



Tutkimuksen digitalisaation politiikkatoimien kokoamisen ja kansallisen datanhallinnan ja laskennan tutkimusinfrastruktuuriympäristön päivittämisen (DL2026) seurantaryhmä

Seurantaryhmän 3. kokous

Aika: perjantai 30.8.2024 klo 12.30–14.00

Paikka: opetus- ja kulttuuriministeriö, kh. Labor, Meritullinkatu 10

Osallistujat:

johtaja Erja Heikkinen, opetus- ja kulttuuriministeriö, puheenjohtaja
tutkimusjohtaja Annukka Pakarinen, ammattikorkeakoulujen rehtorineuvosto Arene ry (Hämeen ammattikorkeakoulu)
~~yksikön johtaja Niko-Matti Ronkonmäki, liikenne- ja viestintäministeriö (virkavapaa)~~
erityisasiantuntija Tuire Valkonen, liikenne- ja viestintäministeriö
johtava tiedeasiantuntija Anu Nuutinen, Suomen Akatemia
tutkimuspalveluiden johtaja Ella Bingham, Suomen yliopistojen rehtorineuvosto Unifi ry (Aalto-yliopisto)
professori Minna Palmroth, Suomen Tiedeakatemit (Helsingin yliopisto)
johtaja Pekka Lehtovuori, Tieteen tietotekniikan keskus CSC
yksikönjohtaja Saku Anttila, Tutkimuslaitosten yhteenliittymä Tulanet (Suomen ympäristökeskus)
opetusneuvos Petteri Kauppinen, opetus- ja kulttuuriministeriö
opetusneuvos Laura Taajamaa, sihteeri

Pysyvät asiantuntijat
professori Heikki Mannila, Aalto-yliopisto
CSC:n hallituksen puheenjohtaja Jukka Mönkkönen, Itä-Suomen yliopisto

1. Kokouksen avaus ja asialistan hyväksyminen

Puheenjohtaja avasi kokouksen kello 12.30, asialista hyväksyttiin.

2. Edellisen kokouksen muistion hyväksyminen

Hyväksyttiin 27.5.2024 kokouksen muistio.

3. Käynnissä olevat politiikkatoimet tutkimuksen digitalisaation edistämiseksi

Taustoittaakseen ryhmän keskusteluja puheenjohtaja kertoi tiiviisti käynnissä olevista digitalisaation edistämiseen liittyvistä politiikkatoimista. Tällaisia ovat esimerkiksi: hallituksen tekoälyiltakoulu; eurooppalaiset ja kansalliset panokset datanhallinnan sekä laskennan infrastruktuuriin; digitoimisto, digikompassi ja digisalkku; tutkimuksen lippulaivat, mm. FCAI, kvantti, 6G; tohtorikoulutuksen pilotit: mm. Finnish Doctoral Program Network in Artificial Intelligence, Quantum sekä Doctoral pilot in software engineering, Intelligent Work Machines, MICroELectronics doctoral school pilot; eurooppalaisen tekoälytutkimuksen Ellis-verkosto ja -instituutti; nk. toisilain korjaaminen (STM); jatkuvan oppimisen digitaalinen palvelukokonaisuus ja

digivisio osaamisen kehittämisen ja löytämisen alustoina sekä LUMA(TE)-strategian ja –toimenpidesuunnitelman toteuttaminen.

Ajankohta ryhmän työlle on otollinen, käynnistymässä on korkeakoulutuksen ja tutkimuksen visiotyö, johon voidaan nivoa varautuminen tarvittaviin muutoksiin. Ministeriössä valmistaudutaan myös syksyn yliopistojen sopimusneuvotteluihin. Ryhmässä todettiin, että perustana työssä on dialogi ja toimijoiden työnjako.

4. Eurooppalaisen kehityksen ajankohtaiset: EuroHPC

Petteri Kauppinen kertoi EuroHPC-yhteistyön ajankohtaisista kuulumisista. EuroHPC-asetukseen lisättiin toukokuussa 2024 uusi tekoälykokonaisuus (ns. tekoälytehdas) eli mahdollisuus uusien tekoälypaineistettujen supertietokoneiden hankintaan, nykyisten päivittämiseen sekä laskentapalveluiden tarjoamiseen ja osaamisen kehittämiseen. LUMIn kilpailukykyinen käyttöaika päättyvässä vuonna 2027 ja Suomi on varautunut uuden EuroHPC-supertietokoneen hankintaan enintään 250 M€:n kansallisella rahoituksella. LUMI-konsortion mailta on tiedusteltu halukkuutta osallistua rahoitukseen. Tekoälytehtaiden isännöintihaku avautuu syyskuussa ja hakemusten ensimmäinen määräaika on 4.11.2024 (ajankohta tarkentunut kokouksen jälkeen). Esityksessä käytiin läpi myös tilastoja Suomen LUMI-hankkeista sekä tarkasteltiin esimerkkejä LUMIn hyödyistä huippututkimuksen mahdollistamista laajemmin, kuten datakeskusekosysteemin edistäminen. Keskustelussa nostettiin esiin, että EuroHPC on korkealla komission politiikkatavoitteissa, pohdittiin datan käytettävyyttä sekä yhteyttä isoihin data-arkistoihin, tarkennettiin tekoälytehtaan määritelmää sekä todettiin, että CSC vastaa hakemuksen valmistelusta, ja että aikataulu on erittäin tiukka.

5. Katsaus tekoälyohjeisiin

Keskustelua alustivat Ella Bingham ja Anu Nuutinen. Eettisen tekoälymallin kysymys on toistaiseksi vielä ratkaisematon, ja haasteista on tutkijoiden hyvä olla tietoisia. Suosituksia tekoälyn vastuullisesta käytöstä tutkimuksen tukena on valmisteltu paljon, mutta Binghamin alustuksessa nostettiin erityisesti esiin ERA-foorumin ohjeet, jotka pohjaavat ALLEAN työhön. Suomessa TENKin ohjeistuksen pohjana on myös ALLEAN suositukset. Ohjeissa todetaan mm. että tutkimuksen arviointia ei pidä tehdä tekoälyllä ja että käytössä pitää huomioida yksityisyydensuoja, luottamuksellisuus sekä aineettomien oikeuksien suoja. Tutkimusorganisaatioilta toivotaan työkalujen vastuullisen käytön edistämistä sekä kehityksen seurantaa. Rahoittajia kehoitetaan tukemaan hakijoita tekoälyn läpinäkyvässä käytössä. Toisena kansainvälisenä esimerkkinä oli KU Leuvenin laadukkaat, konkreettiset ja jatkuvasti ylläpidetyt ohjeet.

Aallossa on opetuksen ja oppimisen ohjeiden ohella laadittu ohjeistus tutkijoille, ja työhön ovat osallistuneet lippulaiva FCAIn tutkijat ja juristit. Ohjeet on hyväksytty yliopiston ylimmissä päättävissä elimissä ja ne koskevat konkreettista tutkimusprosessia. Valmistelussa todettiin, että vastuullisen tutkimuksen laavat periaatteet sekä yleiset tietosuojan ohjeet toimivat erinomaisena ohjenuorana tässäkin. Tekoälyä käytettäessä on tärkeää suojata tutkijan omat immateriaalioikeudet, esimerkiksi tutkimussuunnitelmia ei laiteta harkitsemattomasti eri työkaluihin. Tekoäly ei voi olla tekijänä, vaan se on työkalu ja tutkijalla itsellään säilyy vastuu kirjoituksista. Muiden tekijyyttä on kunnioitettava ja viitattava relevantein osin edeltäviin tutkimuksiin. Tutkijan tulisi myös tiedostaa

mihin lähteisiin tekoälyjärjestelmä nojautuu. Tutkijan tulee kuvata avoimesti, miten tekoälyä on hyödynnetty tutkimusprosessissa. Yleistä tietosuoja-asetusta (GDPR) on myös noudatettava, eli henkilötietoja sisältäviä tutkimusaineistoja tai henkilötietoja ei tule syöttää tekoälytyökaluihin. Julkaisemattomien tulosten suojaamiseksi niitä ei tule ladata verkkoon, varsinkaan jos ne ovat arkaluonteisia. Aallossa on rakennettu GPT-tyyppinen työkalu, jolla voi talon sisällä käsitellä luottamuksellisia tietoja. Lopuksi Bingham kertoi mielenkiintoisesta Britannian Royal Societyn laatimasta raportista ja sen suosituksista tutkimuksen mahdollistajille. Raportissa pohditaan miten tekoäly tulee muuttamaan tiedettä ja tutkimusta.

Anu Nuutinen toi omassa katsauksessaan esiin tutkimusrahoittajan näkökulmaa tekoälyn käyttöön. Suomen Akatemia on selvittänyt tekoälyn käyttöä Akatemian ja verrokkioorganisaatioiden toiminnassa. Julkista sektoria ohjaava lainsäädäntö Suomessa kiteytyy siihen, että tekoäly ei korvaa viranomaisharkintaa, asiantuntemusta eikä hyvän hallinnon periaatteita eikä muuta asiakirjajulkisuutta tai virkavastuuta tuotettujen selvitysten ja asiakirjojen sisällöstä. Selvityksen mukaan on todennäköistä, että julkishallinnossa tulee arvioitaviksi uusia eettisiä ongelmia, jotka voivat liittyä niin työntekijöiden kuin kansalaistenkin oikeuksiin ja velvoitteisiin. Nuutinen totesi, että yliopistojen ohjeistus nivoutuu osaksi hyvän tieteellisen käytännön ohjeistusta ja että eurooppalainen tutkimusetiikkaohjeistus korostaa tieteenalan käytäntöjen noudattamista ja läpinäkyvyyden periaatetta. TENKin HTK-ohje 2023 mainitsee tekoälyn työympäristöön vaikuttavana tekijänä mutta ei ota kantaa sen käyttöön. Rahoittajien osalta selvityksessä todettiin, että hakemusten arvioinnissa tekoälysovellusten käytön katsotaan vaarantavan luottamuksellisen ja vastuullisen arvioinnin. Hakemusten kirjoittamisessa tekoälyn käyttö on sallittua, mikäli se ei vaaranna hyvän tieteellisen käytännön periaatteita, mm. sovellusten käytöstä tulee kertoa avoimesti.

Suomen Akatemia ei ole toistaiseksi ohjeistanut hakijoita tekoälysovellusten käytöstä hakemusten valmistelussa eikä ole koettu tarkoituksenmukaisena kieltää yksittäisten sovellusten käyttöä. HTK-asiantuntijoiden kanssa on laadittu sisäistä ohjeistusta tilanteeseen, että arviointipaneelissa nousee tekoälysovellusten käyttö hakemuksen arvioinnin kannalta negatiivisessa valossa. Arvioitsijoiden ohjeisiin on lisätty huomautus siitä, että hakemuksia ja arviointidokumentteja koskevat salassapitosäädökset koskevat myös tekoälytyövälineiden käyttöä. Kansainvälisistä verrokeista esimerkiksi Vetenskapsrådet, NordForsk, NIH ja NSF ovat laatineet ohjeet tekoälyn käytöstä. Nuutinen totesi lopuksi, että Akatemia hyödyntää tiedepoliittista valmistelua tukevassa analyysissään koneoppimiseen pohjaavaa aihemallinnusta.

Keskustelun aluksi Annukka Pakarinen tiivisti ammattikorkeakoulujen tekoälyn käytön sanoilla ”rohkea kokeilu yhdistettynä viisaaseen varovaisuuteen.” Arene on laatinut yhteiset suositukset tekoälyn hyödyntämisestä, joiden pohjalta osa ammattikorkeakouluista on laatinut omat laajemmat ohjeet. Yli puolet ammattikorkeakouluista on laatinut ohjeistuksia koulutuksen käyttöön (opiskelijat ja/tai opettajat), ja nämä sisältävät usein myös yleisempiä ohjeistuksia. Tutkijoiden käyttöön on vähemmän ohjeistuksia, mutta suunnitelmia ja tilanteen seurantaan tehdään aktiivisesti. Saku Anttila totesi, että vain muutamalla tutkimuslaitoksella on omat ohjeet tekoälyn hyödyntämisestä tutkimuksen tueksi. Nämä ovat yleisempiä raameja, jonka alle on koottu eri työkalujen lisäohjeita. Keskustelu tutkimuslaitoksissa keskittyy generatiiviseen tekoälyyn. Erilaisia pilotteja on kehitetty erityisesti tietoturvan ja tietosuojan näkökulmasta. Tuotiin esiin, että SYKE:lla on oma tietoturvallinen ympäristö ja VTT:ssä nähdään, että käyttöönotto edellyttää tutkijoiden aktiivisuutta.

Puheenjohtaja kysyi tekevätkö mahdollisesti Unifi, Arene, Tulanet tai Finn-ARMA tähän liittyen jo yhteistyötä. Todettiin, että tässä olisi oiva kansallinen vertaisoppimisen mahdollisuus. Osallistujat totesivat, että yhteistyölle ei ole varsinaista foorumia ja että olisi hyödyllistä tunnistaa tahoja, joilla olisi osaamista aiheen tiimoilta. Pohdittiin, että kun ohjeita on jo paljon saatavilla, mitä Suomessa pitäisi tehdä ja tarve olisi ohjeiden konkretisointi edelleen eli niiden tuominen arjen tasolle sekä tietoisuuden kasvattaminen erilaisista jo valmistelluista dokumenteista.

6. Monimuotoinen data

Heikki Mannila alusti keskustelua. Tieteellisen tutkimuksen kannalta kiinnostavat aineistot ovat hyvin erityyppisiä ja tutkimuksen digitalisaatio vaikuttaa tutkimusprosessiin monin tavoin, ei pelkästään varsinaisen laskennan ja datan tallennuksen teknisen toteutuksen kautta. Nousee uusia mahdollisuuksia, uusia ongelmia ja uusia osaamistarpeita. Mahdollisia politiikkatoimia ovat laskennan ja tallennuksen infrastruktuurin koko kirjosta huolehtiminen sekä muuttuviin osaamistarpeisiin vastaaminen, erityisesti datan ja mallinnuksen näkökulmasta.

Tutkimuksessa käytetään tai syntyy erilaista tietoa ja sitä on tarve yhdistellä. Käytettävät aineistot ovat monimuotoisia: eri luonnonilmiöistä ja rakenteista tai yhteiskunnallista ilmiöistä kerätään tai mitataan dataa, joskus simuloidaankin. Tietoa voidaan tuottaa tai kerätä uudestaan, kyseessä voi olla pitempään ylläpidettävä ja kartutettava datakokoelma tai sitten yksittäisen tutkimuksen tai jotakin muuta käyttöä varten muodostettu data. Lopulta tieto saattaa vanhentua ja sitä pitää poistaa. Pohdittavaksi tulee myös käyttäjämäärät, tallennus, kustannukset, sekä manuaalisen kuratoinnin määrä. Datanäkökulmasta tutkimusprosessiin liittyvät eri alojen tarpeet (esim. menetelmätieteet vs. empiiriset tieteet, kvalitatiivinen vs. kvantitatiivinen). Dataan liittyvä osaaminen tulee huomioida esimerkiksi tutkimusaiheen ja kysymysten muodostamisessa, aineistojen hallinnassa sekä yhdistelyssä, datojen kohteen sekä saatavuuden osalta, aiemman tutkimuksen hyödyntämisessä, datojen ominaisuuksien kuten virhelähteiden ja analyysin sekä johtopäätösten hahmottamisessa.

Keskustelussa todettiin, että data tulee keskeiseksi kaikille aloille ja pohdittiin mm. urapolkuja tieteellisen laskennan ja ohjelmoinnin tekniselle tuelle (research software engineer) sekä sitä, että ihmistä tarvitaan edelleen datan keruuseen ja putsamiseen. Lisäksi pohdittiin miten korkeakoulujen johto resursoi työtä. Tulee tiedostaa, että ”data” käsitteenä ei ole homogeeninen. Minkä tyyppiset alat kukoistavat? Nostettiin esiin, että alat, joilla on runsaasti dataa saatavilla saattavat jatkossa dominoida ja mikäli toisilla aloilla on vähän dataa, erilaisten työkalujen hyödyt voivat jäädä vähäisiksi.

7. Muut asiat

Keskusteluun ei noussut muita asioita, mutta todettiin, että seuraavaan kokoukseen on syytä hahmotella väljempi asialista ja jatkaa tämän keskustelun prosessointia.

8. Seuraava kokous

Seuraava kokous pidetään 11.11.2024 sosiaali- ja terveysministeriössä (Meritullinkatu 8).

9. Kokouksen päättäminen

Puheenjohtaja päätti kokouksen kello 14.02.