

ASIANUMERO	STM113:00/2022	
AIKA	maanantai 16.1.2023 klo 10.00–12.00	
PAIKKA	Teams-kokous	
Kutsuttu		
Jäsen	Organisaatio	lisätietoa
Joni Komulainen	STM	puheenjohtaja
Hanna Tervonen	Findata	sihteeri
Jarmo Harju	FIMM	
Meri-Tuulia Kaarakainen	Turun yliopisto	x
Kimmo Kaski	Aalto yliopisto	x
Kristiina Lappalainen	Traficom	x
Jussi Leinonen	Bayer	x
Päivi Lindström	Findata	
Tomi Mäkelä	Helsingin yliopisto	x
Tuomo Pentikäinen	Veil.ai	x
Jarkko Reittu	THL	x
Tom Southerington	FinBB	x
Hannele Svanström	Aluehallintovirastot	
Kirsi Talonen	Helsingin yliopisto	x
Janika Tarkoma	Tilastokeskus	
Arho Virkki	VSSHP	x
Pysyvät asiantuntijat		
Toni Ahvas	Helsingin yliopisto	x
Antti Honkela	Helsingin yliopisto	x
Andrei Lauren	STM	x

1.	Kokouksen avaus Puheenjohtaja avasi kokouksen klo. 10.00.
2.	Edellisen kokouksen pöytäkirja ja yhteystietojen tarkistus Hyväksyttiin edellisen kokouksen pöytäkirja. Sovittiin, että kokousmateriaalit tulevat saataville THL:n Yhteistyötiloissa olevalle sivulle. Hanna Tervonen toimittaa seuraavaan kokoukseen mennessä ohjeet siitä, kuinka THL:n ulkopuoliset käyttäjät pääsevät Yhteistyötiloihin.

	<p>Käytiin läpi ryhmän tämänhetkinen yhteystietolista. Jos yhteystietoihin on tullut muutoksia, asiasta voi informoida Hanna Tervosta. Kokouksessa ilmoitetut muutokset on jo huomioitu.</p>
3.	<p>Linjauspyyntö kuva- ja signaalitiedon anonymisoinnista</p> <p>Jatkettiin edellisessä kokouksessa käsittelyyn otetun linjauspyynnön käsittelyä. Hanna Tervonen selvensi edellisessä kokouksessa esille tulleita pyyntöön liittyviä kysymyksiä.</p> <p>Keskustelussa tuotiin esille, että yleensä pelkällä signaalitiedolla ei voi tehdä koneoppimismallin treenausta, vaan moderniin koneoppimiseen tarvitaan myös vastemuuttajat. Koneoppimismallien opetuksessa usein käytettävät lääkärin lausunnot ovat henkilötietoa. Keskusteltiin siitä, voiko sydänkäyräjoukosta tunnistaa yksittäisen henkilön. Riski tunnistamiselle voi tulla siitä, että tätä tietoa yhdistellään johonkin muuhun tietoon. Todettiin, että GDPR ei kuitenkaan lähde teoreettisesta tunnistettavuudesta, vaan siitä, onko käsittelijällä kohtuudella käytettävissään laillisia keinoja tunnistaa yksilöitä.</p> <p>Keskusteltiin siitä, millaisissa tilanteissa kuvantamisdataa käyttävän koneoppimismallin treenaus on tutkimustoimintaa ja milloin tietojohdamista tai mahdollisesti innovaatio- ja kehittämistoimintaa. Jos tällaista toimintaa tehdään tietojohdamisen alla, ei vaatimusta auditoidulle käyttöympäristölle ole, vaikka tällaisenkin käytön on tapahduttava tietoturvaselästi. Tutkijakentällä koetaan tietoturvaselisen käyttöympäristön vaatimus hankalaksi toteuttaa. Ryhmän näkemys oli, että riittävä syy tietoturvan vähentämiseen ei ole se, että joku asia voidaan tehdä helpommin tietoturvaselisen käyttöympäristön ulkopuolella. Tietoturvaselisen käyttöympäristön vaatimus voi kuitenkin tehdä tietynlaiset kansainväliset tutkimukset lähes mahdottomiksi, samoin tietynlaisen innovaatio- ja kehittämistoiminnan. Eurooppalainen suuntaus on menossa samaan suuntaan kuin Suomen tilanne eli toisiokäyttö on mahdollista vain tietoturvaselissä käyttöympäristöissä. Linjauspyynnön käsittelyä jatketaan alatyöryhmissä. Linjauspyynnön kysyjille tarjotaan mahdollisuutta osallistua keskusteluun seuraavassa kokouksessa.</p>
4.	<p>Findatan tilannekatsaus</p> <p>Hanna Tervonen kertoi esityksessään kahdesta Findatalle selvitykseen tulleesta kysymyksestä:</p> <ol style="list-style-type: none"> 1. Koneoppimismallin anonymiteetti <ul style="list-style-type: none"> - Findatalle on tullut käsittelyyn koneoppimismallin anonymiteetin varmistus. - Todettiin, että luonnollista kieltä käyttävät mallit voivat oppia asioita, joita on vaikea huomata /arvioida. Koneoppimismallien riskialttiutta paljastumiselle on selvitetty. Mallit, jotka käyttävät potilaskertomustekstejä, ovat korkean riskin malleja. Potilaskertomustekstit ovat yksilöivää tietoa, vaikka suorat tunnisteet poistetaan opetusmateriaalina käytettävästä aineistosta. - Findatan tulisi kehittää prosessi, jolla arvioida koneoppimismallien anonymiteettia. Arviointi täytyy kuitenkin todennäköisesti tehdä tapauskohtaisesti. Tilastollisen tietosuojan alatyöryhmä voi jatkaa tämän asian käsittelyä. - Ryhmä ei pystynyt annetun pohjatiedon perusteella arvioimaan esimerkkitapauksena olleen koneoppimismallin anonymiteettia tarkemmin,

	<p>mutta se arvioitiin korkean riskin malliksi. Ryhmässä ehdotettiin, että malli voitaisiin luovuttaa käyttöön suljettuun käyttöympäristöön, jolloin ympäristö olisi lisäsuojatoimi, jos perusteet tähän riittävät.</p> <p>2. Hajautettu laskenta ja tietoturvallisten käyttöympäristön tietoturva</p> <p>Findata toi keskusteluun kysymyksen siitä, näkeekö ryhmä Swarm learning-parametritietojen vaihdon vaarantavan käyttöympäristön tietoturvaa, huomioiden että arviointilaitoksen tulee auditoida kyseinen toteutus ennen käyttöönottoa. Ryhmässä keskusteltiin seuraavaa:</p> <ul style="list-style-type: none"> - Mietittiin malliparametrien anonymiteettia ja tähän liittyviä paljastumisriskejä, koska parametrejä estimoidaan iteratiivisesti. - Lohkoketjuteknologia koettiin riskialttiiksi ja mietittiin, voisiko tämän sijasta käyttää muuta teknologiaa. Lohkoketjuteknologiaan on selvityksessä olevassa tapauksessa päädytty, koska on haluttu täydellinen loki. Käytössä oleva lohkoketju ei ole julkinen. - Mietittiin mahdollisuuksia sille, että tietoturvallisia käyttöympäristöjä laitetaan verkostoksi. Mahdollistaako nykyinen lainsäädäntö tämän? - Tietoturvan näkökulmasta riskiä ei pidetty suurena, jos toteutus tapahtuu teknisesti yleisesti tunnetuilla käytännöillä. Lohkoketjutoteutuksen katsottiin vaativan kunnollisen riskianalyysin.
6.	<p>EHDS (ja TEHDAS) tilannekatsaus</p> <p>Joni Komulainen kertoi EHDS-neuvotteluiden tilanteesta.</p>
7.	<p>Alatyöryhmien tilanne</p> <ul style="list-style-type: none"> - Tilastollinen tietosuoja Tämä ryhmä alkaa työstämään koneoppimiskysymyksen liittyviä asioita. Hanna Tervonen kysyy ryhmälle vetäjää. - Juridinen tietosuoja Ryhmä ei ole kokoontunut. - Organisatorinen tietoturva Andrei Lauren ja Jarkko Reittu organisoivat tämän ryhmän toiminnan. - Tekninen tietoturva (kyberturvallisuus) Ryhmä ei ole kokoontunut.
8.	<p>Ilmoitusasiat</p> <p>Janika Tarkoma ei osallistu asiantuntijaryhmän toimintaan seuraavan kalenterivuoden aikana. Sijaisena toimii Annu Cabrera (Tilastokeskus).</p>
9.	<p>Seuraavat kokoukset</p> <p>Kokouskutsut seuraaviin kokouksiin on jo lähetetty. Seuraava kokous on 27.2.</p>
10.	<p>Muut asiat</p> <p>Keskusteltiin mahdollisuudesta pitää joku seuraavista kokouksista lähitapaamisena.</p>

11.	Kokouksen päätös Puheenjohtaja päätti kokouksen klo. 11.35.