kuitenkaan ole yksinkertaista kammata auki. Tekoälyn hyö-dyntämät mallit ovat monimutkaisia, eivätkä mallinnukset ole välttämättä muodossa, joka olisi ihmisen ymmärrettävissä. Lisäksi suuri osa tekoälyalgoritmeista on yksityisyriyten omaisuutta, jolloin niiden toimintaa ei haluta paljastaa algoritmin tuoman markkinaedun vuoksi. [4] [17] Tekoälyteknologia sekä sitä koskeva tietotaito onkin vahvasti keskittynyt-tä juuri yrityssectorille [4]. Tämä aiheuttaa huolta suurten yritysten kyvyistä väärinkäyttää tekoälyalgoritmeja omien etujensa mukaisesti [15]. Erityisesti nämä syyt ovat nostaneet tekoälyn selitettävyyden esille alan keskusteluissa.

Selitettävyyden haasteena on kuitenkin se, että jo yksinkertaiset koneoppimisen mallit muodostuvat nopeasti liian monimutkaisiksi ymmärtää. Ne koostuvat useista päällekkäi-sistä algoritmeista, tuhansista päätöksentekoon liittyvistä rakenteista, sekä yhdistelmis-tä satoja tuhansia tai miljoonia datapisteitä. [7] Lisäksi tekoälyn toimintamallit ovat usein epäintuitiivisia [15], eivätkä ne käsittele dataa muodossa, joka olisi ihmisen ymmärrettä-vissä. Asiantuntijat pyrkivät kuitenkin keksimään keinoja valaista tekoälyn toimintalogiik-kaa.

2.1 Ehdotettuja keinoja tekoälyn toiminnan selkeyttämiseen

Algoritmin tapahtumien kirjaaminen lokiin:

Yksi mahdollinen ratkaisu algoritmin toiminnan selkeyttämiseen on pitää lokia perusteluista, joiden avulla algoritmi on päätyneet kuhunkin päätökseen. Tällöin algoritmin niin sanottu ajatuksenkulku olisi mahdollista jäljittää tarkasteltaessa mallin antamia päätöksiä. Lokijärjestelmä ei ole vain toimivaksi todettu, vaan myös verrattaen yksinkertainen ja halpa implementoida. [7]

Lokitapahtumien kirjaaminen ei ole kaikkien tekoälymallien kohdalla kuitenkaan mahdollista. Tekoälymallien eräs suuri haaste on se, että monimutkaisemmat mallit kehittävät ja muuttavat toimintaansa itsenäisesti. Tällöin ei voida etukäteen määritellä, mitkä päätökseen vaikuttavat parametrit lokiin tulisi kirjata, sillä ne voivat vaihdella jatkuvasti. Jos sopivien parametrien kirjaaminen puolestaan jätetään tekoälyalgoritmille itselleen, soti se läpinäkyvyyden ideaa vastaan. [7]

Lisäksi pienetkin muutokset tekoälymalliin voivat muuttaa tekoälyn tekemiä päätöksiä. Tämän vuoksi tulisi pystyä ensin selvittämään, miten eri parametrit ja niiden eri yhdistelmät vaikuttavat mallin toimintaan. Parametrien painoarvoa on mahdollista testata analysoimalla, miten tekoälyn antama data muuttuu eri parametrien arvoilla. Tekoälymallien monimutkaisuudesta johtuen tämä vie kuitenkin paljon laskentatehoa. [15] Kaiken kaikkiaan pelkkien parametrien kirjaaminen ylös ei riitä selittämään tekoälyn päätöksentekotekniikkaa [7].

Alaa koskevan viitekehyksen kehittäminen:

Toinen mahdollinen ratkaisuehdotus on kehittää alaa koskeva viitekehys (*framework*). Viitekehys olisi standardeista ja yleisistä käytänteistä koostuva ohjeistus, jonka avulla tekoälysovelluksen voisi sertifioida. Tekoälymallin sertifiointi tarkoittaisi tällöin, että mallin tekemiä päätöksiä voidaan pitää oikeellisina, ja että tekoälysovelluksen suunnittelu, kehitys ja toiminta seuraa yleisesti hyvinä pidettyjä periaatteita. Näin myös tekoälystä vastuussa oleva taho saisi suojaa, mikäli tekoälyn tekemä päätös riitautettaisiin. [7]

Tällaisen viitekehyksen kehittäminen tulee esille useissa akateemisissa julkaisuissa. Tässä ratkaisussa piilee kuitenkin tekoälyn lainsäädännön yksi keskeisimmistä ongelmista: miten tekoälyn ymmärrettävyys toteutetaan teknisesti? Miten tekoälyn päätöksenteon oikeellisuus voidaan taata? Näihin kysymyksiin on vaikea löytää tyydyttäviä vastauksia. Tekoälysovelluksen sertifiointi kohtaa myös saman ongelman kuin muut IT-järjestelmien sertifiointit: se on yksinkertaisesti kallista ja aikaavievää [7]. Seuraavassa luvussa tarkastellaan, mitä keinoja EU:n tekoälylainsäädös hyödyntää ja voiko se vastata esitettyihin ongelmiin tekoälyn läpinäkyvyydestä.

3. YLEISKATSAUS EUROOPAN UNIONIN TEKOÄLYSÄÄDÖKSEEN

3.1 Säädöksen taustaa

Euroopan unioni on ottanut selkeän pyrkimyksen suojella kansalaisten oikeuksia myös digitaalisessa maailmassa [12]. Erityisesti yleinen tietosuoja-asetus (*General Data Protection Regulation, GDPR*) on toiminut suunnannäyttäjänä kansainvälisesti [19]. Jo GDPR:n mukaan Euroopan kansalaisilla on oikeus tietää, mitä tietoja heistä käsitellään ja miksi. Tämä tieto tulisi olla saatavilla pyynnöstä ymmärrettävässä ja selkeässä muodossa. [7]

Sopimus Euroopan unionin toiminnasta (*Treaty on the Functioning of the European Union, TFEU*) pitää myös sisällään periaatteen, että päätöksenteko tulisi pitää läpinäkyvänä. Läpinäkyvyys mahdollistaa kansalaisten kyvyn tarkastella päätöksentekoprosesseja, joka on tärkeää demokratian kannalta. On mahdollista perustella, että EU:n kansalaisten tulisi saada myös tarkempaa tietoa automatisoidusta päätöksenteosta.

Moni Euroopan maa oli aloittanut valmistelut tekoälyä koskevien lakien osalta ennen EU:n tekoälysäädöstä. Koko EU:n markkina-alueen laajuinen säädös nähtiin kuitenkin tarpeellisenä, sillä se mahdollistaa yhtenäisen toiminnan ja rajoittaa yritysten mahdollisuuksia valikoida, minkä maan lain mukaan toimia. Tekoälysäädös sitookin kaikkia EU:n markkinoille tulevia teknologioita riippumatta siitä, mihin maahan yritys on rekisteröity. [15]

3.2 Säädöksen tavoitteet ja toimintaperiaatteet

Tekoälysäädöksen laatijat kohtasivat merkittävän haasteen pyrkiessään hallitsemaan teknologiaa, joka kattaa hyvin erilaisia käyttökohteita ja jonka tulevaisuuden kehitys on vaikeasti ennustettavissa. Lisäksi säädös pyrkii yhdistämään kaksi erilaista lainopin alaa: kuluttajatuotteiden turvallisuuden (*product safety*) ja ihmisten perusoikeuksien turvaamisen. [2] Säädöksen onnistuminen tässä on herättänyt paljon keskustelua, johon palataan tarkemmin luvussa 4. Perehdytään ensin tekoälysäädöksen toimintaperiaatteisiin.

Tekoälysäädöksen tarkoitus on tekoälyteknologian luotettavuus, läpinäkyvyys ja turvallisuus. Tekoälysäädös ja GDPR ajavat monilta osin samoja periaatteita, kuten läpinäkyvän päätöksenteon ja kansalaisten oikeuden ymmärtää miten heidän dataansa käsitellään.

Toisin kuin GDPR, tekoälysäädöksen vaikutusalue ei koske vain henkilökohtaisten tietojen käsittelyä, vaan säädös kattaa ensisijaisesti tekoälyn käytön päätöksenteossa. Tekoälysäädöksen vaatimukset selitettävyydestä koskevat kuitenkin vain teknologioita, joiden arvioidaan voivan aiheuttaa suuren potentiaalisen riskin ihmisoikeuksille, turvallisuudelle ja yksityisyydelle. [13]

Säädös koskee sekä tekoälyn kehittäjiä, että kaikkia tahoja, jotka hyödyntävät tekoälyteknologioita muuhun kuin yksityiskäyttöön. Vaaditut toimenpiteet, joita kehittäjiltä ja tekoälyä käyttäviltä tahoilta vaaditaan, perustuvat tekoälyn potentiaalisiin riskeihin. Tekoälysovellukset luokitellaan säädöksessä neljään kategoriaan:

- IV Kielletyt tekoälyteknologiat
- III Suuririskiset tekoälyjärjestelmät
- II Avoimuuteen liittyvän riskin tekoälyjärjestelmät
- I Minimaalisen riskin tekoälyjärjestelmät

Kiellettyihin tekoälyteknologioihin luokitellaan sovellukset, joilla on liian suuret riskit ihmisten oikeuksien ja yksityisyyden kannalta. Tarkemmin sanottuna tekoälysäädös puhuu "kielletyistä tekoälyyn liittyvistä käytännöistä", joihin on kuitenkin olemassa poikkeuksia. Kategorian II avoimuuteen liittyvällä riskillä tarkoitetaan, että siihen lukeutuvien järjestelmien kohdalla tulee käyttäjälle olla selvää, että hän kommunikoi tekoälyn kanssa tai katsoo tekoälyn tuottamaa sisältöä. Esimerkkejä eri kategorioiden tekoälyjärjestelmistä on esitelty taulukossa 3.1. Taulukko ei kuitenkaan kata kaikkia sovelluskohteita.

| Kategoria | Esimerkkisovelluksia |
|---------------|--|
| Kategoria IV | Sosiaaliset pisteytysjärjestelmät (kuten Kiinan <i>social scoring</i>), biometrinen tunnistus ja kategorisointi, reaaliaikaiset tunnistusjärjestelmät julkisissa tiloissa |
| Kategoria III | Tekoäly jota käytetään kriittisen infrastruktuurin ylläpitoon, lääkinnällisissä laitteisiin, rajavalvontaan tai osana autoja |
| Kategoria II | Palvelubotit, synteettistä materiaalia tuottavat järjestelmät, syväväärnökset |
| Kategoria I | Roskapostisuodattimet, videopelit |

Taulukko 3.1. Esimerkkejä eri tekoälysäädöksen kategorioihin kuuluvista tekoälyjärjestelmistä.

Tekoälyteknologiat kategorisoidaan siis sen perusteella, kuinka vakavaa haittaa teknologia voi mahdollisesti aiheuttaa. Koska riskitasojen määrittäminen on yksi säädöksen keskeisistä haasteista, palataan siihen tarkemmin luvussa 4.3. Tekoälysäädöksen vaatimukset keskittyvät kategorian III suuren riskin järjestelmiin, joten myös tässä työssä keskitytään niiden tarkasteluun. Säädös asettaa näille järjestelmille useita vaatimuksia:

Riskinhallintajärjestelmä (artikla 9): Suuririskisille tekoälyjärjestelmille täytyy perustaa riskienhallintajärjestelmä, jota tulee ylläpitää ja päivittää tarpeen tullen koko järjestelmän käytön ajan. Tämä tarkoittaa, että järjestelmän aiheuttamat mahdolliset riskit tulee analysoida ja minimoida mahdollisuuksien mukaan.

Data ja datanhallinta (artikla 10): Järjestelmän opetusdatassa tulee huomioida mahdolliset vinoumat, jotka johtavat syrjintään, tai joilla on todennäköisiä vaikutuksia terveyteen ja turvallisuuteen.

Dokumentaatio (artikla 11): Järjestelmästä on laadittava tekninen dokumentaatio, jossa esitetään selkeässä ja ymmärrettävässä muodossa, että tekoälyjärjestelmä täyttää asetetut vaatimukset.

Lokitiedot (artikla 12): Lisäksi tekoälysäädöksessä hyödynnetään luvussa 2 esille tuotua ehdotusta lokitietojen keräämisestä, jonka tarkoituksena on varmistaa tekoälyn toiminnan jäljitettävyyttä. Kuten todettua, lokitietojen keräämiseen liittyy kuitenkin haasteita. Lisäksi säädös velvoittaa vain teknisten tapahtumien tallentamisen, joka ei vielä kuvaa sitä, miten algoritmi tekee päätöksiä.

Avoimuus ja tietojen antaminen käyttäjille (artikla 13): Erityisesti tekoälyn toiminnan avoimuuteen liittyy säädöksen artikla 13. Tämä artiklan keskeisenä periaatteena, on varmistaa tekoälyjärjestelmältä riittävä läpinäkyvyys tahoille, jotka ottavat tekoälyjärjestelmän käyttöön. [1][14] [10] Tekoälysäädös ei kuitenkaan määrittele, miten tämä on tarkoitus varmistaa teknologian näkökulmasta.

3.3 Tekoälysäädöksen tukena toimiva lainsäädäntö

Tekoälysäädöksen oli tarkoitus toimia harmoniassa erilaisten EU direktiivien kanssa, erityisesti tuotevastuudirektiivin (*Product Liability Directive, PLD*) ja tekoälyn vastuudirektiivin (*Artificial Intelligence Liability Directive*) kanssa [19]. Direktiivit ovat Euroopan unionin asettamia tavoitteita, joihin kukin jäsenmaa sitoutuu. Jäsenmaat joutuvat kuitenkin oma-toimisesti laatimaan paikallisen lainsäädännön tavoitteiden saavuttamiseksi. [20] Heinäkuussa 2025 EU kuitenkin ilmoitti luopuneensa tekoälyn vastuudirektiivin valmistelusta kokonaan [11].

Direktiivien ohella, säädöksen on tarkoitus toimia yhdessä eri teknologiastandardien kanssa. Standardit eivät kuitenkaan ole laillisia vaatimuksia, vaan alan yleisiä ohjenuoria, joissa myös lainsäädännön ulkopuolisilla tahoilla on sananvaltaa. Konkreettisten toimenpiteiden jättäminen standardien varaan onkin herättänyt kritiikkiä EU:n toiminnassa. Teknologian nopeassa kehityksessä standardit ovat kuitenkin päätyneet keskeiseen asemaan, sillä niiden aikaansaaminen on lainsäädäntöä nopeampaa. [12] Siirrytään nyt lukuun 4 ja standardoinnin herättämiin huoliin.

4. SÄÄDÖKSEEN TOIMIVUUTEEN LIITTYVÄT HAASTEET

Kuten jo luvussa 2 tuotiin esille, tekoälyjärjestelmien ymmärrettävyys ei ole yksinkertainen ongelma, johon olisi olemassa yksi selkeä ratkaisu. Tekoälynsäädös pyrkii asettamaan raamit sille, että tekoälyn käyttö Euroopan alueella olisi mahdollisimman turvallista. Tämän tavoitteen saavuttaminen riippuu kuitenkin siitä, kuinka hyvin säädös pystyy velvoittamaan eri tahoja kantamaan vastuun teknologiasta. Säädestä on kritisoitu erityisesti vaatimuksista, jotka jättävät runsaasti tulkinnanvaraa tekoälyä kehittävien tahojen harkintaan. [18] Tätä tarkastellaan tässä luvussa tarkemmin.

4.1 Säädöksen lakitekkinen lähestymistapa ja standardeihin nojaaminen

EU:n tekoälynsäädöksen tarkoitus on yhtenäistää Euroopan lainsäädäntöä ja ylläpitää yhtenäistä markkina-alueita. Erityisesti tekoälyn sääntely on kuitenkin monimutkaista, sillä vaikutusvaltaa on lainsäädännön lisäksi myös standardoinnista vastaavilla toimijoilla, tekoälyn toimitusketjuun osallistuvilla, sekä kuluttaja- ja perusoikeuksia suojelevilla tahoilla. [12]

Vaikka säädöksen tavoitteissa esitetään vahvasti pyrkimys turvata kansalaisten perusoikeuksia, on laki kuitenkin sisällöltään lähempänä tuoteturvallisuuden lainsäädäntöä. Lisäksi, kuten aiemmin mainittua, tekoälynsäädös on jättänyt tarkemmat määrittelyt eri teknisten standardien, joko nykyisten tai kehitteillä olevien, varaan. Tämä lähestymistapa on herättänyt keskustelua siitä, miten eettiset huolet voidaan huomioida standardoinnissa ja onko se ylipäätään mahdollista. [5]

Eri standardit ovat tieto- ja sähkötekniikan alalla historiallisesti merkittävässä asemassa, ja ne ovat edesauttaneet yhtenäisten järjestelmien kehittämistä. Asiantuntijat ovat kuitenkin kyseenalaistaneet standardien merkitystä ja painoarvoa lainsäädännössä. Standardien kehittämiseen osallistuu sekä voittoa tavoittelemattomia järjestöjä, eri maiden delegaatioita kuin myös suuryritysten edustajia [5]. Esimerkiksi ETSI (*European Telecommunications Standards Institute*) sallii yritysedustajien suoran osallistumisen. ETSI on riippumaton eurooppalainen standardointijärjestö, mutta sen jäseniin kuuluu myös Euroo-

pan unionin ulkopuolisia yrityksiä, kuten Amazon ja Huawei. Lisäksi ETSI antaa äänestysissä painoarvoa osallistuvien tahojen rahallisille kontribuutioille. [5] Tämä herättää kysymyksen voivatko yritykset ostaa vaikutusvaltaa, jolla muotoilla standardeja itselleen sopiviksi.

Tällä on merkittäviä vaikutuksia, sillä vaikka standardit eivät teknisesti ottaen ole sitovia lakeja, on Euroopan unioni tuomioistuin jo hyödyntänyt niitä oikeustapauksissa. Tämä tarkoittaa, että standardeista muodostuu epäsuorasti myös osa lakia. [5] Mikäli yrityksillä siis on mahdollisuus vaikuttaa merkittävästi standardien kehitysprosessiin, vaikuttavat ne myös lainsäädäntöön.

4.2 Säädöksen monitulkintaisuus

Tekoälysäädöksen laatiminen ei ollut suoraviivaista ja Euroopan parlamentti joutui käymään pitkiä keskusteluja saadakseen säädöksen lopulliseen muotoilun läpi [16]. Tekoälysäädöksen määrittelemättä jääneet sanavalinnat ja monitulkintaisuus ovat herättäneet kritiikkiä ja keskustelua, sillä jo itse tekoälyn lakitekkinen määrittely aiheutti ongelmia EU-tasolla [18].

Säädöksessä määritellään esimerkiksi se, että tekoälyalgoritmin käyttämällä opetusdatailla tulee olla "asianmukaiset tilastolliset ominaisuudet". Se mitä "asianmukaisella" tässä yhteydessä tarkoitetaan, jätetään kuitenkin tulkinnanvaraiseksi. Säädöksessä käytetyt termit, joille ei voida antaa tarkkaa teknistä määritelmää, ovat vaikeatulkintaisia ja niiden toteutumista on hastavaa arvioida. [18]

Epämääräisyys voi heikentää tekoälysäädöksen kykyä saavuttaa haluttu teknologian turvallisuuden taso. Mikäli vaatimuksia ei määritellä laissa selkeästi, nojaa lain tavoitteiden toteutuminen vahvasti standardointiin. Kuten edelle todettua, standardointiin pystyy vaikuttamaan monia eri tahoja. Mikäli standardointi ei heijasta lainsäätäjien alkuperäisiä tavoitteita, tulevatko ne lopulta toteutumaan kunnolla? Erityisesti huomioitaessa, että suurin osa tekoälysovelluksista tulee yksityisyrittäjiltä, jotka tavoittelevat ensisijaisesti markkinaetua. Tekoälysäädöksen seuraamisesta aiheutuu yrityksille lisäkuluja, joten uhkana on, että yritykset pyrkivät minimoimaan tarvittavat toimenpiteet.

4.3 Riskitasojen määrittelystä vastaava taho

Kuten luvussa 3 esiteltiin, tekoälysäädös luokittelee teknologiat neljään kategoriaan perustuen niiden riskeistä terveydelle, turvallisuudelle ja perusoikeuksille. Tekoälysovellusten onnistunut luokittelu on keskeistä säädöksen toimivuudelle, mutta haasteena kuitenkin on se, mikä taho riskitasojen määrittelystä vastaa.

Jo varhaisessa vaiheessa huomattiin, että tekoälyä kehittävät tahot todennäköisesti tah-

tovat välttää korkean riskitason luokittelun, sillä se aiheuttaa eniten vaadittuja toimenpiteitä [18]. Tästä aiheutuu yrityksille kustannuksia muun muassa lisääntyneiden työtuntien myötä, joten voidaan väittää yritysten olevan puolueellisia teknologian luokittelun suhteen. Tämänhetkiset puitteet kuitenkin implikoivat, että tekoälyn riskitason määrittäminen on sitä kehittäväen yrityksen vastuulla [18] [19].

Riskitasojen määrittelyssä ilmenee myös aiemmin todettu tulkinnanvaraisuuteen liittyvä ongelma. Vaikka tekoälyteknologiaa hyödynnettäisiin korkean riskitason tapauksissa, on mahdollista välttää kyseisen kategorian luokitus, mikäli teknologiaa käytetään esimerkiksi vain valmisteleviin tehtäviin [19]. Säädöksessä jätetään kuitenkin jälleen paljon harkinnanvaraisuutta sille, miten tällaiset tehtävät erotellaan tekoälyn muusta käytöstä. Mikäli tekoälystä vastaava taho toteaa, että teknologiaa ei käytetä riskialttiisiin tehtäviin, tulee sen tarjota vain yhteenveto perusteluista EU:n tietokantaan, joka ei ole julkisesti saatavilla. [18]

Dokumentaatio siitä, miten riskitaso on määriteltä, tulee pystyä esittämään viranomaisille tarvittaessa. Tämä aiheuttaa kuitenkin lisätyötä säädöksen valvomiselle ja on mahdollista, ettei viranomaisten resurssit riitä kyseenalaistamaan tehtyä riskiarviointia [19]. Tällöin markkinoille päätyy teknologiaa, jonka valvonta on puutteellista.

Kun nyt huomioidaan siis seuraavat seikat:

- yrityksen motivaatiot ovat lähtökohtaisesti korkean riskitason kategorisointia vastaan,
- tekoälyn riskiluokitus voidaan hoitaa yrityksen sisällä ja
- yrityksille annetaan jo lähtökohtaisesti mahdollisuus välttää tekoälyteknologian kuuluminen korkean riskitason kategoriaan,

herää epäily siitä, miten luotettavana teknologian luokittelua voidaan pitää ilman ulkoista valvontaa. On siis riskinä, että tämän varjolla korkean riskitason tekoälysovellukset onnistuvat välttämään oikean luokittelun ja täten myös sen tuomat velvollisuudet.

Lisähaasteena valvonnassa on tietotaidon painottuminen yksityisyrittäjien puolelle. Julkisella sektorilla on harvoin resursseja ja taitoa kehittää tekoälysovelluksia, vaan ne hankkivat käyttämänsä teknologian yrityksiltä esimerkiksi lisenssien avulla [4]. Tämä aiheuttaa huolia siitä, että julkiset toimijat ovat teknologian suhteen liian riippuvainen yksityisestä sektorista [18]. Riskitasot myös määritellään tekoälyteknologian aiotun sovelluskohteen mukaan; on kuitenkin vähän näyttöä siitä, että riskitaso voitaisiin määrittellä etukäteen vain tarkoituksen perusteella. [2]

4.4 Yhteenveto tekoälysäädöksen tavoitteista ja haasteista, sekä tulevaisuuden kehityssuunnat

Edellä käsiteltiin tekoälysäädöksen toimintaa ja ongelmakohtia laajemmin. Kuten luvussa 3 esiteltiin, tekoälysäädös asettaa monia tavoitteita suuririskisille tekoälyjärjestelmille. Alla olevaan taulukkoon 4.1 on koottu yhteenveto sekä tekoälysäädöksen vaatimuksista että niitä koskevista haasteista.

| Tekoälysäädöksen asettama vaatimus | Yhteys algoritmin läpinäkyvyyteen ja oikeellisuuteen | Käytännön haasteet |
|---|---|---|
| Riskinhallintajärjestelmä | Järjestelmän toimivuus testattava Riskien analysointi ja tunnistaminen | Ei määrittele konkreettisia menetelmiä algoritmin oikeellisuuden arviointiin |
| Data ja datanhallinta | Velvoittaa estämään mahdolliset vinoumat opetusdatassa, mikä edesauttaa algoritmin päätöksenteon oikeellisuutta | Miten datan eheydestä, oikeellisuudesta ja asiallisista tilastollista ominaisuuksista huolehditaan? Opetusdatan eheys ei ratkaise algoritmin ymmärrettävyyteen liittyviä haasteita |
| Lokitiedot | Lokien avulla voidaan seurata päätöksentekoprosessia | Toimiva implementoiminen ei mahdollista kaikille järjestelmille Säädös velvoittaa vain teknisten tapahtumien tallentamisen, mikä ei vielä kerro miten malli tekee päätöksen |
| Avoimuus ja käyttäjälle annettavat tiedot | Velvoittaa tarjoamaan käyttöönottajalle ymmärrettävää tietoa järjestelmän toiminnasta | "Ymmärrettävä tieto" jää tulkinnanvaraiseksi |
| Ihmisen suorittama valvonta | Ihmisellä oltava kyky ymmärtää ja seurata järjestelmän toimintaa, sekä mahdollisuus puuttua tekoälyn päätöksiin | Miten käyttäjä pystyy arvioimaan järjestelmän toimivan oikein? Miten tekoälyn päätösten oikeellisuutta tulkitaan? |

Taulukko 4.1. *Tekoälysäädöksen vaatimukset ja niihin liittyviä käytännön huolia.*

Säädöksen tavoitteet ovat monelta osin algoritmien ymmärrettävyyden puolella. Vaatimusten toteutuminen käytännössä tuo kuitenkin monia ongelmakohtia, joihin tekoälyjärjestelmiä kehittävä tahot joutuvat pureutumaan. Erityisesti nämä kysymykset jäävät tämän työn puitteissa avoimiksi:

- Onko olemassa tarpeeksi tietotaitoa ymmärtää tekoälyalgoritmien toiminta?
- Miten tekniset toteutukset tapahtuvat?

- Minkälainen prosessi tekoälyn standardointi on?

Näiden myötä sekä tekoälysäädöksen, että tekoälyä koskevan lainsäädännön tulevaisuuden kehitys tulee olemaan monimutkainen haaste, johon on useita eri lähestymistapoja. On todennäköistä, että muu maailma seuraa Euroopan unionia tekoälysäädöksen kannalta. Näin tapahtui muun muassa yleisen tietosuoja-asetuksen (GDPR) kanssa, joka on nyt noussut globaaliin asemaan [19]. Teknologiateollisuuden johtavat maat, kuten Yhdysvallat, Japani ja Etelä-Korea ovat ottaneet tekoälyn turvallisuuden ja säätelyn tavoitteeksi. Erityisesti Etelä-Korea asetti loppuvuodesta 2024 laajan tekoälyä koskevan säädöksen, joka seuraa EU:n tekoälysäädöksen jalanjäljissä. [9]

EU:n tekoälysäädös on myös Euroopassa vasta alkua tekoälysovelluksia koskevalle lainsäädännölle. Esimerkiksi Tanskan hallitus on aloittanut valmistelut tekoälyllä generoitujen väärennösten (*deepfake*) rajoittamiseksi määrittelemällä jokaiselle kansalaiselle eräänlaisen tekijänoikeussuojan omille piirteilleen. Tämän lain varalla henkilöllä olisi oikeus vaatia internetissä olevan, hänestä itsestään luodun, väärennöksen poistoa alustalta.[8]

Tutkimus tekoälyn säätelyn tueksi on kuitenkin vielä tällä hetkellä vähäistä. Tekoälyä koskeva akateeminen kirjallisuus on vahvasti painottonutta tekniseen lähestymiskulmaan ja tekoälyn eettiset ongelmat jäävät vähälle huomiolle. Tekoälyn nopea kehitys yhdistettynä vähäiseen eettiseen tutkimukseen aiheuttaa sen, että tekoälyn lainsäädäntö uhkaa jäädä uuden teknologian jalkoihin. [9]

Tekoälyä koskevia huolia on myös huomattavasti enemmän, kuin tässä työssä on voitu tuoda esille. Erityisesti generatiivisen tekoälyn, kuten ChatGPT:n, toiminta jää tekoälysäädöksen puitteissa hyvin vapaaksi. Siihen liittyy kuitenkin esimerkiksi tekijänoikeuksia koskevia eettisiä kysymyksiä.

5. KESKUSTELU JA POHDINTA

Aiheen tarkastelu osoittautui odotettua vaikeammaksi, sillä kyseessä ei ole puhtaasti objektiivinen teknologinen sovelluskohde, päinvastoin. Tekoälydirektiivi on vuorovaikutuksessa usean eri tieteenalan kanssa. Lainsäädäntö kumpuaa yhteiskunnallisista arvoista, huolista ja eettisistä kysymyksistä. Nämä yhdistyvät tekoälydirektiivissä monimutkaiseen teknologiaan, sekä lainsäädännön yleisiin haasteisiin.

Työn pohjimmaisena tarkoituksena on herätellä kriittistä ajattelutapaa käyttämiämme teknologioita kohtaan. Tekoälyllä on monia mahdollisuuksia parantaa työntekoa, sekä vapauttaa resursseja ihmiselle mielekkäisiin tehtäviin. Automatisoinnilla voidaan vähentää työntekijöille jääviä yksitoikkoisia rutiinitöitä ja jopa parantaa tehokkuutta niin, että tulevaisuudessa työviikot lyhenevät ja ihmisille jää enemmän vapaa-aikaa.

Tekoälyllä on kuitenkin myös kyseenalaisia käyttökohteita. Sitä käytetään jo avoimesti mielipidevaikuttamiseen eri sosiaalisen median alustoilla. Tekoälyllä luodut käyttäjätilit kommentoivat Ukrainan sotaa ja mainostavat tuotteita. Samalla suuret teknologiayritykset optimoivat algoritmeja omien etujensa mukaisesti. Erityisesti nämä maalaavat uhkakuva tekoälyn mahdollisesta väärinkäytöstä, jonka vuoksi työssä on tuotu esille teknologian läpinäkyvyyden tärkeys.

Työssä pyritään myös haastamaan ajattelutapaa, jonka mukaan tekniikka on objektiivista. Olen huomannut monilla alan opiskelijoilla sekä ammattilaisilla tavan irrottaa teknologinen kehitys sitä ympäröivästä poliittisesta ja eettisestä kontekstista. Tämä on kuitenkin mielestäni virheellinen ajatustapa, joka voi pahimmillaan aiheuttaa sen, ettei teknologian väärinkäyttöä edes tunnisteta. Opetusdatan tai painotetun algoritmin aiheuttamien vääristymien huomaaminen vaatii ymmärrystä ympäröivän maailman ilmiöistä. On erityisen helppo olla huomaamatta vääristymiä, jotka myötäilevät omia näkemyksiä, kokemuksia ja maailmankatsomusta. Esimerkiksi rekryointialgoritmin taipumus suosia vinoutuneen opetusdatan perusteella tiettyä sukupuolta voi jäädä rekrytoijalta huomaamatta, jos hän itse odottaa saavansa enemmän päteviä hakijoita kyseisen sukupuolen edustajista.

Tekoälyn haasteena onkin sen monimuotoisuus ja laajat käyttökohteet. Tekoälyn kehittäjät, algoritmeja lisensoivat tahot, loppukäyttäjät ja lainsäätäjät lähestyvät tekoälyä eri lähtökohdista ja vaikka kyseessä ovat alansa ammattilaiset, on harvalla valmiudet hahmottaa koko kokonaisuutta. Tämän vuoksi eri asiantuntijoiden välinen ymmärrys ja yhteistyö on

kriittistä. Teknologian ja etiikan välille vaikuttaa kuitenkin usein jäävän aukko, ellei jopa aktiivinen konflikti. On kenties hyväksyttävää todeta, ettei esimerkiksi karttasovelluksen kehittäjän tarvitse suorittaa syväluotaavaa eettistä analyysiä lyhimmän reitin algoritmille. Samaa ei kuitenkaan voi sanoa tuomioistuimen, rajavalvontalaitoksen tai pankin päätöksentekoon käyttämästä teknologiasta.

Euroopan unioni on joutunut vastaamaan tähän teknologian nopeaan kehitykseen. Haasteena on mahdollistaa Euroopan jäsenmaiden kilpailu kansainvälisillä markkinoilla ilman, että se uhkaa kansalaisten perusoikeuksia tai että markkinoille päätyy vaarallisia tuotteita. Koska tekoälysäädös ei työn kirjoitushetkellä ole astunut kaikilta osin voimaan, emme voi vielä tietää, miten se tulee vaikuttamaan tekoälyn kehitykseen. Työssä esille tuodut ongelmakohdat kuitenkin osoittavat, ettei tekoälysäädös ole täydellinen ratkaisu esitettyihin huoliin.

Sen sijaan se tuo useita ongelmia tekoälyn kehittäjien ja eri standardointitahojen ratkaistavaksi. On mahdollista, että asetetut vaatimukset vähentävät tekoälyn käyttöä yrityksissä. Vaikka tekoälysäädös tarjoaa pienille yrityksille ja starupeille mahdollisuuden yksinkertaistettuun dokumentaatioon, ovat ne kuitenkin resursseiltaan suuryrityksiä heikommassa asemassa säädöksen veloitteiden noudattamisen suhteen.

Se millaisena tekoälysäädös näyttäytyy Euroopan kansalaisten arjessa jää myös nähtäväksi. Optimistisesti synteettisen sisällön tunnistaminen käy helpommaksi, sillä säädös velvoittaa merkitsemään esimerkiksi tekoälyllä tuotetut kuvat. Toisaalta säädös ei puutu esimerkiksi tekijänoikeushuoliin, joita taiteilijat, muusikot ja kirjoittajat ovat tuoneet yhä enenemmän esille. Tekoäly vaikuttaakin aiheuttaneen varsinaisen teknologiakehityksen, lainsäädännön sekä tietotaidon juoksukilpailun. Todennäköisesti tekoäly, kuten uudet teknologiat kautta historian, asettuu aloilleen aiheuttamansa alkurysäyksen jälkeen. Onkin myös osaltaan mielenkiintoista pohtia, mikä keksintö aiheuttaa seuraavan väittelyn tekniikan alalla.

6. YHTEENVETO

Tässä työssä tarkasteltiin kirjallisuuskatsauksen avulla Euroopan unionin tekoälynsäädöstä ja sen vaikutuksia erityisesti automatisoidun päätöksenteon läpinäkyvyyteen. Työn tavoitteena oli selvittää vastaus tutkimuskysymykseen: Kuinka Euroopan unionin tekoälynsäädös vastaa tekoälyn päätöksenteon selitettävyyteen liittyviin haasteisiin, sekä minkälaiset puitteet se tarjoaa korkean riskin tekoälyjärjestelmien vastuulliselle käytölle?

Euroopan unionin tekoälynsäädös asettaa useita tavoitteita tekoälyn turvalliselle ja eettiselle käytölle. Säädös pyrkii edistämään tekoälynkäytön läpinäkyvyyttä ja varmistamaan sen, että tekoälyalgoritmit toimivat oikein korkean riskin tapauksissa. Työn keskeisenä tuloksena tekoälynsäädös ei kuitenkaan aseta selkeitä konkreettisia vaatimuksia, jotka varmistaisivat säädöksen tavoitteiden toteutumisen.

Sen sijaan säädös jättää paljon tulkinnanvaraisuutta eri vaatimuksille sekä nojaa vahvasti siihen oletukseen, että eri standardointitahot pystyvät kääntämään eettiset tavoitteet toimiviksi teknologisiksi ohjeiksi. Lisäksi tässä työssä tuotiin esille huolia siitä, että lain vaatimusten täyttymisen valvominen jätetään organisaatioiden sisäisille tahoille. Tämä yhdistettynä siihen, kuinka yritykset voivat vaikuttaa teknologiastandardien laatimisprosessiin, tuo lisää epävarmuutta tekoälynsäädöksen kykyyn saavuttaa esitetyt tavoitteet.

Aiheen valinta ja ajankohtaisuus perustui tekoälyn yleistyvään käyttöön yhä useammilla aloilla, jossa erityisesti yritykset käyttävät tekoälyä osana päätöksentekoa. Tekoälyalgoritmien toiminta on kuitenkin vaikeasti hahmotettavissa, sillä algoritmit ovat niin sanottuja mustia laatikoita, joiden toimintalogiikka on vaikeaa, tai lähes mahdotonta, jäljittää. Päätöksenteon läpinäkyvyyden kriittisyyttä perusteltiin työn alussa tuomalla esille Euroopan unionin tavoitteita läpinäkyvästä päätöksenteosta ja kansalaisten oikeuksista ymmärtää miten heidän tietojaan käsitellään.

Työn puitteissa ei löydetty vielä riittävästi keinoja tekoälymallien valaisemiseen. Tekoälymallit muodostuvat nopeasti hyvin monimutkaisiksi ja niissä käytetyt algoritmit hyödyntävät päätöksenteossa miljoonia eri datapisteitä. Lisäksi algoritmien paljastaminen on yritysten intressien vastaista. Tämän todettiin olevan merkittävä ongelmakohta, sillä suuri osa tekoälyteknologiasta sekä sitä koskevasta tietotaidosta on keskittynyt yritys sektorille, joka tekee myös tekoälynsäädöksen valvomisesta haasteen julkishallinnolle.

Ratkaisuja algoritmien toiminnan selkeyttämiseen tai tekoälynsäädöksen toiminnallisu-

teen liittyviin huoliin ei tällä kirjallisuuskatsauksella ilmennyt. Tekoälysäädöksen rinnalle tulevien mahdollisten direktiivien, lisäsäädösten ja teknisten standardien toimivuus selviää vasta tulevaisuudessa, kun tekoälysäädös astuu kokonaisuudessaan voimaan vuonna 2026.

Tällä hetkellä tämä työ laajentaa suomenkielistä tieteellistä kirjallisuutta automatisoidun päätöksenteon sekä erityisesti EU:n tekoälysäädöksen osalta, sekä tarjoaa pohjatiedot tekoälyn lainsäädännön kehityksen seuraamiselle. Työssä on esitelty tekoälysäädöksen keskeiset toimintaperiaatteet erityisesti korkean riskin tekoälyjärjestelmille, mutta tämäkään työ ei pysty antamaan täydellistä kokonaiskuvaa koko säädöksestä. Lisäksi työssä esitetyt kysymykset mustien laatikoiden avaamisesta jäävät vielä suurelta osin avoimiksi.

Tekoälyalgoritmien päätöksenteon ymmärtäminen on tämän työn puitteissa ilmennyt vielä uutena tutkimusalana, joka sisältää useita merkittäviä haasteita. Tekoälyalgoritmien jatkuva kehitys ja monimutkaisuus tekee algoritmien mustien laatikoiden avaamisesta yhä haastavampaa, mikäli algoritmien kehittäjät eivät huomioi ymmärrettävyyttä jo suunnitteluvaiheessa. Tällä hetkellä akateemisessa keskustelussa on tahoja, jotka pyrkivät ajamaan eteenpäin ajatusta selitettävistä tekoälymalleista (*eXplainable AI, XAI*). Tämä yhdistettynä toimivaan teknologiastandardointiin voi auttaa puskemaan tekoälysäädöksen tavoitteita eteenpäin, jolloin tämän työn perimmäiseen kysymykseen "*Miten tekoälyn päätöksentekoon voi luottaa?*" olisi mahdollista saada vihdoinkin tyydyttävä vastaus.

LÄHTEET

- [1] Euroopan parlamentin ja neuvoston asetus (EU) 2024/1689. saatavilla: https://eur-lex.europa.eu/legal-content/FI/TXT/HTML/?uri=OJ:L_202401689.
- [2] Marco Almada ja Nicolas Petit. "THE EU AI ACT: BETWEEN THE ROCK OF PRODUCT SAFETY AND THE HARD PLACE OF FUNDAMENTAL RIGHTS". *Common Market Law Review* 62.1 (2025), s. 85–120. DOI: <https://doi.org/10.54648/cola2025004>.
- [3] Eurostat. *Use of artificial intelligence in enterprises*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises. [viitattu 31.5.2025].
- [4] Katarina Foss-Solbrekk. "Searchlights across the black box: Trade secrecy versus access to information". *Computer Law Security Review* 50 (2023), s. 105811. DOI: <https://doi.org/10.1016/j.clsr.2023.105811>.
- [5] Marta Cantero Gamito ja Christopher T. Marsden. "Artificial intelligence co-regulation? The role of standards in the EU AI Act". *International journal of law and information technology* (32 2024). DOI: <https://doi.org/10.1093/ijlit/eaee011>.
- [6] Tom Gerken. *Bacon ice cream and nugget overload sees misfiring McDonald's AI withdrawn*. <https://www.bbc.com/news/articles/c722gne7qngo>. [viitattu 9.8.2025]. 2024.
- [7] Jarek Gryz ja Marcin Rojszczak. "Black box algorithms and the rights of individuals: no easy solution to the "explainability" problem". *Internet policy review* 10 (2021). DOI: <https://doi.org/10.14763/2021.2.1564>.
- [8] Jack Guy. *Denmark plans to thwart deepfakers by giving everyone copyright over their own features*. <https://edition.cnn.com/2025/06/27/business/denmark-ai-law-scli-intl>. [viitattu 3.8.2025]. 2025.
- [9] Byeong-Je Kim et al. "AI Governance in the Context of the EU AI Act". *IEEE Access* (2025). DOI: <https://doi.org/10.1109/ACCESS.2025.3598023>.
- [10] Euroopan komissio. *Tekoälysäädös tulee voimaan*. https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_fi. [viitattu 1.9.2025]. 2024.
- [11] Cynthia Kroet. *EU Commission confirms ditching of AI liability and patents proposals*. <https://www.euronews.com/next/2025/07/31/eu-commission-confirms-ditching-of-ai-liability-and-patents-proposals>. [viitattu 3.8.2025]. 2025.
- [12] Sabrina Kutscher. "The EU AI Act: Law of Unintended Consequences?" *Technology and Regulation* (2025). DOI: <https://doi.org/10.71265/krne7205>.

- [13] Ljubiša Metikoš ja Jef Ausloos. "The right to an explanation in practice: insights from case law for the GDPR and the AI Act". *Law, Innovation and Technology* 17 (2025), s. 205–240. DOI: <https://doi.org/10.1080/17579961.2025.2469349>.
- [14] Euroopan parlamentti. *EU AI Act: first regulation on artificial intelligence*. <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. [viitattu 31.5.2025].
- [15] Georgios Pavlidis. "Unlocking the black box: analysing the EU artificial intelligence act's framework for explainability in AI". *Law, Innovation and Technology* 16 (2024), s. 293–308. DOI: <https://doi.org/10.1080/17579961.2024.2313795>.
- [16] Nicoletta Rangone ja Luca Megale. "Risks Without Rights? The EU AI Act's Approach to AI in Law and Rule-Making". *European Journal of Risk Regulation* (2025). DOI: <https://doi.org/10.1017/err.2025.13>.
- [17] Cynthia Rudin. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". *Nature machine intelligence* 1 (2019), s. 206–215. DOI: <https://doi.org/10.1038/s42256-019-0048-x>.
- [18] Ludivine Sarah Stewart. "The regulation of AI-based migration technologies under the EU AI Act: (Still) operating in the shadows?" *European law journal* 30 (2024), s. 122–135. DOI: <https://doi.org/10.1111/eulj.12516>.
- [19] Sandra Wachter. "Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond". *Yale Journal of Law and Technology* 26 (2024). DOI: <https://doi.org/10.2139/ssrn.4924553>.
- [20] Euroopan unionin verkkosivut. https://european-union.europa.eu/institutions-law-budget/law/types-legislation_en. [viitattu 13.8.2025].