

# Laiteturvallisuus- ja DigiNet -verkostojen webinaari

## Sote-tekoälyn ekosysteemin riskienhallintaryhmä

Andrei Laurén, STM  
Mika Pihlajamäki, THL  
13.02.2025

# Riskienhallintaryhmän tausta ja tavoitteet edellytyksiä luovassa toiminnassa



- Vapaaehtoinen riskienhallintatyöryhmä sijoittuu STM:n tekoälyn ekosysteemin koordinaatioryhmän alatyöryhmäksi
- Työryhmän työskentely on alkanut 11/2024 ja työ jatkuu ainakin 31.12.2025 asti
- Riskienhallintaryhmän tehtävänä on tukea tekoälyn turvallisen käytön kehittämistä ja edistää kokeiluja
- Tekoälyn turvallisesta käytöstä ja kokeiluista saadut riskihavainnot ja muut turvalliseen käyttöön liittyvät kokemukset kerätään keskitetysti ryhmän ja sote- toimijoiden käyttöön
- *Ryhmässä ei arvioida toisen organisaation projekteja vaan opittuja asioita ja havaintoja jaetaan yhteiseen hyvään*

# Riskienhallintaryhmän tarkennetut tavoitteet



## Vaikuttavuuteen keskittyminen

- Periaate: keskitytään mihin voidaan vaikuttaa
- Keskitytään korkeariskisissä järjestelmissä riskeihin ja esteisiin, koska niihin on akuutein tarve
- Pyritään löytämään balanssi riskien ja potentiaalisten hyötyjen välillä, siten että mahdollistetaan käyttö (täysi säätely vs. ei mitään säätelyä)
- 20/80-ajattelua eli 20% oikealla tekemisellä saadaan hallittua 80% riskeistä

## Tuotokset

- Tuotoksia jaetaan sote-tekoälyn ekosysteemissä kaikille
- Missä käytöissä voidaan edetä nopeasti, missä pitää edetä varovaisemmin
- Periaatteiden selkeyttäminen, onko kyseessä automaattinen päätöksentekijä vai ei (mielipiteet eroavat nyt kentällä, esim. lääkinnälliset laitteet)
- AI vaikutusten arvioinnin tueksi osana DPIA:a ja AIA:a (AI Assessment), sisältäen suositukset riskien ja esteiden vähentämiseksi tai poistamiseksi

# Tekoälyn turvallisen käytön tähän asti tunnistetut periaatteet



Esimerkkejä tekoälyn käytön turvallisuudesta (edellytyksiä luova tekoälyn käyttö riskien näkökulmasta)



## Vastuullisuus

Vaikutuksen arviointi:  
Haitallisten vaikutusten valvonta  
Sopivuuden arviointi (tulee sopia tarkoitukseen)  
Lähdetietojen hallinta  
Toteutuksen valvonta (Ihmisiä ymmärtävä ja kunnioittava)  
Tekoälyn rajoitusten ymmärtäminen



## Läpinäkyvyys

Mahdollistaa tarkastelun (avoimuus):  
Tunnistetaan ja ymmärretään miten järjestelmätasolla AI toimii, esim. päätösten tukena  
Valvonta ja kyvykkyys, kuten järjestelmän tekemien päätösten arvioiteja avoimesti  
Sidosryhmät ja vuorovaikutus ja viestintä  
Sosiaalinen kestävyys



## Oikeudenmukaisuus

Toimii tasa-arvoisesti:  
Palvelun laatu (AI:n käyttäjän tunne)  
Resurssien ja mahdollisuuksien kohdentaminen  
Stereotyyppien minimointi ja perusoikeudet sekä eettisyys  
Osallistava johtaminen



## Luotettavuus, tietoturva ja tietosuojaja

Luotettavuus- ja turvallisuusohjeet sekä riskienhallinta:  
Yksityisyyden ja tietosuojan noudattaminen  
Turvallisuuden noudattaminen ja seuranta sekä standardit  
Havaintojen seuranta ja korjausprosessit  
Tekninen läpinäkyvyys



## Osallistaminen

Saavutettavuusvaatimusten noudattaminen:  
AI-kulttuurin ja toiminnan edistäminen sekä ymmärtäminen



# Esimerkkejä riskeistä:



## Riskinä kielimallien vinoumat

Vastausten pohjana tulisi käyttää huolellisesti kuratoituja dokumentteja eikä pelkästään yleisen kielimallin tietämystä. On tärkeää tunnistaa, että kaikki data on lähtökohtaisesti jossain määrin vinoutunutta, mikä vaikuttaa myös mallien tuottamiin lopputuloksiin.

Synteettinen data voi auttaa korjaamaan vinoumia, jos ne on ensin tunnistettu. Esimerkiksi, jos datassa esiintyy sukupuolivinoumaa – kuten miesten yliedustus – synteettisellä datalla voidaan tasapainottaa aineistoa ja parantaa mallin objektiivisuutta. Tämä lähestymistapa tukee oikeudenmukaisempia ja tarkempia päätöksiä erityisesti sosiaali- ja terveysalalla, jossa vinoumien vaikutukset voivat olla merkittäviä.

## Riskinä etiikka tekoälystä saatavia tietoja käytetään väärin tai sitä käytetään väriin asioihin

Tekoälyn käyttö sosiaali- ja terveysalalla edellyttää selkeitä ohjeistuksia ja oikeuksia sekä jatkuvaa eettistä arviointia. Tätä varten olisi hyödyllistä perustaa tekoälyn eettistä toimintaa seuraava ryhmä, joka voisi ottaa mallia tutkimuspuolen eettisistä toimikunnista ja mittausmenetelmistä. Sairaaloilla on ollut eettisiä toimikuntia, joten tästä kokemuksesta voidaan ammentaa myös tekoälyn soveltamiseen.

Tekoälyn eettinen arviointi (AI Assessment), joka määrittelee, mikä on sallittua ja millaisia käytäntöjä tulisi noudattaa. Keskeinen tavoite on vinoumien poistaminen ja ihmisten tasa-arvoinen kohtelu. Eettisellä arvionilla voidaan varmistaa paremmin, että tekoäly tukee yhdenvertaisia, turvallisia ja eettisesti kestäviä ratkaisuja sosiaali- ja terveydenhuollossa.

# Riskienhallintaryhmän jäsenet



Riskienhallintaryhmän jäsenet löytyy ao. linkin kautta

[Riskienhallintaryhmä – DigiFinland](#)

Lisätietoa:

[andrei.lauren@gov.fi](mailto:andrei.lauren@gov.fi)

[mika.pihlajamaki@thl.fi](mailto:mika.pihlajamaki@thl.fi)