

**DigiFinland**Sosiaali- ja
terveysministeriöTerveystieteiden
laitos

Tekoälyä hyödyntävien palveluiden riskienhallinta sosiaali- ja terveydenhuollossa

SOTE-TEKOÄLYN KOKEILUPROJEKTIT 2025ANDREI LAURÉN (STM), MIKA PIHLAJAMÄKI (THL), TERHI PIETILÄ
(DIGIFINLAND), TERO MÄKIRANTA (DIGIFINLAND)

Sisällysluettelo

1. Johdanto.....	3
2. Riskienhallintaryhmä.....	3
3. Riskien tunnistus	3
4. Riskienhallintatyökalu	4
5. EU-säädökset	4
5.1 Datanhallinta-asetus (Data Governance Act)	5
5.2 Datasäädös (Data Act)	5
5.3 Tekoälyasetus (AI Act)	5
5.4 Tietosuojalainsäädäntö (GDPR)	6
5.5 Verkko- ja tietoturvadirektiivi (NIS 2)	6
5.6 Lääkinnällisiä laitteita koskeva asetus (MDR)	6
5.7 In vitro -diagnostiikkaan tarkoitettuja lääkinällisiä laitteita koskeva asetus (IVDR)	6
6. Riskiluokat.....	6
6.1 Vastuullisuus.....	6
6.2 Läpinäkyvyys	7
6.3 Luotettavuus	7
6.4 Oikeudenmukaisuus	7
6.5 Osallistaminen	7
7. Eettinen käyttö.....	8
8. Turvallinen käyttö	9
8.1 Vastuullisuus.....	9
8.1.1 Havainnot kokeiluista: vastuullisuus riskienhallinnassa	10
8.2 Läpinäkyvyys	10
8.2.1 Havainnot kokeiluista: läpinäkyvyys riskienhallinnassa	11
8.3 Luotettavuus	12
8.3.1 Havainnot kokeiluista: luotettavuus riskienhallinnassa	12
8.4 Oikeudenmukaisuus	13
8.4.1 Havainnot kokeiluista: oikeudenmukaisuus riskienhallinnassa	14
8.5 Osallistaminen	15
8.5.1 Havainnot kokeiluista: osallistaminen riskienhallinnassa	15

9. Tekoälyn käytön esteet	16
10. Käytännöt havainnot hyvinvointialueiden tekoälypilottien riskienarvioinneista	18
10.1 Vakavimmat kokonaisriskit	18
10.2 Vakavimmat kokonaisriskit hallintakeinojen jälkeen	18
10.3 Johtopäätös merkittävistä jäännösriskeistä	19
11. Huomiot tekoälyn vaikutusten arvioinnissa riskien pohjalta	20
12. Tekoälyn hallintoratkaisut	21
12.1 Tarkemmin AI governance -työkaluista	22
13. Toimenpidesuosituksia	23
13.1 Kansallisen ohjauksen ja hallintamallin kehittäminen	23
13.2 Tekoälyn käyttöönoton yhteiskunnallisten vaikutusten arviointi sote-sektorilla: kehityskohteet	23
13.3. Tekoälyosaamisen vahvistaminen ja tekoälyn käytön edistäminen.....	24
14. Kvanttitekkoäly ja robotiikka – tulevaisuuden riskienhallinnan näkökulmia.....	25
15. Yhteenveto.....	25
Liitteet	27
Liite 1. Huomiota vaikutusten arviointiin.....	27
Liite 2. Tekoälyn turvallisen käytön arviointi.....	28

Raportissa esitetyt kaaviot on tuotettu generatiivisen tekoälyn (Google Gemini) avulla.

1. Johdanto

Osana sosiaali- ja terveysministeriön käynnistämää SOTE-tekoälyn ekosysteemin toimintaa perustettiin lokakuussa 2024 työryhmä, joka keskittyy sosiaali- ja terveydenhuoltoalan tekoälyratkaisujen riskienhallintaan. Riskienhallintaryhmän tehtävänä oli:

- Tukea tekoälyn turvallisen käytön kansallista kehittämistä osana vaikutustenarviointeja
- Vahvistaa erityisesti hyvinvointialueiden riskinhallintaa
- Suositella konkreettisia kehittämistoimenpiteitä riskienhallintaan ja kerätä tietoa riskien hallintakeinoista

Yhtenä osana Riskienhallintaryhmän työskentelyä oli riskienhallintatyökalun kehittäminen sosiaali- ja terveydenhuoltoalan tekoälyä hyödyntävien ratkaisujen riskiarviointeja varten. Tämä dokumentti pohjautuu ekosysteemin riskienhallintaryhmän toimintaan, ekosysteemin verkostoihin ja kokeilupprojektien havaintoihin.

2. Riskienhallintaryhmä

Riskienhallintatyöryhmä koostui usean eri sosiaali- ja terveydenhuollon toimijan edustajasta, jotta mahdollisia riskejä ja niiden hallintakeinoja havaittaisiin kattavasti. Mukana oli edustajia Sosiaali- ja terveysministeriöstä (STM), Terveys- ja hyvinvoinnin laitokselta (THL), DigiFinland, julkisen sektorin erityistehtäväyhtiöstä, julkisoikeudellisesta yhteisöstä sekä yksityisen sektorin toimijoista, jotka kehittävät ja tuottavat ratkaisuja sosiaali- ja terveydenhuoltoon.

Riskienhallintaryhmä lähestyi tekoälyä hyödyntävien sosiaali- ja terveydenhuollon ratkaisujen riskien tunnistamista ja niiden hallintaa lähinnä eettisyyden ja tekoälyn turvallisen käytön näkökulmasta.

3. Riskien tunnistus

Riskienhallintaryhmä tunnisti tekoälyn käytön aiheuttamia riskejä sosiaali- ja terveydenhuoltoon liittyvissä palveluissa mm. työryhmäpalavereissa, ryhmässä vierailevien asiantuntijoiden esityksistä sekä hyödyntämällä vuoden 2025 aikana toteutettavien tekoälyn kokeilupprojektien laatimia riskiarviointeja. Tyypillisiä riskejä sote-tekoälysovelluksissa ovat mm. virheellinen päätös, datan laatu, osaamisen puute ja toimintakulttuuri, syrjivyyys, ymmärryksen puute, ylikuormitus ja tietosuojariskit. Riskin tunnistuksella pyritään ennakoivasti havaitsemaan mahdollisia vaaroja, uhkia tai epävarmuustekijöitä, jotka voivat vaikuttaa haitallisesti sote-palvelun toteutumiseen eettisesti ja turvallisesti.

Riskienhallintaryhmä tarkasteli toimintaympäristöjä, joissa tekoälyratkaisuja mahdollisesti tullaan käyttämään ja arvioinut mitä mahdollisia uhkia tai epävarmuuksia niihin liittyy. Näistä riskeistä on koottu taulukko ja niiden pohjalta tehtiin työkalu riskien arvioinnin tueksi. Tätä työkalua on mahdollista kehittää eteenpäin lainsäädännön ja tekoälyn ratkaisujen kokemusten ja käytötapausten myötä.

4. Riskienhallintatyökalu

Riskienhallintatyökalu on SOTE-tekoälyn ekosysteemin riskienhallintaryhmän kehittämä työkalu, joka pohjautuu ryhmän työskentelyyn ja aihepiiriä käsitteleviin esityksiin. Sen tarkoituksena on tunnistaa, arvioida ja hallita tekoälyn käyttöön liittyviä riskejä jo varhaisessa vaiheessa, jotta järjestelmää voidaan kehittää ja käyttää hallitusti ja ennakoivasti.

Työkalulla voidaan jäsentää ja hallita sosiaali- ja terveydenhuollon digitaalisissa tekoälyratkaisuissa esiintyviä eettisiä riskejä ja uhkia. Se kokoaa yhteen yleisesti tunnistettuja eettisiä ja turvalliseen käyttöön liittyviä riskitekijöitä, mutta rajaa pois yksityiskohtaiset tietotekniset riskit. Dokumentissa esitetyt yleiset riskihavainnot perustuvat SOTE-tekoälyn ekosysteemin kokeiluprojekteista saatuihin kokemuksiin.

Riskityökalua voidaan hyödyntää myös EU-säädösten toimeenpanon tukena. Seuraavassa luvussa käsitellään näitä EU-vaatimuksia tarkemmin.

5. EU-säädökset

Euroopan unionin sääntelykehys tekoälylle muodostuu useista toisiaan täydentävistä asetuksista ja direktiiveistä, jotka ohjaavat datan hallintaa, tietosuojaa, turvallisuutta sekä erityisalojen vaatimuksia. Tavoitteena on varmistaa, että tekoälyratkaisut ovat läpinäkyviä, turvallisia ja luotettavia sekä noudattavat perusoikeuksia ja eettisiä periaatteita. Keskeisiä säädöksiä ovat:

- Datanhallinta-asetus (Data Governance Act)
- Datasäädös (Data Act)
- Tekoälyasetus (AI Act)
- Tietosuojalainsäädäntö (GDPR)
- Verkko- ja tietoturvadirektiivi (NIS 2)
- Lääkinnällisiä laitteita koskeva asetus (MDR)
- In vitro -diagnostiikkaan tarkoitettuja lääkinällisiä laitteita koskeva asetus (IVDR)

Nämä säädökset muodostavat kokonaisuuden, joka ohjaa datan, teknologian ja turvallisuuden hallintaa sekä varmistaa potilasturvallisuuden ja luotettavuuden sosiaali- ja terveydenhuoltoalalla.

EU-säädökset: Luottamusta ja turvallisuutta tekoälyyn ja dataan

Euroopan unionin sääntelykehys muodostuu useista toisiaan täydentävistä säädöksistä. Niiden tavoitteena on varmistaa, että tekoäly ja data ovat läpinäkyviä, turvallisia ja noudattavat perusoikeuksia erityisesti kriittisellä sosiaali- ja terveydenhuoltoalalla.

Tekoäly ja datan hallinta



5.1 Datanhallinta-asetus (Data Governance Act)

Asetus edistää datan jakamista EU:ssa luomalla luottamusta ja selkeitä sääntöjä datan hallinnalle. Se koskee erityisesti julkisen sektorin dataa ja datan välityspalveluita ja tukee datatalouden kehitystä. Eettisesti tämä vahvistaa luottamusta ja läpinäkyvyyttä, mikä on tärkeää sosiaali- ja terveydenhuollossa, jossa käsitellään arkaluonteisia potilastietoja. Riskienhallinnan näkökulmasta selkeät säännöt vähentävät väärinkäytön riskiä ja varmistavat, että dataa jaetaan vain luotettavien toimijoiden kesken.

5.2 Datasäädös (Data Act)

Datasäädös täydentää hallinta-asetusta ja keskittyy datan käyttöoikeuksiin ja jakamiseen yritysten ja kuluttajien välillä. Se pyrkii varmistamaan reilun pääsyn dataan ja edistämään innovaatioita datan hyödyntämisessä. Eettisesti tämä estää datan monopolisoitumista ja tukee oikeudenmukaista kilpailua. Riskienhallinnan kannalta selkeät käyttöoikeudet vähentävät väärinkäytön riskiä ja tukevat potilaan oikeuksia datan hallinnassa.

5.3 Tekoälyasetus (AI Act)

EU:n tekoälyasetus luo riskiperusteisen sääntelykehyksen tekoälyjärjestelmille ja määrittelee vaatimukset korkeaa riskiä sisältäville sovelluksille kuten läpinäkyvyys, turvallisuus ja ihmisen valvonta sekä kieltää tietyt riskialttiit käytöt. Eettisesti tämä varmistaa, että tekoälyä käytetään vastuullisesti ja ihmisen päätöksenteko säilyy keskiössä. Riskienhallinnan näkökulmasta asetuksen vaatimukset vähentävät virheiden ja haitallisten vaikutusten riskiä.

5.4 Tietosuojalainsäädäntö (GDPR)

Tietosuojaj-asetus säätelee henkilötietojen käsittelyä ja varmistaa yksilön oikeudet kuten oikeuden tietojen poistamiseen ja läpinäkyvyyteen sekä asettaa velvoitteita organisaatioille tietoturvan ja tietosuojan varmistamiseksi. Eettisesti tämä tukee yksityisyyden suojaa ja potilaan autonomiaa. Riskienhallinnan kannalta GDPR:n velvoitteet vähentävät tietovuotojen ja väärinkäytön riskiä.

5.5 Verkko- ja tietoturvadirektiivi (NIS 2)

NIS 2 -direktiivi vahvistaa EU:n kyberturvallisuutta ja asettaa velvoitteita kriittisten palveluiden ja infrastruktuurien toimijoille riskienhallinnasta, tietoturvasta ja ilmoitusvelvollisuuksista tietoturvaloukkauksissa. Eettisesti tämä suojaa potilasturvallisuutta ja hoidon jatkuvuutta. Riskienhallinnan näkökulmasta direktiivit parantavat reagointikykyä ja vähentävät kyberhyökkäysten vaikutuksia.

5.6 Lääkinnällisiä laitteita koskeva asetys (MDR)

Medical Device Regulation säätelee lääkinällisten laitteiden turvallisuutta ja suorituskykyä EU:ssa ja sisältää tiukat vaatimukset laitteiden arvioinnista, jäljitettävyydestä ja markkinoille pääsystä. Eettisesti tämä varmistaa, että tekoälyä sisältävät laitteet ovat luotettavia ja turvallisia. Riskienhallinnan kannalta asetuksen vaatimukset vähentävät virheiden ja väärinkäytön riskiä.

5.7 In vitro -diagnostiikkaan tarkoitettuja lääkinällisiä laitteita koskeva asetys (IVDR)

In Vitro Diagnostic Regulation koskee diagnostisia laitteita, joita käytetään näytteiden analysointiin ja varmistaa laitteiden luotettavuuden, turvallisuuden ja kliinisen suorituskyvyn. Eettisesti tämä tukee oikeiden hoitopäätösten tekemistä ja potilasturvallisuutta. Riskienhallinnan näkökulmasta asetuksen vaatimukset vähentävät virheellisen diagnostiikan riskiä.

6. Riskiluokat

Riskienhallintaryhmä tunnisti seuraavat viisi riskiluokkaa: vastuullisuus, läpinäkyvyys, luotettavuus, oikeudenmukaisuus ja osallistaminen. Riskienhallintaryhmässä esiin nostetut riskiluokat kuuluvat kategorisesti eettisiin ja sosiaalisiin riskeihin seuraavasti.

6.1 Vastuullisuus

Riskit kohdistuvat siihen, että tekoälyn käyttöön liittyvät vastuut ja roolit ovat epäselviä sekä siihen, että tekoälyä voidaan käyttää epäeettisesti tai väärin. Ilman selkeitä eettisiä periaatteita ja valvontaa tekoäly voi toimia väärinkäytösten mahdollistajana, mikä vaarantaa

asiakkaiden oikeudet ja heikentää luottamusta arvioitavaan järjestelmään. Tekoälyä saatetaan esimerkiksi hyödyntää päätöksenteossa, jossa painottuvat taloudelliset hyödyt ammattilaisten ja asiakkaiden edun sijaan.

6.2 Läpinäkyvyys

Tekoälyn rooli päätöksen tukemisessa voi olla vaikeasti ymmärrettävä tai selittämätön, mikä heikentää asiakkaiden ja ammattilaisten mahdollisuuksia arvioida päätösehdotusten perusteluita. Tämä voi lisätä riskiä siitä, että tekoälyä käytetään kritiikittä ja ei ymmärretä, mistä tekoälyn tuottama tieto koostuu, ja mikä taustalogiikka ja perustelut ovat olleet. Tämä tunnetaan yleisesti ”Mustan laatikon” -haasteena. Läpinäkyvyys on myös sitä, että asiakkaat ja ammattilaiset tunnistavat käyttävänsä tekoälyä toiminnassaan.

6.3 Luotettavuus

Tekoälyratkaisut voivat antaa virheellisiä, epäyhtenäisiä tai vinoutuneita suosituksia tai tuloksia, erityisesti, jos ne on koulutettu puutteellisella datalla tai niitä ei ole testattu riittävästi käytännön ympäristöissä. Opetusdatan haasteina voivat olla esimerkiksi tiedon ajantasaisuuden puute tai lähdetietoa sovelletaan virheellisesti laajempiin käyttötapauksiin ja datan vähyys voi vinouttaa tuloksia.

Luotettavuus voi aiheuttaa kahdenlaista haastetta ammattilaisen työssä. Ammattilainen voi luottaa tekoälyn tuottamaan tietoon liiallisesti tai ammattilainen voi epäillä tekoälyn tuottamaa sisältöä. Luotettavuuden puute voi johtaa pahimmillaan asiakkaan väärään palveluohjaukseen tai hoitoon/palveluun, ja pahimmillaan horjuttaa palvelujärjestelmän uskottavuutta.

6.4 Oikeudenmukaisuus

Tekoäly saattaa toimia epäoikeudenmukaisesti, jos algoritmit kohtelevat eri väestöryhmiä eriarvoisesti esimerkiksi iän, sukupuolen, etnisen taustan tai sosioekonomisen aseman perusteella. Lisäksi riski piilee siinä, että resurssien kohdentamista ohjataan tehokkuus- tai kustannusperusteisesti ilman riittävää huomiota heikommassa asemassa olevien tarpeisiin. Oikeudenmukaisuus on myös sitä, että asiakkailla ja ammattilaisilla on yhtäläinen mahdollisuus hyödyntää tekoälyä.

6.5 Osallistaminen

Ammattilaisten puutteellinen osallistaminen tekoälyn kehittämiseen, testaamiseen sekä käyttöönottoon voi johtaa riskinä järjestelmiin, jotka eivät vastaa todellisia tarpeita tai vaatimuksia ja käyttöaste jää alhaiseksi. Puutteellinen osallistaminen ja viestintä voi myös heikentää teknologian hyväksyttävyyttä ja mahdollistaa ratkaisujen kehittämisen suljetuissa verkostoissa ilman laajempaa eettistä ja sosiaalista pohdintaa.

Tekoälyn 5 Eettistä Riskiä

Ymmärrä tekoälyn keskeiset eettiset ja sosiaaliset riskit yhdellä silmäyksellä.



7. Eettinen käyttö

Tekoälyn eettinen käyttö sosiaali- ja terveydenhuollossa perustuu periaatteisiin, jotka turvaavat potilaan oikeudet, yksityisyyden ja turvallisuuden sekä varmistavat luottamuksen teknologian hyödyntämiseen. Riskienhallinta on olennainen osa tätä kokonaisuutta, sillä tekoälyratkaisut käsittelevät arkaluonteista dataa ja vaikuttavat suoraan ihmisten terveyteen ja hyvinvointiin.

Eettinen käyttö tarkoittaa, että tekoälyä sovelletaan läpinäkyvästi ja vastuullisesti. Potilaan tietosuoja on keskiössä, ja henkilötietojen käsittelyssä noudatetaan tiukasti GDPR:n vaatimuksia. Datanhallinta-asetus ja datasäädös tukevat luottamusta ja reilua datan jakamista, mikä on välttämätöntä innovaatioiden edistämiseksi ilman, että yksilön oikeudet vaarantuvat. Tekoälyasetuksen riskiperusteinen lähestymistapa varmistaa, että korkean riskin sovellukset, kuten diagnostiikkaa tukevat järjestelmät, täyttävät vaatimukset läpinäkyvyydestä, turvallisuudesta ja ihmisen valvonnasta.

Eettinen käyttö turvataan kyberturvallisuudirektiivillä NIS 2, mikä on tärkeää hoidon jatkuvuuden ja oikean tiedon turvaamiseksi. Lääkinnällisiä laitteita koskevat asetukset MDR ja IVDR varmistavat, että tekoälyä sisältävät laitteet ja diagnostiset järjestelmät ovat luotettavia ja turvallisia, mikä vähentää virheellisen hoidon riskiä ja tukee potilasturvallisuutta.

Eettinen käyttö edellyttää myös oikeudenmukaisuutta ja syrjimättömyyttä. Tekoälyalgoritmien tulee olla läpinäkyviä ja niiden päätöksenteon perusteet selkeitä, jotta vältetään vinoumat ja varmistetaan tasapuolinen kohtelu kaikille potilaille. Ihmisen valvonta säilyy keskeisenä periaatteena, sillä tekoäly ei saa korvata kliinistä harkintaa vaan toimia sen tukena.

Seuraavissa kohdissa on tarkasteltu tekoälyn turvallista käyttöä vastuullisuuden, läpinäkyvyyden, luotettavuuden, oikeudenmukaisuuden ja osallistamisen näkökulmista eettisen käytön pohjaksi.

8. Turvallinen käyttö

Turvallinen tekoälyn käyttö tarkoittaa, että tekoälyä sovelletaan vastuullisesti, harkitusti ja ihmisten sekä yhteiskunnan etua suojellen. Se edellyttää, että yksityisyys ja tietoturva ovat kunnossa, ja että tekoälyn tuottamat tulokset ja päätökset ovat mahdollisimman läpinäkyviä, perusteltuja ja jäljitettävissä. Turvallisuus merkitsee myös oikeudenmukaisuutta, luotettavuutta ja osallistavuutta: sitä, että järjestelmä kohtelee käyttäjiä tasapuolisesti ja on rakennettu siten, että sen toimintaan voivat vaikuttaa ne, joita se koskettaa.

Keskeistä on lisäksi se, että tekoälyn toiminta on selitettävissä, jotta voidaan ymmärtää, millaisista tiedoista ja perusteista ehdotukset tai päätökset muodostuvat. Turvallinen käyttö sisältää myös virheiden ennakoinnin, hallinnan ja korjaamisen sekä väärinkäytön riskien ehkäisyä. Kokonaisuutena kyse on selkeiden rajojen ja ohjeiden asettamisesta: tekoäly toimii ihmisen tukena ja vahvistaa päätöksentekoa, mutta ei koskaan korvaa sitä tavalla, joka voisi johtaa hallitsemattomiin tai haitallisiin seurauksiin.

Liitteessä 2 on kuvattu esimerkki, jonka avulla tekoälyn turvallista käyttöä voi myös arvioida. Turvalliseen käyttöön liittyy vastuullisuus, läpinäkyvyys, oikeudenmukaisuus, luotettavuus sekä osallistamiseen liittyvät riskit.

Liitteessä esitetään kunkin osa-alueen osalta todennäköisyys- ja vaikutustaulukot, joiden avulla voidaan arvioida tekoälyyn liittyviä riskejä. Todennäköisyys ja vaikutus arvioidaan asteikolla 1–5.

Näiden arvojen perusteella voidaan laskea riskiluku, joka saadaan kertomalla todennäköisyys ja vaikutus keskenään ($\text{todennäköisyys} \times \text{vaikutus}$). Riskilukua hyödynnetään tekoälyn riskitason arvioinnissa ja päätöksenteon tukena.

Seuraavissa kohdissa käydään tarkemmin läpi tekoälyn turvallisen käytön periaatteet: vastuullisuus, läpinäkyvyys, oikeudenmukaisuus, luotettavuus ja osallistamiseen liittyvät riskit.

8.1 Vastuullisuus

Tekoälyn vastuullinen käyttö tarkoittaa sen hyödyntämistä tavalla, joka tukee yhteiskunnan hyvinvointia ja ihmisten oikeuksia. Käytännössä tämä merkitsee, että tekoälyn ratkaisut suunnitellaan ihmiskeskeisesti, niiden toimintaa valvotaan ja vaikutuksia arvioidaan jatkuvasti. Vastuullisuus ei ole vain tekninen kysymys, vaan siihen liittyy myös arvopohjaisuuden ja sosiaalisia näkökulmia: tekoälyn tulee edistää luottamusta ja tasa-arvoa sekä välttää syrjiviä tai haitallisia seurauksia.

8.1.1 Havainnot kokeiluista: vastuullisuus riskienhallinnassa

Ennakointi ja korjaavat toimet ovat keskeinen osa vastuullisuutta.

Vastuullinen toiminta edellyttää ennakointia ja varautumista, jotta myös harvinaiset mutta vakavia seurauksia aiheuttavat riskit voidaan estää jo etukäteen. Silloin kun riskin todennäköisyys on suuri, korostuu järjestelmällisten korjaavien toimien ja jatkuvan valvonnan merkitys, koska vaarat voivat toteutua helposti käytännössä.

Vastuullisuus ei ole irrallinen arvo, vaan se toimii läpileikkaavana periaatteena riskienhallinnassa ja tekoälyn kehittämisessä. Se sitoo yhteen todennäköisyydet ja vaikutukset niin, että tekninen riskimatriisi muuttuu eettiseksi velvoitteeksi: miten estää epäoikeudenmukaisuudet ja varmistaa, että tekoälyn turvallinen käyttö tukee käyttäjien yhdenvertaisuutta ja oikeusturvaa. Toisin sanoen vastuullisuus tuo riskienhallintaan inhimillisen ulottuvuuden, joka pakottaa sote-sektoria tarkastelemaan tekoälyä suorituskkyä ja tehokkuutta yhteiskunnallisen oikeudenmukaisuuden näkökulmasta.

Vastuullisuudessa nousee esiin kolme keskeistä huomiota:

1. Ennakointi ja varautuminen:

Vastuullinen toiminta alkaa riskien ennakkoinnista. Jos harvinaiset mutta vakavat riskit jäävät huomioimatta, seuraukset voivat olla merkittäviä. Ennakoiva suunnittelu vähentää todennäköisyyttä ja tuo järjestelmään resilienssiä.

2. Korjaavat toimet ja jatkuva valvonta:

Kun riskin todennäköisyys on korkea, korostuu järjestelmällisten korjaavien toimien merkitys. Vastuullisuus ei ole kertaluonteinen ratkaisu, vaan jatkuvaa seurantaa ja reagointia, jotta vaarat eivät toteudu käytännössä.

3. Eettinen ulottuvuus:

Vastuullisuus ei rajoitu teknisiin ratkaisuihin. Se muuttaa riskienhallinnan eettiseksi velvoitteeksi: miten estetään epäoikeudenmukaisuudet ja varmistetaan yhdenvertaisuus sekä oikeusturva tekoälyn käytössä. Tämä tuo inhimillisen näkökulman tekniseen kehitykseen.

Näin vastuullisuus muodostaa riskienhallinnan ytimen: se toimii läpileikkaavana periaatteena, joka yhdistää tekniset, toiminnalliset ja eettiset näkökulmat yhdeksi kokonaisuudeksi.

8.2 Läpinäkyvyys

Sote-alan tekoälyn käytön riskienhallinnassa **läpinäkyvyys** tarkoittaa sitä, että tekoälyjärjestelmän toiminta, päätöksenteko ja vaikutukset ovat ymmärrettäviä sekä

järjestelmän käyttäjille että niille, joita ratkaisut koskevat. Läpinäkyvyys on vastakohta mustalle laatikolle.

Käytännössä tämä sisältää esimerkiksi sen, että on selkeästi dokumentoitava, mihin tietoihin tekoäly perustuu, millaisia malleja ja menetelmiä on käytetty sekä millaisia rajoitteita ja epävarmuuksia niihin liittyy. Lisäksi käyttäjien tulee saada tietää, milloin he ovat vuorovaikutuksessa tekoälyn tuottaman ratkaisun kanssa ja millä perusteilla järjestelmä on tehnyt ehdotuksensa. Läpinäkyvyys vahvistaa luottamusta, mahdollistaa ulkopuolisen arvioinnin ja tukee vastuullista käyttöä. Muutoin riskienhallinta jää helposti puutteelliseksi, koska piiloon jääneet virheet tai vääristymät voivat johtaa vakaviin seurauksiin potilaiden hoidossa tai sosiaalipalveluiden päätöksenteossa.

8.2.1 Havainnot kokeiluista: läpinäkyvyys riskienhallinnassa

Läpinäkyvyys täydentää riskienhallinnan näkökulmaa siten, että mitä heikompi läpinäkyvyys on, sitä suurempi on todennäköisyys väärinkäsityksille, virheelliselle tekoälyn käytölle ja hallitsemattomille seurauksille. Tällaisia seurauksia voivat olla esimerkiksi virheelliset analyysit, väärät johtopäätökset palveluiden kohdentamisesta, epäluottamus päätöksentekoon tai jopa haitalliset vaikutukset potilasturvallisuuteen ja sosiaalipalveluiden oikeudenmukaisuuteen. Jos tekoälyjärjestelmän toimintaperusteita ei avata, käyttäjät voivat luottaa siihen liaksi tai väärällä tavalla, mikä kasvattaa riskiä virheellisistä päätöksistä. Läpinäkyvyys vähentää riskiä lähtötiedon, väärin painotusten tai muiden vastaavien vinoumien johtamisesta virheellisiin tuloksiin.

Vaikutukset voivat ulottua yksittäisistä virheistä yhteiskunnallisesti merkittäviin päätöksiin, jos tekoälyä sovelletaan laajassa mittakaavassa. Läpinäkyvyyden puute siis sekä lisää riskien todennäköisyyttä että kasvattaa niiden vaikutusten vakavuutta, koska ongelmia on vaikeampi havaita ajoissa ja vastuukysymykset jäävät epäselviksi.

Läpinäkyvyys voidaan näin nähdä sekä itsenäisenä riskialueena että riskienhallintaa vahvistavana tekijänä. Kun se toteutuu riittävällä tasolla, se alentaa kokonaisriskiprofiilia merkittävästi ja luo edellytyksiä luotettavalle ja vastuulliselle tekoälyn käytölle sote-sektorilla. Läpinäkyvyys edistää tekoälyn käyttöä sote-alalla, koska se vahvistaa luottamusta, mahdollistaa ammatillisen harkinnan ja selkiyttää vastuita.

Läpinäkyvydessä nousee esiin kolme keskeistä huomiota:

1. Selkeät toimintaperusteet ja ymmärrettävyys:

Jos tekoälyjärjestelmän logiikkaa ei avata riittävästi, käyttäjien on vaikea arvioida sen luotettavuutta. Tämä voi johtaa virheellisiin päätöksiin ja heikentää luottamusta järjestelmään. Vaikutus on korkea, koska epäselvyys lisää väärinkäytön riskiä.

2. Väärinkäsitysten ja vinoumien hallinta:

Läpinäkyvyyden puute kasvattaa todennäköisyyttä, että lähtötiedon virheet tai painotukset jäävät huomaamatta. Todennäköisyys voi olla keskitasoa, mutta vaikutus merkittävä, sillä virheelliset analyysit voivat johtaa haitallisiin seurauksiin sote-palveluiden kohdentamisessa.

3. Vastuiden ja luottamuksen selkiyttäminen:

Kun järjestelmän toimintaa ei avata, vastuukysymykset jäävät epäselviksi ja luottamus päätöksentekoon heikkenee. Hyvä läpinäkyvyys parantaa ammatillisen harkinnan mahdollisuutta ja vähentää riskiä hallitsemattomista seurauksista.

Läpinäkyvyys toimii sekä itsenäisenä riskialueena että riskienhallintaa vahvistavana tekijänä. Riittävä läpinäkyvyys alentaa kokonaisriskiprofiilia merkittävästi ja luo edellytykset luotettavalle, vastuulliselle tekoälyn käytölle sote-alalla. Se lisää luottamusta, tukee oikeudenmukaista päätöksentekoa ja parantaa potilasturvallisuutta.

8.3 Luotettavuus

Luotettavuus sote-alan tekoälyn riskienhallinnassa tarkoittaa sitä, että järjestelmä toimii johdonmukaisesti, tarkasti ja ennakoitavalla tavalla erilaisissa käyttötilanteissa. Luotettava tekoäly perustuu laadukkaaseen ja ajantasaiseen lähtötietoon, sen toimintaa on testattu kattavasti, ja sen tuottamat tulokset ovat toistettavia. Lisäksi luotettavuus tarkoittaa sitä, että järjestelmä ei romahda yllättävissä tilanteissa, vaan tunnistaa omat rajoituksensa ilmoittaen epävarmuudesta, kun varmoja vastauksia ei voida antaa. Riskienhallinnan näkökulmasta luotettavuus vähentää virhepäätösten todennäköisyyttä, tukee sote-ammattilaisten luottamusta järjestelmään ja varmistaa, että tekoälyn käyttö vahvistaa eikä vaaranna potilasturvallisuutta tai sosiaalipalveluiden oikeudenmukaisuutta.

8.3.1 Havainnot kokeiluista: luotettavuus riskienhallinnassa

Luotettavuuden kannalta riskit eivät synny vain yksittäisistä virheistä, vaan epäluotettava tekoälyjärjestelmä voi tuottaa jatkuvaa ja vaikeasti ennakoitavaa epävarmuutta. Virheiden riski kasvaa erityisesti silloin, kun järjestelmää ei ole testattu riittävän monipuolisesti todellisissa käyttötapauksissa tai kun sen toiminta on osittain läpinäkymätöntä. Tällaiset puutteet voivat heikentää ammattilaisten ja potilaiden/asiakkaiden luottamusta sekä voi johtaa virheellisiin ratkaisuihin terveydenhuollossa, sosiaalipalveluissa tai hallinnossa, missä päätösten vaikutukset ovat laajoja.

Luotettavuuteen liittyvät riskit näkyvät usein vaikutusten kautta. Vaikka virheen todennäköisyys ei olisi suuri, yksikin väärä vastaus tai epä johdonmukainen toiminta voi vääristää päätöksentekoa, heikentää järjestelmän uskottavuutta tai vaarantaa tietoturvan ja tietosuojan. Jos järjestelmä toimii epä johdonmukaisesti poikkeustilanteissa, seuraukset voivat vaihdella harmittomista

väärinymmärryksistä jopa vakaviin päätöksentekovirheisiin tai arkaluonteisten tietojen paljastumiseen. Todennäköisyys riippuu järjestelmän kypsyydestä ja testauksen kattavuudesta, mutta riskiä ei voi pitää vähäisenä, koska monimutkaiset järjestelmät voivat tuottaa ennalta-arvaamattomia virheitä.

Luotettavuus on keskeinen edellytys tekoälyn turvalliselle käytölle. Sen puutteet kertautuvat ja heijastuvat muihin turvallisuuden osa-alueisiin, mikä tekee luotettavuudesta koko riskienhallinnan kannalta kriittisen tekijän.

Luotettavuudessa nousee esiin kolme keskeistä huomiota:

1. Johdonmukaisuus ja ennustettavuus:

Jos järjestelmä ei toimi samankaltaisissa tilanteissa samalla tavalla, riski vääristä lopputuloksista kasvaa. Tämä voi saada korkean vaikutuspisteytyksen, koska epäluotettava järjestelmä heikentää koko ratkaisun käyttökelpoisuutta.

2. Tietoturva ja manipulointiriski:

Epäluotettava tekoäly voi olla herkempi hyökkäyksille (esim. syötteiden manipulointi). Todennäköisyys voi olla keskitasoa, mutta vaikutus korkea, sillä järjestelmän tulokset voivat muuttua harhaanjohtaviksi.

3. Tietosuoja:

Luotettavuus kattaa myös sen, että järjestelmä ei tahattomasti vuoda henkilötietoja. Vaikka tietosuojariskin todennäköisyys voidaan pitää matalana hyvillä teknisillä ratkaisuilla, sen vaikutus on merkittävä, jos järjestelmä kuitenkin epäonnistuu.

Näin luotettavuus muodostaa riskien arvioinnin ytimen: sen painoarvo on suuri, sillä sekin toimii läpileikkaavana tekijänä, joka yhdistää tekniset, toiminnalliset ja oikeudelliset riskit yhdeksi kokonaisuudeksi.

8.4 Oikeudenmukaisuus

Oikeudenmukaisuus sote-alan tekoälyn riskienhallinnassa tarkoittaa sitä, että järjestelmä kohtelee kaikkia yksilöitä ja ihmisryhmiä tasapuolisesti eikä vahvista tai luo uusia eriarvoisuuksia. Tämä edellyttää, että tekoälyn käyttämät lähtötiedot ovat mahdollisimman kattavia ja edustavia niin, että mallit on rakennettu tunnistamaan ja minimoimaan ennakkoluuloista tai puutteellisesta datasta johtuvat vinoumat. Oikeudenmukaisuus liittyy myös päätösten läpinäkyvyyteen ja perusteltavuuteen: ammattilaisten ja potilaiden/asiakkaiden on voitava ymmärtää, millä perusteella tekoäly on tehnyt ehdotuksia tai arvioita. Riskienhallinnan näkökulmasta oikeudenmukaisuuden puute voi johtaa esimerkiksi palveluiden epätasaiseen saatavuuteen, syrjiviin käytäntöihin tai heikentyneeseen luottamukseen järjestelmää ja päätöksentekoa kohtaan. Näin ollen oikeudenmukaisuus on keskeinen periaate, joka varmistaa, että tekoälyn hyödyntäminen sote-alalla tukee yhdenvertaisuutta.

8.4.1 Havainnot kokeiluista: oikeudenmukaisuus riskienhallinnassa

Oikeudenmukaisuuden näkökulmasta merkittävin riski liittyy koulutusdatan vinoumiin ja niiden heijastumiseen tekoälyn päätöksiin. Todennäköisyys tälle riskille on korkea, sillä lähes kaikki tekoälymallit perustuvat laajoihin aineistoihin, joissa yhteiskunnalliset ennakkoluulot ja epätasapainot ovat läsnä. Vastaavasti suppeiden aineistojen tilanteessa tietopohja ei välttämättä ole tarpeeksi edustava.

Koulutusdatan vaikutukset voivat vaihdella yksittäisistä virheellisistä arvioista aina laajamittaisiin syrjinnän mekanismeihin, jotka vahvistuvat tekoälyn toistamana. Oikeudenmukaisuuden ja läpinäkyvyyden puute kasvattavat yhdessä riskiä entisestään, sillä jos järjestelmän toimintaa ei voi selittää, virheiden tunnistaminen ja korjaaminen vaikeutuu, mikä lisää vaikutusten vakavuutta.

Oikeudenmukaisuuden keskeinen tehtävä on varmistaa yhdenvertainen kohtelu ja syrjimättömyys. Riskit korostuvat erityisesti silloin, kun tekoälyn päätöksillä on suoria yhteiskunnallisia vaikutuksia, kuten sote-palveluiden kohdentamisessa, resurssien jakamisessa tai asiakkaiden arvioinnissa. Vaikutukset nousevat kriittisiksi etenkin silloin, kun epäoikeudenmukaisuus kohdistuu jo valmiiksi haavoittuvassa asemassa oleviin ryhmiin. Näin ollen oikeudenmukaisuuteen liittyviä riskejä ei voi pitää vain teknisinä haasteina, vaan ne heijastuvat laajasti yhteiskunnan luottamukseen tekoälyn käyttöä kohtaan.

Oikeudenmukaisuudessa nousee esiin kolme keskeistä huomiota:

1. Koulutusdatan vinoumat ja niiden vaikutus:

Jos tekoälymallin pohjana oleva data sisältää ennakkoluuloja tai epätasapainoja, riski syrjivistä päätöksistä kasvaa merkittävästi. Tämä saa korkean vaikutuspisteytyksen, sillä vinoumat voivat johtaa järjestelmälliseen epäoikeudenmukaisuuteen.

2. Läpinäkyvyyden puute ja virheiden korjattavuus:

Kun järjestelmän toimintaa ei voida selittää, virheiden tunnistaminen ja korjaaminen vaikeutuu. Todennäköisyys tälle riskille on keskitasoa, mutta vaikutus korkea, koska selittämättömät päätökset heikentävät luottamusta ja lisäävät virheiden vakavuutta.

3. Haavoittuvien ryhmien erityinen riski:

Epäoikeudenmukaisuus korostuu erityisesti silloin, kun tekoälyn päätökset kohdistuvat jo valmiiksi heikossa asemassa oleviin ryhmiin. Vaikka todennäköisyys voi vaihdella, vaikutus on kriittinen, sillä seuraukset voivat olla laajamittaisia ja yhteiskunnallisesti merkittäviä.

Näin oikeudenmukaisuus muodostaa riskienhallinnan ytimen: sen painoarvo on suuri, koska se vaikuttaa suoraan luottamukseen tekoälyn käyttöä kohtaan ja heijastuu teknisten ratkaisujen lisäksi laajemmin yhteiskunnalliseen oikeudenmukaisuuteen.

8.5 Osallistaminen

Osallistaminen sote-alan tekoälyn riskienhallinnassa tarkoittaa sitä, että tekoälyn suunnitteluun, käyttöönottoon ja arviointiin pyritään ottamaan mukaan eri näkökulmia edustavat toimijat. Kun järjestelmän kehittämisessä kuullaan niitä, joihin ratkaisut käytännössä vaikuttavat, voidaan tunnistaa riskejä, joita tekninen tai hallinnollinen näkökulma ei yksin toisi esiin. Osallistaminen lisää järjestelmän hyväksyttävyyttä, tukee luottamuksen rakentumista ja parantaa mahdollisuuksia tunnistaa sekä ehkäistä epätoivottuja seurauksia jo varhaisessa vaiheessa. Riskienhallinnan kannalta se on keskeinen keino varmistaa, että tekoäly palvelee eri käyttäjäryhmien tarpeita, eikä luo uusia esteitä, epäluottamusta tai eriarvoisuutta palveluiden saatavuudessa. Hyvä viestintä motivoi osallistamista.

8.5.1 Havainnot kokeiluista: osallistaminen riskienhallinnassa

Laaja osallistujajoukko luo pohjan kattavalle riskienhallinnalle. Läpinäkyvyyden puute tai datan vääristymät ovat riskejä, joiden esiintymistä voidaan ehkäistä, kun järjestelmien suunnitteluun ja arviointiin otetaan mukaan eri taustoista tulevia ihmisiä. He voivat tunnistaa ongelmia, joita pelkät tekniset kehittäjät eivät välttämättä huomaa. Samalla vaikutusten vakavuus pienenee, koska osallistaminen luo ennakoivia korjausmekanismeja, koska riskeihin voidaan puuttua jo ennen laajaa käyttöönottoa.

Osallistamiseen liittyy kuitenkin myös omia riskejään. Jos se toteutetaan näennäisesti tai vain rajatulle joukolle, syntyy väärä turvallisuuden tunne ja kriittisiä näkökulmia voi jäädä huomioimatta. Osallistaminen ei ole tekninen ominaisuus, vaan turvallisuuden hallinnan menetelmä, jonka arvo riippuu sen systemaattisuudesta ja aidosta sitoutumisesta eri ryhmien kuulemiseen. Hyvin toteutettuna se vähentää epäluottamusta ja vastustusta, parantaa riskien tunnistamisen kattavuutta ja voi muuttaa joidenkin riskien luonteen hallittavammaksi. Huonosti toteutettuna se puolestaan voi lisätä vastareaktioita, jos osallistumista pidetään vain muodollisuutena ja suppeana.

Osallistamisessa nousee esiin kolme keskeistä huomiota:

1. Monimuotoisuus ja näkökulmien kattavuus:

Kun riskienhallintaan osallistetaan eri taustoista tulevia ihmisiä, voidaan tunnistaa ongelmia laaja-alaisemmin teknisten riskien lisäksi. Tämä vähentää riskiä, että järjestelmään jää piileviä virheitä tai vinoumia, ja parantaa ratkaisun kokonaisvaltaista luotettavuutta.

2. Aito sitoutuminen:

Osallistaminen ei ole pelkkä muodollisuus. Jos se toteutetaan näennäisesti tai vain rajatulle joukolle, syntyy väärä turvallisuuden tunne ja kriittisiä näkökulmia voi jäädä huomioimatta. Rajatun joukon ulkopuolelle jääneet voivat kokea epäluottamusta ja sitoutumisen puutetta järjestelmään. Lisäksi riski vääristä päätöksistä kasvaa, koska prosessi ei ole läpinäkyvä eikä kattava.

3. Ennakoivat korjausmekanismit:

Hyvin toteutettu osallistaminen luo mahdollisuuden puuttua riskeihin jo ennen kuin riskit realisoituvat. Tämä pienentää riskien toteutumisen vakavuutta, koska ongelmat havaitaan ja korjataan varhaisessa vaiheessa, mikä tekee riskienhallinnasta proaktiivista eikä reaktiivista.

Osallistaminen toimii riskienhallinnassa keskeisenä tekijänä: se ei ole tekninen ominaisuus, vaan menetelmä, joka vahvistaa turvallisuutta ja luottamusta. Kun se toteutetaan systemaattisesti ja aidosti, se vähentää epäluottamusta, parantaa riskien tunnistamisen kattavuutta ja tekee monista riskeistä hallittavampia. Huonosti toteutettuna se voi kuitenkin kääntyä itseään vastaan ja lisätä vastareaktioita.

Tekoälyn turvallisen käytön 5 periaatetta



Vastuullisuus on eettinen velvoite.

Se vaatii riskien ennakkointia, jatkuvaa valvontaa ja korjaavia toimia haittojen estämiseksi.



Läpinäkyvyys rakentaa luottamusta.

Tekoälyn toiminnan ja päätösten on oltava ymmärrettäviä ja perusteltuja käyttäjille.



Luotettavuus takaa johdonmukaisuuden.

Järjestelmän tulee toimia tarkasti ja ennakoitavasti sekä tunnistaa omat rajoituksensa.



Oikeudenmukaisuus estää syrjintää.

Tekoälyn on kohdeltava kaikkia tasapuolisesti eikä se saa vahvistaa olemassa olevia vinoumia.



Osallistaminen tuo eri näkökulmia.

Eri käyttäjäryhmien kuuleminen auttaa tunnistamaan ja ehkäisemään riskejä jo varhaisessa vaiheessa.

9. Tekoälyn käytön esteet

Sosiaali- ja terveydenhuollossa tekoälyn käyttöä Suomessa rajoittavat useimmiten samat tekijät. Ensinnäkin tiedot ovat hajallaan eri järjestelmissä, niiden laatu vaihtelee ja

lainsäädäntö asettaa rajoituksia niiden käyttöön. Toiseksi tietosuoja ja vastuuongelmat herättävät epävarmuutta: kuka kantaa vastuun, jos tekoäly tekee virheen? Kolmanneksi monilla organisaatioilla ei ole riittävästi osaamista, resursseja tai luottamusta uuteen teknologiaan. Myös ohjeistukset ja sääntelyn tulkinnat voivat olla epäselviä, mikä lisää varovaisuutta käyttöönnotossa. Lisäksi tekoälyratkaisujen hankinta, kustannukset ja sovittaminen vanhoihin järjestelmiin tuottavat käytännön haasteita. Vaikeutena on myös hajanaiset hallintamallit sote-sektorilla. Suurimmat esteet eivät siis ole vain teknisiä, vaan liittyvät myös sääntöihin, toimintatapoihin ja varovaisiin asenteisiin.

Sote-sektorilla tekoälyn käyttöä hidastavat esteet muodostuvat monimutkaisesta kokonaisuudesta, jossa juridiset tulkinnat, tekniset rajoitteet ja hallinnollinen epäselvyys kietoutuvat toisiinsa. Esimerkiksi THL:n ja hyvinvointialueiden erilaiset linjat pilvipalveluiden käytössä paljastavat, miten tietosuojan tulkinta voi käytännössä vaikeuttaa yhteisten alustojen rakentamisen ja on voinut johtaa datojen rajattuun käyttöön. Tämä ei ole vain tekninen kysymys, vaan se heijastuu suoraan tutkimuksen ja tulosten hyödynnettävyyteen. Samalla EU:n tekoälyasetuksen riskiluokittelu voi luoda kustannuksia ja epävarmuutta, kun varmuuden vuoksi kohdistetaan raskaita vaatimuksia myös kevyisiin sovelluksiin lain tulkinnanvaraisuuden takia. Tekoälyasetus nivoutuu myös muuhun ympäröivään lainsäädäntöön aiheuttaen lisävaatimuksia ja laajempaa riskienhallintaa.

Kun tähän lisätään lääkinnällisten laitteiden ja tekoälyn sääntelyn tiukat vaatimukset, muodostuu kokonaisuus, jossa riskien hallinta voi nousta innovoinnin edelle. Kyse ei siis ole yksittäisistä haasteista, vaan rakenteellisesta esteestä: miten yhdistää tietosuoja ja potilasturva ja toisaalta ketterä kehittäminen tietosuojan ja korkeamman riskin lääkinnällisten laitteiden yhteen toimivuudessa. Tämän esteen ratkaiseminen edellyttää johdonmukaista, riskiperusteista ohjausta ja lääkinnällisten laitteiden tunnistamista.

Profiloinnin sääntelyn tulkinnanvaraisuus voi puolestaan estää tekoälyn soveltamisen palvelujen kohdentamiseen, vaikka juuri ennaltaehkäisevässä terveydenhuollossa ja palvelujen kohdentamisessa voitaisiin saavuttaa merkittäviä hyötyjä. Rekisterinpitäjiä on paljon ja yhteisrekisterinpitäjyyttä ei ole mahdollista toteuttaa laajemmin lainsäädäntöesteiden vuoksi, mikä tekee yhdistellyn datan käytöstä vaikeaa. Tämä voi johtaa siihen, että kehittäminen perustuu pienempiin tietolähteisiin, jos yhteistyötä on vähemmän eri rekisterinpitäjien kanssa. Tämä voi esiintyä datan vinoutumisena ja vertailukelpoisuuden puutteena.

Matalan riskin sote-sovelluksille tulisi avata kevyempi ja nopeampi kehityspolku, joka voisi rakentua yhteisten tietomallien ja kansallisen pilvi-infrastruktuurin varaan, jotka voivat hyödyntää olemassa olevia palveluita ja tietoturvaratkaisuja ilman raskaita lupaprosesseja. Näin innovaatiot voisivat edetä hallitusti prototyypeistä käytäntöön ja yksittäiset hankkeet kytkeytyisivät luonnollisesti osaksi koko sote-järjestelmän toimintaa, esimerkiksi laajemmin yhteisrekisterinpitäjien mahdollistumisen kautta.

10. Käytännöt havainnot hyvinvointialueiden tekoälypilottien riskienarvioinneista

Seuraavaksi luvassa on yhteenveto siitä, mitä riskienarvioinnit toivat esiin tekoälypilottien käyttöönotosta sosiaali- ja terveydenhuollossa. Luvussa käsitellään vakavimpia kokonaisriskejä ennen ja jälkeen hallintakeinojen ja niiden jäännösriskien luonnetta. Näiden havaintojen avulla tunnistetaan, mitkä riskit ovat hallittavissa ja mitkä jäävät merkittäviksi haasteiksi.

10.1 Vakavimmat kokonaisriskit

Kokonaisriskit vaihtelivat kokeiluprojekteissa riskitulona välillä 0–25 pistettä (todennäköisyys x vaikutus).

Vakavimpien kokonaisriskien (16 pistettä) projekteja yhdisti se, että ne käsittelivät tekoälyä yhteiskunnallisesti herkillä alueilla, joissa epäonnistuminen ei tarkoita vain huonoa teknistä tulosta, vaan luottamuksen, hyväksyttävyyden ja legitimitietin menetystä. Riskit liittyivät erityisesti seuraaviin aiheisiin:

- EU AI Act -tulkinta- ja luokitteluriskit. Seurauksena voi olla järjestelmän virheellinen riskiluokitus, josta voi tulla merkittäviä oikeudellisia ja toiminnallisia seurauksia.
- Luottamuksen puute tekoälyn tuottamiin tuloksiin. Seurauksena voi estää käytön, heikentää päätöksenteon legitimitettä ja voi romuttaa kokeilun vaikuttavuuden.
- Sidosryhmäviestinnän ja vuorovaikutuksen epäonnistuminen. Seurauksena voi johtaa vastustukseen, väärinymmärryksiin ja poliittiseen tai organisatoriseen torjuntaan.
- Sosiaalisen kestävyyspeittäminen. Seurauksena tekoäly voidaan kokea epäoikeudenmukaisena, syrjivänä ja yhteiskunnallisesti hyväksymättömänä.

Muita korkeita riskejä (9–12 pistettä) olivat mm.:

- Datan laatu ja riittävyys.
- Osaamisvaje käyttäjillä tai henkilöstöllä.
- Vastuiden epäselvyys tekoälyn kehittämisessä ja käytössä.

Näiden korkean riskin kohdalla hallintakeinoina mainittiin mm. koulutus, sidosryhmäosallistaminen, riskiluokitusten prosessien dokumentointi ja tekoälypolitiikan määrittely. Näistä huolimatta osassa riskeistä jäännösriski jäi merkittäväksi.

10.2 Vakavimmat kokonaisriskit hallintakeinojen jälkeen

Jäännösriskit vaihtelivat riskituloltaan 0–9 pistettä (ennen hallintakeinoja vakavimmat olivat 16 pistettä), mikä kertoo, että useimmat riskit saatiin hallintakeinoilla merkittävästi pienemmiksi.

Kuitenkin jäännösriskit (8–9 pistettä) osoittivat, että osa ongelmista säilyy vaikeasti hallittavina.

Jäännösriski (9 pistettä):

- Tekoälyn käytön kulttuuri ja ymmärrys ei kehity
- Käyttäjien tai henkilöstön osaaminen ei riitä
- Järjestelmä luokitellaan virheellisesti korkean riskin järjestelmäksi (EU AI Act)
- Palvelun laatu ei vastaa tavoitteita tai käyttäjien tarpeita
- Luottamus tekoälyn tuottamiin tuloksiin ei synny

Nämä ovat ihmisiin, osaamiseen, tulkintaan ja hyväksyttävyyteen liittyviä riskejä, joita ei voi täysin poistaa prosesseilla tai teknisillä ratkaisuilla.

Jäännösriski (8 pistettä):

- EU AI Act -sanktoriskit

Vaikka hallintakeinoja on määritelty, sääntelyyn liittyvä epävarmuus jää pysyväksi riskiksi, koska tulkinnat ja käytäntö kehittyvät.

10.3 Johtopäätös merkittävistä jäännösriskeistä

Hallintakeinojen jälkeen vakavimmat riskit eivät ole enää ensisijaisesti teknisiä, vaan liittyvät ihmisten osaamiseen, toimintakulttuuriin, luottamukseen ja sääntelyn tulkintaan. Ne ovat luonteeltaan pitkäkestoisia ja vaativat jatkuvaa johtamista, eivät pelkkää projektivaiheen riskienhallintaa.



11. Huomiot tekoälyn vaikutusten arvioinnissa riskien pohjalta

Tekoälyn käyttöönotto sosiaali- ja terveydenhuollossa edellyttää perusteellista riskien arviointia, jossa DPIA eli tietosuojaa koskeva vaikutustenarviointi on keskeinen työkalu. DPIA:n avulla tunnistetaan henkilötietojen käsittelyyn liittyvät riskit, arvioidaan vaikutukset yksityisyyteen ja varmistetaan, että ratkaisut ovat GDPR:n ja muun lainsäädännön mukaisia. Arvioinnissa on huomioitava myös EU AI Actin vaatimukset sekä mahdolliset muutokset sääntelyssä ja tulkinnoissa, sillä nämä voivat vaikuttaa käyttöönoton aikatauluun ja taloudellisiin vaikutuksiin.

Lääkinnällisiä laitteita koskevan asetuksen (MDR) soveltuvuus on toinen keskeinen kysymys. On selvittävä, kuuluuko tekoälyratkaisu MDR:n piiriin, sillä tämä vaikuttaa sertifiointiin, markkinoille pääsyyn ja turvallisuusvaatimuksiin. Samalla on arvioitava teknologian luotettavuus ja integraatio olemassa oleviin potilastietojärjestelmiin, jotta järjestelmät toimivat saumattomasti ja potilasturvallisuus säilyy.

Eettinen näkökulma korostuu erityisesti vinoumien hallinnassa. Tekoälymallit voivat vahvistaa ennakkoluuloja tai syrjiviä käytäntöjä esimerkiksi eri väestöryhmien hoidossa. Vinoumien tunnistamiseen ja korjaamiseen on määriteltävä menetelmät, kuten vastausten pisteytys, mallien vertailu ja läpinäkyvyyden varmistaminen. Tarvittaessa voidaan palata mallin aikaisempiin versioihin tai arvioida tuloksia toisella kielimallilla. Näin varmistetaan, että tekoäly toimii oikeudenmukaisesti ja tukee tasa-arvoa.

Tietosuoja on keskeinen osa arviointia. Potilastietojen käsittely lisää riskiä tietovuodoille ja väärinkäytöksille, joten anonymisointi ja pseudonymisointi ovat välttämättömiä keinoja riskien hallinnassa. Samalla on varmistettava datan laatu, sillä huono laatu voi heikentää tekoälyn luotettavuutta ja johtaa virheellisiin hoitosuosituksiin.

Organisaation näkökulmasta tekoälyn käyttöönotto ei ole pelkästään tekninen hanke, vaan se vaatii toiminnan muutoksia ja prosessien uudistamista. Ammattilaiset ja muu henkilöstö voivat kokea työnkuvan muuttuvan henkilökohtaisista syistä, tekoälyn tai tekoälyn käyttämisen myötä. Tämä voi vaikuttaa tekoälyn hyväksyttävyyteen ja henkilöstön jaksamiseen ja motivaatioon. Lisäksi on määriteltävä vastuut virhetilanteissa, esimerkiksi kuka kantaa vastuun, jos tekoäly tekee virheen hoitopäätöksessä.

Toisin sanoen luottamus ja hyväksyttävyys ovat ratkaisevia tekijöitä tekoälyn käytössä. Potilaiden ja henkilöstön luottamus tekoälyjärjestelmiin vaikuttaa hoidon laatuun ja käyttöönoton onnistumiseen. Viestinnän ja koulutuksen riittävyys on varmistettava, jotta tekoälyn käyttö koetaan turvalliseksi ja hyödylliseksi. On myös huomioitava potilaan oma tahto: yksilöllä tulee olla mahdollisuus tulla kuulluksi, jos hän ei halua tekoälyä käytettävän omassa hoidossaan.

Kaikkien näiden näkökulmien yhdistäminen on välttämätöntä. Vaikutusten arvioinnissa tulee kattaa kliininen laatu ja potilasturvallisuus, eettisyys ja oikeudenmukaisuus, tietosuoja ja lainsäädäntö, työntekijöiden ja potilaiden hyväksyttävyyss sekä teknologinen luotettavuus. Vain näin voidaan varmistaa, että tekoälypalvelut tukevat sosiaali- ja terveydenhuollon tavoitteita ilman, että potilasturvallisuus tai yksilön oikeudet vaarantuvat.

Liitteessä 1 on tuotu havaintoja, mitä vaikutusten arviointi voisi sisältää.

12. Tekoälyn hallintoratkaisut

Tekoälyn hallintaa tulee tarkastella hallintamallin kautta, joka viittaa organisaation sisäisiin toimintaperiaatteisiin ja käytäntöihin, sekä AI governance -näkökulmasta, joka kattaa laajemmat periaatteet, sääntelyn ja rakenteet, joilla tekoälyn käyttöä ohjataan yhteiskunnan, toimialan ja organisaation tasoilla.

Tekoälyn hallinta eli AI governance on keskeinen osa riskienhallintaa sosiaali- ja terveydenhuollossa, koska se varmistaa, että tekoälyratkaisut toimivat läpinäkyvästi, vastuullisesti ja sääntelyn mukaisesti. Markkinoilla on useita työkaluja, jotka tukevat tekoälyn hallintaa. Tällaiset ratkaisut voivat auttaa organisaatioita hallitsemaan tekoälyn elinkaarta ja varmistamaan, että käytössä olevat mallit ovat luotettavia ja eettisiä.

AI governance -ratkaisujen rooli on laajempi kuin pelkkä tekninen valvonta. Ne linkittävät riskit tekoälypolitiikkaan ja organisaation sisäisiin prosesseihin. Tämä tarkoittaa, että riskienhallinta ei ole irrallinen toiminto, vaan osa strategista johtamista. Sosiaali- ja terveydenhuollossa tämä on erityisen tärkeää, koska riskit liittyvät potilasturvallisuuteen, tietosuojaan ja eettisiin kysymyksiin. Hallintaratkaisut voivat sisältää ominaisuuksia, kuten mallien auditointi, vinoumien tunnistus, datan laadun seuranta ja dokumentointi, jotka tukevat DPIA:n ja muiden arviointien toteuttamista.

Organisaatiot voivat ottaa käyttöön AI governance -työkaluja, jotka mahdollistavat riskien jatkuvan seurannan ja hallinnan. Näiden avulla voidaan varmistaa, että tekoälyratkaisut täyttävät EU AI Actin, GDPR:n ja MDR:n vaatimukset. Lisäksi voidaan luoda sisäisiä politiikoita, jotka määrittelevät vastuut, hyväksyttävyyden kriteerit ja toimintamallit virhetilanteissa. Ero eri työkalujen välillä liittyy usein siihen, kuinka kattavasti ne tukevat sääntelyä, läpinäkyvyyttä ja integraatiota olemassa oleviin järjestelmiin.

Isossa kuvassa tarvitaan järjestelmä, joka yhdistää teknologisen hallinnan, eettiset periaatteet ja organisaation prosessit. Tämä tarkoittaa, että AI governance ei ole vain ohjelmisto, vaan kokonaisvaltainen toimintamalli, joka tukee potilasturvallisuutta, eettisyyttä, oikeudenmukaisuutta ja luottamusta tekoälyn käytössä.

12.1 Tarkemmin AI governance -työkaluista

AI governance ei ole pelkästään vaatimustenmukaisuuden (compliance) varmistamista. Se muodostaa kokonaisuuden, joka mahdollistaa tekoälyprojektien toteuttamisen sekä organisaation tekoälykehityksen ja -valvonnan koko tekoälymallien elinkaaren ajan. Tavoitteena on varmistaa sekä laatu että haluttu palvelutaso ja tehokkuus.

Keskeisiä AI governance -kokonaisuuden osa-alueita ovat:

Dokumentaatio

Mallien systemaattinen dokumentointi sekä sallitun käytön määrittely, validointi ja auditoitavuuden varmistaminen. Dokumentaatio luo perustan läpinäkyvyydelle ja jäljitettävyydelle.

Monitorointi

Kehitysvaiheessa monitorointi tukee mallien arviointia ja varmistaa, että ne täyttävät vaaditun tarkkuustason. Tuotantoon siirron jälkeen monitorointi mahdollistaa mallien jatkuvan valvonnan ja laadun seurannan.

Riskienhallinta

Riskienhallinta yhdistää monitoroinnin, dokumentaation ja vastuunjaon. Se tarjoaa kokonaiskuvan organisaatiossa käytössä olevista tekoälyratkaisuksista ja mahdollistaa automatisoidut työnkulut, joissa asian omistajat pysyvät tietoisina tekoälyn käyttöön ja kehitykseen liittyvistä muutoksista ja riskeistä.

Yhdessä nämä kolme osa-aluetta mahdollistavat organisaation hallitun ja läpinäkyvän tekoälyn käytön ja kehittämisen sekä varmistavat valmiuden sisäiseen ja ulkoiseen auditointiin.

Vastuullinen tekoäly: Hallinnan 3 kulmakiveä

Systemaattinen dokumentointi



Luo perustan läpinäkyvyydelle ja jäljitettävyydelle, mahdollistaen mallien auditoinnin.

Jatkuva monitorointi



Varmistaa tekoälymallien laadun ja suorituskyvyn koko niiden elinkaaren ajan.

Kokonaisvaltainen riskienhallinta



Yhdistää dokumentaation ja valvonnan osaksi strategista johtamista ja vastuunjakoa.

13. Toimenpidesuosituksia

13.1 Kansallisen ohjauksen ja hallintamallin kehittäminen

Yksi keskeinen este tekoälyn hyödyntämiselle sote-sektorilla on yhteisen hallintamallin puuttuminen. Tällä hetkellä vastuut ja ohjaus jakaantuvat monille toimijoille (STM, THL, hyvinvointialueet, Kela ja eri viranomaiset), mutta selkeää kansallista mallia ei ole. Tämä johtaa siihen, että eri alueet ja organisaatiot tekevät omia ratkaisujaan, mikä voi aiheuttaa hajanaisuutta ja eritahtisuutta tekoälyn käyttöönotossa. Käytännössä tämä näkyy esimerkiksi siinä, että pilottihankkeet jäävät paikallisiksi, eivätkä tulokset skaalaudu kansallisesti, koska yhteisiä pelisääntöjä ei ole. Myös ohjeistusten tulkinnat vaihtelevat, mikä luo epävarmuutta hankinnoissa ja hidastaa päätöksentekoa.

Ratkaisuksi suositellaan SOTE-tekoälyn ekosysteemin puitteissa tehtäväksi kansallista SOTE-tekoälyn koordinaatiomallia, joka loisi selkeän viitekehyksen tekoälyn käytölle ja määrittelisi vastuunjaon viranomaisten, hyvinvointialueiden ja palveluntuottajien välillä. Tällainen malli voisi kattaa esimerkiksi riskiluokittelun, tietomallien harmonisoinnin ja pilvipalvelujen hyväksyttävyyden linjaukset. Yhteinen hallintamalli vahvistaisi luottamusta järjestelmiin, vähentäisi epävarmuutta ja loisi edellytykset sille, että tekoälyratkaisut voitaisiin ottaa käyttöön yhdenmukaisesti koko Suomessa.

Kehitysehdotuksena perustetaan SOTE-tekoälyn ekosysteemiin alatyöryhmä, joka vastaa kansallisen ohjauksen ja hallintamallin kehittämisestä:

- Sovitaan/laaditaan yhteinen hallintamalli, joka määrittelee ohjauksen, riskienhallinnan ja vastuut selkeästi.
- Luodaan valtakunnalliset hankinnan periaatteet ja kehitysperiaatteet piloteille, jotta ratkaisuja voidaan skaalata alueelta toiselle.
- Kehitetään yhteinen seurantamalli, joka tuottaa tietoa tekoälyratkaisujen hyödyistä, riskeistä ja kustannuksista.
- Sovitaan hallintamallin ylläpitovastuut

13.2 Tekoälyn käyttöönoton yhteiskunnallisten vaikutusten arviointi sote-sektorilla: kehityskohteet

Tekoälyn käyttöönotto sosiaali- ja terveydenhuollossa tarjoaa suuria mahdollisuuksia palvelujen laadun, tehokkuuden ja saavutettavuuden parantamiseen. Suomessa kuitenkin puuttuu toistaiseksi kattava yhteiskunnallinen vaikuttavuus selvitys, joka arvioisi tekoälyn laajemmat vaikutukset sote-järjestelmään, potilaisiin/asiakkaisiin ja ammattilaisiin. Tämä tekee kansallisesti haasteelliseksi ymmärtää esimerkiksi, miten tekoäly muuttaa työnkuvia, resurssien kohdentamista tai eri väestöryhmien yhdenvertaisuutta kokonaisuudessaan. Tällä hetkellä tekoälyratkaisut arvioidaan pistemäisesti ja kansallisella tasolla oleva strategista

vaikutustenarviointia ei ole tehty, miten tekoäly tulee vaikuttamaan poikkileikkaavasti sote-sektorilla.

Ilman vaikutusten arviointia on myös vaikea ennakoida eettisiä riskejä, kuten tekoälyn luotettavuuteen liittyviä, syrjintää ja päätöksenteon läpinäkyvyyden ja vastuunjaon haasteita. Ilman kansallista vaikutusten arviointia lainsäädännön ja eettisten ohjeistusten kehittäminen voi jäädä jälkeen suhteessa teknologian nopeaan kehitykseen. Sote-sektorin onnistunut tekoälyintegraatio edellyttääkin, että vaikutukset arvioidaan kokonaisvaltaisesti kliinisten, sosiaalisten ja taloudellisten näkökulmien lisäksi myös yhteiskunnallisen vaikuttavuuden kautta. Tämä mahdollistaa paitsi turvallisen ja yhdenvertaisen käytön myös sen, että tekoäly tukee kestävästä kehitystä ja yhteiskunnan hyvinvointia tulevaisuudessa.

Kehitysehdotuksena toteutetaan kartoitus tekoälyn käyttöönoton yhteiskunnallisten vaikutusten arvioinnista sote-sektorilla:

- Miten tekoäly vaikuttaa hoitopäätösten tarkkuuteen, potilasturvallisuuteen ja hoitoprosessien virheettömyyteen sekä, miten positiiviset hyödyt ja mahdolliset riskit vaikuttavat, esimerkiksi virheelliset suositukset tietyissä potilasryhmissä.
- Miten tekoälyjärjestelmien mahdolliset ennakkoluulot, syrjivät vaikutukset ja päätösten läpinäkyvyys ovat tunnistettavissa. Arvioinnissa on varmistettava, että teknologia tukee oikeudenmukaisuutta ja yhdenvertaista hoitoa kaikille väestöryhmille.
- Miten tekoäly muuttaa työnkuvia ja mitkä ovat koulutustarpeet. Samalla on huomioitava potilaiden ja henkilöstön luottamus sekä järjestelmien selitettävyys eli perustelujen käyttäminen.

13.3. Tekoälyosaamisen vahvistaminen ja tekoälyn käytön edistäminen

Ehdotuksena on tehdä kansallinen tiekartta (roadmap), jonka avulla mahdollistetaan Suomeen tekoälykulttuuri, joka sisältää turvallisen ja ennakoivan käytön. Tämä edellyttää samanaikaisesti tekoälyosaamisen vahvistamista, datan laadun järjestelmällistä parantamista, lainsäädännöllisten ja tulkinnallisten esteiden poistamista. Lisäksi tarvitaan valtakunnallista linjausta siitä, mitä tietoja voidaan käsitellä tekoälyllä pilvessä ja mikä vaatii sensitiivisemmän ympäristön. Tiekartta voisi sisältää kehityspolun vaiheittain usealle vuodelle, ja se soveltuisi sekä kansalliseen ohjaukseen että organisaatiotason toteutuksen suunnitteluun.

Tekoälykulttuuri syntyy, kun ihmisillä on lupa, kyky ja syy käyttää tekoälyä vastuullisesti. Kehitysehdotuksena tekoälyosaamisen vahvistamiseksi ja tekoälyn käytön edistämiseksi tiekartta-aiheiksi:

- Tiekartan linkittyminen kansallisiin periaatteisiin ja strategioihin
- Data- ja tekoälyosaamisen koulutusohjelmat käyntiin sote-sektorilla.
- Datan laadun strateginen kartoitus.

- Yhtenäinen tulkinta sote-sektorilla pilvessä toimivan tekoälyn käytöstä ja tietosuojasioista.
- Aloitetaan ensimmäiset skaalautuvat pilotit pistemäisten kokeilujen sijaan.

14. Kvanttitekoäly ja robotiikka – tulevaisuuden riskienhallinnan näkökulmia

Vaikka tässä dokumentissa ei vielä käsitellä kvanttitekoälyä ja robotiikkaa, on tärkeää tunnistaa näiden teknologioiden kasvava merkitys sosiaali- ja terveydenhuollon digitalisaatiossa. Kvanttitekoäly yhdistää kvanttilaskennan ja tekoälyn mahdollisuudet, mikä voi tulevaisuudessa mullistaa datan käsittelyn, diagnostiikan ja hoitoprosessien optimoinnin.

Robotiikka puolestaan tuo uusia ulottuvuuksia automaatioon, hoitotyön tukemiseen ja palvelujen saavutettavuuteen. Näiden teknologioiden käyttöönotto sote-sektorilla edellyttää ennakoivaa riskienhallintaa, jossa arvioidaan muun muassa päätöksenteon läpinäkyvyyttä, tietoturvaa, eettisyyttä ja vastuunjakoa.

SOTE-tekoälyn ekosysteemin riskienhallintaryhmän olisi suositeltavaa perehtyä kvanttitekoälyyn ja robotiikkaan sekä niiden erityispiirteisiin tulevien riskienhallinnan haasteiden ymmärtämiseksi. Sote-toimiala on sensitiivinen ja henkilötiedon turvaaminen ja suojaaminen vaativat erityistä huomiota.

15. Yhteenveto

Tässä raportissa on tarkasteltu tekoälyä hyödyntävien palveluiden riskienhallintaa sosiaali- ja terveydenhuollossa SOTE-tekoälyn ekosysteemin kokeiluprojektien, riskienhallintaryhmän työn ja ajankohtaisen EU-sääntelyn pohjalta. Keskeinen lähtökohta on, että tekoälyn käyttö sote-sektorilla edellyttää systemaattista, ennakoivaa ja eettisesti kestävä riskienhallintaa, koska tekoälyratkaisut vaikuttavat suoraan ihmisten oikeuksiin, palvelujen laatuun ja potilasturvallisuuteen. Riskienhallinta ei ole erillinen tekninen vaihe, vaan osa kokonaisvaltaista vaikutusten arviointia ja toiminnan ohjausta.

Raportissa tunnistetut viisi keskeistä riskiluokkaa, vastuullisuus, läpinäkyvyys, luotettavuus, oikeudenmukaisuus ja osallistaminen, muodostavat viitekehyksen, jonka avulla tekoälyn riskejä voidaan jäsentää ja hallita eri käyttötapauksissa. Käytännön havainnot hyvinvointialueiden tekoälypiloteista osoittavat, että vakavimmat riskit liittyvät usein inhimillisiin ja organisatorisiin tekijöihin, kuten osaamiseen, luottamukseen, sääntelyn tulkintaan ja datan laatuun. Vaikka hallintakeinot pienentävät kokonaisriskejä merkittävästi, osa jäännösriskeistä jää väistämättä olemassa oleviksi ja vaatii jatkuvaa seurantaa.

EU:n sääntelykehys, erityisesti tekoälyasetus, tietosuojalainsäädäntö ja lääkinnällisiä laitteita koskevat asetukset luovat selkeät reunaehdot tekoälyn käytölle, mutta samalla se tuo

mukanaan tulkinnanvaraisuutta ja hallinnollista kuormaa. Tämä korostaa tarvetta kansalliselle ohjaukselle ja yhtenäisille toimintamalleille, jotta tekoälyn kehittäminen ja käyttöönotto eivät pysähdy varovaisuuteen, vaan etenevät hallitusti ja yhdenmukaisesti. Riskiperusteinen lähestymistapa on keskeinen keino sovittaa yhteen potilasturvallisuus, tietosuojat ja innovaatioiden edistäminen.

Johtopäätöksenä voidaan todeta, että tekoälyn vastuullinen hyödyntäminen sote-sektorilla edellyttää paitsi toimivia teknisiä ratkaisuja myös selkeitä hallintamalleja, osaamisen vahvistamista ja avointa vuoropuhelua eri sidosryhmien välillä. Riskienhallintatyökalu tarjoaa käytännöllisen välineen riskien tunnistamiseen ja arviointiin, mutta sen vaikuttavuus riippuu siitä, miten systemaattisesti ja aidosti sitä hyödynnetään osana päätöksentekoa. Pitkällä aikavälillä luottamukseen perustuva, läpinäkyvä ja oikeudenmukainen tekoälyn käyttö luo edellytykset sille, että tekoäly voi tukea sosiaali- ja terveydenhuollon kestävästä kehityksestä ja yhteiskunnallista hyvinvointia.

Liitteet

Liite 1. Huomiota vaikutusten arviointiin.

- **Oikeudellinen ja sääntelyyn liittyvä arviointi**
 - Onko kyseessä lääkinällinen laite tai muu korkean riskin tekoälysovellus
 - Mitä lainsäädäntöä ja viranomaisvaatimuksia sovellus edellyttää
 - Miten varaudutaan sääntelyn ja tulkintojen muutoksiin
- **Tietosuoja ja tietoturva**
 - Henkilötietojen anonymisointi tai pseudonymisointi
 - Potilas- ja asiakastietojen vuoto- ja väärinkäyttöriskit
 - Käyttötarkoituksen rajausta ja tiedon minimointi
- **Vinoumien ja syrjinnän hallinta**
 - Riskit ennakkoluulojen tai syrjivien käytäntöjen vahvistumiselle
 - Vaikutusten seuranta eri väestöryhmissä
 - Menetelmät vinoumien tunnistamiseen ja korjaamiseen
- **Yksilön rooli ja vastuu**
 - Tekoälyn käyttö päätöksenteon tukena, ei itsenäisenä päättäjänä
 - Mahdollisuus ohittaa, korjata tai hylätä tekoälyn ehdotukset
 - Selkeä vastuunjako, jos tekoäly tekee virheen
- **Kliininen laatu ja potilasturvallisuus**
 - Hyötyjen ja riskien arviointi diagnoosissa, hoitosuosituksissa tai resurssien kohdentamisessa
 - Virheiden mahdollinen vakavuus ja niiden hallinta
 - Käytön rajaaminen korkean riskin tilanteissa
- **Hyväksyttävyys ja luottamus**
 - Potilaiden ja henkilöstön luottamus tekoälyjärjestelmään
 - Ymmärrettävä informointi tekoälyn käytöstä
 - Vaikutukset hoidon laatuun ja ammattilaisten työhön
- **Organisatorinen käyttöönotto**
 - Tekoälyn roolin ja rajojen määrittely osana työnkulkuja
 - Koulutus ja vaiheittainen käyttöönotto
 - Jatkuva seuranta ja vaikutusten arviointi
- **Teknologinen luotettavuus**
 - Järjestelmän toimintavarmuus ja virhetilanteiden hallinta
 - Integraatio olemassa oleviin tietojärjestelmiin
 - Teknologiset riippuvuudet ja jatkuvuus

Liite 2. Tekoälyn turvallisen käytön arviointi.

Vastuullisuus					
TODENNÄKÖISYYS			VAIKUTUKSET		
1	Riskit epätasapuolisista ja epäeettisistä toiminnoista vähäisiä	Matala riski	1	Loppukäyttäjät saavat oikeudenmukaisen käsittelyn ja käsittelyn sääntöjen toimivuus on varmistettu, käyttäjien tasapuolinen kohtelu toteutuu hyvin	Pieni vaikutus
2	Riskit epätasapuolisista ja epäeettisistä toiminnoista ovat melko vähäisiä tai ne ovat paikallisia	Kohtalainen riski	2	Loppukäyttäjät saavat oikeudenmukaisen käsittelyn ja käsittelyn säännöt toimivat mutta pientä vinoumaa voi esiintyä tekoälyn käytössä, tasapuolisuutta ei voida täysin taata	Kohtalainen vaikutus
3	Riskit epätasapuolisista ja epäeettisistä toiminnoista voivat toteutua tai ne ovat paikallisesti laajemmin	Keskitasoinen riski	3	Loppukäyttäjät voivat saada epäoikeudenmukaiset käsittelyn ja käsittelyn säännöt eivät välttämättä aina toimi täysin ja vinoumaa voi esiintyä laajemmin tekoälyn käytössä, tasapuolisuutta ei voida taata	Keskisuuri vaikutus
4	Riskit epätasapuolisista ja epäeettisistä toiminnoista ovat melko todennäköisiä tai ne ovat melko laajoja	Melko suuri riski	4	Loppukäyttäjät voivat saada epäoikeudenmukaista käsittelyä ja käsittelyn säännöt eivät välttämättä aina toimi ja vinoumaa voi esiintyä laajemmin tekoälyn käytössä, epätasapuolisuutta esiintyy jokin verran	Melko suuri vaikutus
5	Riskit epätasapuolisista ja epäeettisistä toiminnoista ovat todennäköisiä tai ne ovat laajoja	Korkea riski	5	Loppukäyttäjät saavat todennäköisesti epäoikeudenmukaista käsittelyä ja käsittelyn säännöt eivät yleensä toimi ja vinoumaa esiintyy tekoälyn käytössä, epätasapuolisuutta esiintyy	Suuri vaikutus

Läpinäkyvyys					
TODENNÄKÖISYYS			VAIKUTUKSET		
1	Riskit tekoälyn toiminnan ymmärtämisessä ja osallistamisen toteutuksessa ovat vähäisiä.	Matala riski	1	Loppukäyttäjälle toimintaperiaatteet ja viestintä ovat selkeitä, koska ne kohdistuvat oikeaan käyttäjäryhmään.	Pieni vaikutus
2	Riskit tekoälyn toiminnan ymmärtämisessä on melko vähäistä. Läpinäkyvyys on osittaista ja osallistaminen hiukan rajattua.	Kohtalainen riski	2	Loppukäyttäjille toimintaperiaatteet ovat melko selkeitä, mutta niitä ei välttämättä pystytty viestimään laajasti tai ne eivät ole saavutaneet kaikkia osapuolia. Päätöksentekoon osallistuu mahdollisesti vain suppea joukko, mikä voi johtaa sokeisiin pisteisiin.	Kohtalainen vaikutus
3	Riskit tekoälyn toiminnan ymmärtämättömyydestä voivat toteutua ja osallistaminen voi epäonnistua. Eilätkinnällisten laitteiden auditointikäytäntö ja CE merkintä epäselvä korkeariskisissä tapauksissa.	Keskitasoinen riski	3	Loppukäyttäjille toiminnasta voidaan viestiä puutteellisesti, ja osallistujien määrä voi olla liian laaja. Tämä kasvattaa riskiä siitä, että päätöksenteossa ilmenee virheellisiä oletuksia tai väärinkäsityksiä. Tämä virheellisyys voi heijastua myös viestintään. Vaikutukset voivat ulottua organisaatioon tai ryhmään.	Keskisuuri vaikutus
4	Riskit tekoälyn toiminnan ymmärtämättömyydestä ovat melko todennäköisiä tai ne ovat melko laajoja. Osallistaminen melko todennäköisesti epäonnistuu.	Melko suuri riski	4	Loppukäyttäjille toimintaperiaatteet voivat olla monimutkaisia ja vaikeasti ymmärrettäviä, eikä niiden perusteista osata välttämättä viestiä kunnolla. Osallistaminen voi olla heikkoa, mikä voi lisätä päätöksenteon epävarmuutta ja käytön vastustamista	Melko suuri vaikutus
5	Riskit tekoälyn toiminnan ymmärtämättömyydestä ovat todennäköisiä tai ne ovat laajoja. Osallistaminen todennäköisesti epäonnistuu.	Korkea riski	5	Loppukäyttäjille tekoäly toimii läpinäkyvättömästi ("musta laatikko"), eikä sen päätöksentekoa voida selittää edes asiantuntijoille. Viestintä ja osallistaminen voi olla puutteellisia, mikä voi johtaa merkittävään epäluottamukseen ja päätöksentekorisien kasvuun, erityisesti kun vaikutukset ulottuvat laajasti yhteiskuntaan.	Suuri vaikutus

Oikeudenmukaisuus						
TODENNÄKÖISYYS				VAIKUTUKSET		
1	Riskit tekoälyn oikeudenmukaisuuden toteutumisessa ovat vähäisiä eikä syrjintää tapahdu ja toiminta on tasa-arvoista. Mahdolliset vaikutukset ovat marginaalisia eikä merkittäviä eriarvoistavia vaikutuksia odoteta.	Matala riski		1	Loppukäyttäjät kokee käytön turvallisiksi ja oikeudenmukaiseksi ja tasa-arvoiseksi. Loppukäyttäjälle ei muodostu palvelun laatuun liittyviä ongelmia.	Pieni vaikutus
2	Riskit tekoälyn oikeudenmukaisuuden toteutumisessa ja syrjinnässä ovat melko vähäisiä. Toiminta on melko tasa-arvoista. Oikeudenmukaisuus voi osittain toteutua, mutta joitain lieviä eriarvoistavia vaikutuksia saattaa ilmetä. Toiminta on käytettävissä. Tekoälysovellusten ja palveluiden käyttöohjeet ovat epäselvät ja harhaanjohtavat.	Kohtalainen riski		2	Loppukäyttäjille toimintaperiaatteet ovat pääosin selkeitä, mutta tekoälyn käyttö voi herättää epäilyksiä erityisesti yhdenvertaisuuden ja oikeudenmukaisen kohtelun osalta. Vaikka palvelun laadussa esiintyy lieviä vaihteluita, käyttäjät kokevat sen edelleen turvallisiksi käyttää.	Kohtalainen vaikutus
3	Riskit tekoälyn oikeudenmukaisuuden toteutumisessa voivat realisoitua ja syrjintää voi tapahtua ja tasa-arvo toteutuu siedettävästi. Joitain merkittäviä eriarvoistavia vaikutuksia saattaa esiintyä, mikä voi vaikuttaa osapuolten kokemaan oikeudenmukaisuuteen. Toiminta voi kokonaisuudessa olla silti siedettävää.	Keskitasoinen riski	*	3	Loppukäyttäjille toimintaperiaatteet voivat olla epäselviä, ja tekoälyn käyttö herättää kasvavaa epäilyä erityisesti yhdenvertaisuuden ja oikeudenmukaisuuden suhteen. Palvelun laatu voi vaihdella enemmän, mutta sitä pidetään yhä pääosin turvallisena käyttää.	Keskisuuri vaikutus
4	Riskit tekoälyn oikeudenmukaisuuden toteutumatta jäämisestä ovat melko todennäköisiä tai vaikutukset voivat olla melko laajoja. Syrjintä voi olla tahallista tai melko laajaa ja tasa-arvo toteutuu vain satunnaisesti. Eriarvoistavat vaikutukset voivat olla huomattavia. Säädot voidaan tehdä hyödyttämään vain osaa käyttäjistä.	Melko suuri riski		4	Loppukäyttäjille toimintaperiaatteet ovat melko epäselviä, ja tekoälyn käyttö epäilyttää erityisesti yhdenvertaisuuden ja oikeudenmukaisuuden suhteen. Palvelun laatu vaihtelee, ja sitä pidetään melko turvattomana käyttää.	Melko suuri vaikutus
5	Riskit tekoälyn oikeudenmukaisuuden toteutumatta jäämisestä ovat erittäin todennäköisiä, ja vaikutukset ovat laajoja. Syrjintä on haluttua tai laajaa, eikä tasa-arvo toteudu käytännössä juuri lainkaan. Eriarvoistavat vaikutukset ovat merkittäviä ja voivat vakavasti heikentää yhteiskunnallista oikeudenmukaisuutta.	Korkea riski		5	Loppukäyttäjille toimintaperiaatteet ovat epäselvät, ja tekoälyn käyttö koetaan haastavaksi erityisesti yhdenvertaisuuden ja oikeudenmukaisuuden näkökulmasta. Palvelun laatu on vaihteleva, ja käyttäjät kokevat sen turvattomaksi. Tämä voi heikentää merkittävästi käyttäjien turvallisuuden tunnetta.	Suuri vaikutus

Osallistaminen						
TODENNÄKÖISYYS				VAIKUTUKSET		
1	Riskit osallistamisen toteutumisessa ovat vähäisiä, ja sidosryhmät pääsevät vaikuttamaan aidosti päätöksentekoon. Osallistaminen on monipuolista ja kattavaa, ja eri käyttäjäryhmät tulevat huomioituiksi. Mahdolliset negatiiviset vaikutukset ovat marginaalisia. Vaikuttaminen on todellista, ei vain teknistä.	Matala riski		1	Loppukäyttäjä kokee osallistumismahdollisuudet riittäviksi, ja hänen näkemyksensä huomioidaan päätöksenteossa. Osallistaminen parantaa palvelun laatua, eikä merkittäviä ongelmia ilmene. Lisäksi loppukäyttäjä kokee voivansa vaikuttaa päätöksentekoon ja olevansa osa prosessia.	Pieni vaikutus
2	Riskit osallistamisen toteutumisessa ovat melko vähäisiä. Suurin osa sidosryhmistä voi vaikuttaa päätöksentekoon, mutta joitakin puutteita voi esiintyä. Osallistaminen on pääosin kattavaa, mutta joillakin ryhmillä voi olla vaikeuksia tuoda näkemyksiään esiin. Ohjeistus ja viestintä voivat olla osin puutteellisia. Vaikuttaminen osallistamisessa on edelleen todellista.	Kohtalainen riski		2	Loppukäyttäjille osallistumismahdollisuudet ovat pääosin selkeitä, mutta joitakin epäilyksiä voi herätä erityisesti vaikutusmahdollisuuksien todellisesta merkityksestä. Vaikka osallistamisessa voi olla lieviä puutteita, käyttäjät kokevat prosessin silti toimivaksi.	Kohtalainen vaikutus
3	Riskit osallistamisen toteutumisessa voivat realisoitua, ja osa sidosryhmistä voi jäädä prosessien ulkopuolelle. Osallistaminen voi olla muodollista tai rajoitettua, mikä heikentää vaikutusmahdollisuuksia. Joitain merkittäviä osallistamisen puutteita voi esiintyä, mutta järjestelmä on silti käytettävissä.	Keskitasoinen riski		3	Loppukäyttäjät kokevat osallistumismahdollisuudet osittain epäselviksi, ja huoli heidän vaikutusmahdollisuuksistaan kasvaa. Palvelun kehityksessä voi olla vaihtelevuutta, mutta käyttäjät kokevat sen yhä osittain toimivaksi. Prosessin toiminta voi arveluttaa.	Keskisuuri vaikutus
4	Riskit osallistamisen toteutumisessa ovat melko todennäköisiä tai vaikutukset voivat olla melko laajoja. Osallistaminen on rajattua ja vain tietyillä ryhmillä on merkittävä vaikutus. Käyttäjien näkemykset jäävät suurelta osin huomiotta, mikä voi heikentää luottamusta ja hyväksyntää. Prosessit voivat olla läpinäkyvättömiä tai vaikeasti saavutettavia. Sidosryhmät voivat kokea ettei heihin luoteta eivätkä he ole hyväksyttyjä.	Melko suuri riski		4	Loppukäyttäjät kokevat osallistumismahdollisuudet melko epäselviksi, ja he epäilevät niiden todellista vaikutusta päätöksentekoon. Käyttäjien kokemus palvelun laadusta heikkenee, ja järjestelmän luotettavuus saatetaan kyseenalaistaa.	Melko suuri vaikutus
5	Riskit osallistamisen toteutumatta jäämisestä ovat erittäin todennäköisiä, ja vaikutukset ovat laajoja. Osallistaminen on käytännössä olematonta, ja päätöksenteko tapahtuu ilman käyttäjien tai sidosryhmien panosta. Tämä voi johtaa järjestelmän epäluottamukseen, vastustukseen ja käytön heikkenemiseen. Vaikka osallistaminen onnistuisi teknisesti, niin käyttäjä kokee jäävänsä kuulematta ja hyväksymättä.	Korkea riski		5	Loppukäyttäjät eivät koe osallistumismahdollisuuksia riittäviksi tai ne puuttuvat kokonaan. Käyttäjät eivät tunne voivansa vaikuttaa palvelun kehitykseen, mikä johtaa luottamuksen katoamiseen ja palvelun koetaan olevan epäkäytettävä. Tämä voi merkittävästi heikentää käyttäjien sitoutumista ja hyväksyntää. Käyttäjät tuntevat olevansa vain dataa prosessissa.	Suuri vaikutus

Luotettavuus (tietoturva ja tietosuoja)						
TODENNÄKÖISYYS				VAIKUTUKSET		
1	Riskit tietoturvan ja/tai tietosuojan toteutumisessa ovat vähäisiä, eikä merkittäviä haavoittuvuuksia esiinny. Järjestelmä suojaa tiedot tehokkaasti, eikä tietovuotoja odoteta ja riskit ovat vähäisiä. Mahdolliset vaikutukset ovat marginaalisia.	Matala riski		1	Loppukäyttäjä kokee järjestelmän turvalliseksi ja tietosuojan hyvin toteutetuksi. Käyttäjän tiedot ovat suojattuja, eikä tietoturvaan liittyviä ongelmia esiinny.	Pieni vaikutus
2	Riskit tietoturvan ja/tai tietosuojan toteutumisessa ovat melko vähäisiä. Järjestelmä on pääosin turvallinen, mutta joitakin lieviä haavoittuvuuksia voi esiintyä. Käyttäjien tiedot ovat kohtalaisen hyvin suojattuja, mutta satunnaiset tietovuodot tai väärinkäytökset ovat mahdollisia mutta kohtalaisen matalalla riskillä. Käyttöohjeet voivat olla paikoin epäselviä tai puutteellisia, mutta eivät ole haitaksi toiminnalle.	Kohtalainen riski		2	Loppukäyttäjille järjestelmän tietoturva- ja tietosuojaoperaahteet ovat pääosin selkeitä, mutta joitakin huolenaiheita voi herätä erityisesti tietojen käsittelyn ja suojauksen osalta. Käyttäjät pitävät palvelua kuitenkin edelleen turvallisena käyttöä.	Kohtalainen vaikutus
3	Riskit tietoturvan ja tietosuojan toteutumisessa voivat realisoitua ilman hyvää ylläpitoa. Järjestelmässä voi olla joitakin haavoittuvuuksia, jotka voivat johtaa tietovuotoihin tai luvattomaan pääsyyn tietoihin ilman hyvää ylläpitoa. Käyttäjät voivat kokea epävarmuutta tietoturvan ja tietosuojan riittäväyydestä, mutta kokonaisuus on vielä siedettävä ja käytettävissä. Järjestelmät on auditoitu tai tarkastettu.	Keskitasoinen riski		3	Loppukäyttäjille järjestelmän tietoturva- ja tietosuojaoperaahteet voivat olla arveluttavia, ja epäilykset tietojen turvallisuudesta lisääntyvät. Käyttäjät saattavat kokea ajoittaisia tietoturvaongelmia, mutta palvelu nähdään vielä pääosin turvallisena käyttöä.	Keskisuuri vaikutus
4	Riskit tietoturvan ja tietosuojan heikkenemisestä ovat melko todennäköisiä tai vaikutukset voivat olla paikallisia. Tietovuodot tai väärinkäytökset voivat olla yleisiä, ja järjestelmä voi olla altis hyökkäyksille. Käyttäjien tiedot eivät ole riittävästi suojattuja, ja järjestelmän tietoturvatavoimia ei välttämättä ole riittävästi auditoitu. Tietoturva ja tietosuoja maturiteetti ei ole tiedossa kunnolla. Ei voida varmaksi sanoa tuleeko poikkeamia. Suojaus voi silti kestää, mutta riski on suurempi.	Melko suuri riski		4	Loppukäyttäjille järjestelmän tietoturva- ja tietosuojaoperaahteet ovat melko epäselviä, ja tietoturvaan liittyvät huolet kasvavat. Palvelun tietoturvaongelmat voivat näkyä, ja käyttäjät voivat kokea sen turvattomaksi käyttöä.	Melko suuri vaikutus
5	Riskit tietoturvan ja tietosuojan toteutumatta jäämisestä ovat todennäköisiä, ja vaikutukset voivat olla laajoja. Järjestelmään voi tulla vakavia haavoittuvuuksia, ja tietomurrot tai väärinkäytökset ovat mahdollisia. Käyttäjien tiedot voivat olla vaarassa, ja niiden suojaaminen on puutteellista. Tämä voi vakavasti heikentää luottamusta järjestelmään ja sen käyttökelpoisuutta. Voi olla auditoitu, mutta auditoinnissa on puutteita.	Korkea riski		5	Loppukäyttäjille järjestelmän tietoturva- ja tietosuojaoperaahteet ovat epäselviä, ja palvelu koetaan haavoittuvaksi ja turvattomaksi. Käyttäjät eivät luota järjestelmään, ja tietoturvaongelmat voivat vakavasti heikentää sen käyttökelpoisuutta.	Suuri vaikutus