

Tutkimuksen digitalisaation politiikkatoimien kokoamisen ja kansallisen datanhallinnan ja laskennan tutkimusinfrastruktuuriympäristön päivittämisen (DL2026) seurantaryhmän alatyöryhmä / versio 25.8.2025

Työryhmä ja sen toimeksianto

Opetus- ja kulttuuriministeriö asetti 24.1.2025 alatyöryhmän Tutkimuksen digitalisaation politiikkatoimien kokoamisen ja kansallisen datanhallinnan ja laskennan tutkimusinfrastruktuuriympäristön päivittämisen (DL2026) seurantaryhmälle. Alatyöryhmän tehtäväksiinto toteaa

”Tutkimuksen ja innovaatiotoiminnan tekemisen tavat digitalisoituvat teknologisen kehityksen myötä. Tekoälyn kehitys sekä datan ja laskennan roolin kasvu merkitsevät suuria muutoksia kaikilla tieteenaloilla ja koko innovaatiotoiminnassa. Tämä tutkimuksen digitalisaatioksi kutsuttu muutos etenee nopeasti.”

Alatyöryhmän tehtävä on määritelty

”Alatyöryhmän tehtävänä on luonnostella nykytilan kuvaus sekä kolme vaihtoehtoista tulevaisuuden skenaariota tutkimuksen digitalisaation edistämiseksi.”

Alatyöryhmään nimettiin prof. Heikki Mannila (puheenjohtaja, Aalto-yliopisto), Pekka Marttinen (Aalto-yliopisto), Esa Rahtu (Tampereen yliopisto) ja Laura Ruotsalainen (Helsingin yliopisto).

Tausta

Tutkimuksen ja korkeakouluopetuksen toimintaympäristö muuttuu nopeasti. Nopeasti etenevä ilmastonmuutos ja luontokato tulevat vaikuttamaan monin eri tavoin yksilöihin ja yhteiskuntiin. Geopoliittinen kehitys on hankalasti ennustettavissa, mutta viimeksi kuluneiden vuosikymmenten kohtuullisen vakaa tilanne näyttää muuttuneen huomattavasti epävarmemmaksi.

Teknologinen kehitys ja erityisesti tekoälyn nopea kehitys tulee muuttamaan yhteiskuntia. Tekoäly vaikuttaa monien eri toiminta-alueiden tuottavuuteen hyvinkin radikaalilla tavalla ja siten eri maiden kilpailukyky saattaa muuttua merkittävästi. Tekoäly ja muut digitaaliset teknologiat vaikuttavat yhteiskuntien toimintaan myös muilla tavoin.¹

¹ Tekoälyn vaikutusten pohtimisessa on aina hyvä muistaa se, että tekoälymenetelmiin kuuluu laaja kirjo erilaisia lähestymistapoja. Esimerkiksi kielimallit ovat vain yksi, toki merkittävä, osa tekoälymenetelmien kokonaisuudesta.

Yhteiskunnallinen muutos vaikuttaa väistämättä myös tutkimukseen sekä suoraan että muuttuvan toimintaympäristön kautta. Onkin tarpeen pohtia sitä, mitä vaikutuksia teknologisella muutoksella ja murroksella on Suomen talouteen ja yleisemmin suomalaiseen yhteiskuntaan, ja miten nämä vaikutukset heijastuvat Suomessa tehtävään tutkimukseen.

Raportti jakautuu kahteen osaan: ensin tarkastellaan tutkimuksen digitalisaatiota perinteisemmässä merkityksessään ja sitten siirrytään pohtimaan tekoälyn kehittymisen merkitystä tutkimukselle ja tutkimuksen tekemiselle.

Tutkimuksen digitalisaatio

Tutkimuksen digitalisaatioon keskeinen komponentti on tutkimuksen ja innovaatiotoiminnan tekemisen tapojen jo toteutunut ja edelleen jatkuva muutos. Empiiristen havaintojen ja teorianmuodostuksen lisäksi tutkimuksessa ovat aikaisempaa enemmän nousseet esille laskennalliset menetelmät ja suurten tai muuten monimutkaisten aineistojen käsittely. Samalla monitieteinen yhteistyö varsinaisen aiheen tutkijoiden ja laskennallisten menetelmien kehittäjien välillä on muodostunut aiempaa tärkeämmäksi. Tutkimuksen digitalisaation osana ovat uusien mittausten menetelmien kehittyminen ja niiden kautta uudentyyppiset tutkimusaineistot sekä uusien laskentamenetelmien kehittäminen tällaisten aineistojen käsittelyyn.

Esimerkiksi sopii molekyylibiologia, jossa esim. genomien rakennetta koskevan datan määrä on 1980-luvulta alkaen kasvanut eksponentiaalisesti. Uudet molekyylibiologiset mittausten menetelmät tuottavat uudenlaisia tieteellisiä aineistoja, joiden käsittelyä varten on kehitetty uusia laskennallisia menetelmiä. Näiden laskentamenetelmien kehityksellä on ollut merkittävä vaikutus myös tietojenkäsittelytieteen sisällä: uudet kysymyksenasettelut ovat osoittautuneet hedelmällisiksi myös alueen oman kehityksen näkökulmasta.

Molekyylibiologia tarjoaakin yhden erittäin laajan esimerkin siitä, miten tutkimuksen digitalisaatio on vaikuttanut tieteen kenttään. Myös ilmastotutkimus, paikkatiedon käsittely, digitaaliset ihmistieteet ja kvalitatiivisten aineistojen käsittely digitaalisten tekstianalyysimenetelmien avulla ovat esimerkkejä siitä, miten uudentyyppiset aineistot, niihin liittyvät uudet tutkimuskysymykset ja analyysimenetelmät ovat johtaneet tutkimuksen merkittävään sisällölliseen muutokseen.

Tilannekuva Suomessa

Tutkimuksen digitalisaation tilanteesta eri maissa ei ole kovin selkeätä tietopohjaa olemassa, ja siksi pohjautumme jatkossa laadullisiin havaintoihin Suomen tilanteesta.

Suomessa laskennallisia menetelmiä muihin tieteisiin yhdistävä tutkimus käynnistyi varsin aikaisin. Esimerkiksi molekyylibiologian ja tietojenkäsittelytieteen tutkijoiden yhteistyö alkoi jo 1980-luvulla. Tutkimuksen digitalisaation kannalta Suomen Akatemian akatemiaohjelmat, huippuyksiköt ja muut rahoitusmuodot ovat olleet merkittäviä laskentamenetelmien, niiden soveltamisen ja monitieteisen yhteistyön kehittämisessä.

Laskennallisten tieteiden ja tutkimuksen digitalisaation menetelmien kehittämisen lisäksi Suomessa on panostettu laskentainfrastruktuurien kehittämiseen. Aivan keskeinen osa tässä

kehityksessä on ollut Tieteen tietotekniikan keskuksen CSC:n toiminta ja sen resursseista huolehtiminen. Merkittävää on ollut myös se, miten Suomessa on investoitu laskennallisen tieteen laitteistojen koko spektriin, pienemmistä yliopistojen ja tutkimuslaitosten omista laskentalaitteistoista kansallisesti ja kansainvälisesti hyvinkin mittaviin järjestelmiin.

Tutkimuksen digitalisaatio ja korkeakoulujen henkilöstörakenne

Tutkimuksen digitalisaation kehittyessä on tarpeen pohtia myös sitä, minkälaisia tarpeita yliopistojen henkilöstörakenteelle uudentyyppinen tapa tehdä tiedettä aiheuttaa. Ei ole selvää, että perinteinen rakenne, jossa yhden tutkimusryhmän tutkimushenkilöstö muodostuu tietyn alan väitöskirjaopiskelijoista, tutkijatohtoreista ja professoreista, tarjoaa parhaan pohjan tieteelliselle työlle, jossa laskennallisilla menetelmillä on keskeinen rooli. Sekä digitaalisten että muiden tutkimusmenetelmien kehittyminen antaa aihetta pohtia sitä, tarvitaanko kenties enemmän henkilöstöä, jotka keskittyvät pitkäjänteisesti fyysisten (ei-laskennallisten) ja laskennallisten tutkimusinfrastruktuurien kehittämiseen.

Ei-laskennallisten tutkimusinfrastruktuurien osalta tämä tarve on pitkään tunnistettu ja ratkaisuna on käytetty laboratorioinsinööri-tyyppisiä tehtäväkuvia. Tyypillisesti laboratorioinsinöörit ovat kiinteästi yhden laitoksen tai tutkimusryhmän työntekijöitä, joilla on syvä tutkimusalan ja sen mittausten menetelmien tai aineistojen tuntemus. He harvemmin työskentelevät monien eri alojen parissa.

Tilanne laskentamenetelmien erityisosaajien suhteen on osittain erilainen: tyypillisesti yksittäinen tutkimusryhmä tarvitsee heidän osaamistaan vain tutkimusprosessin tietyissä kohdissa. Tällaisen henkilön ja tutkimusorganisaation kannalta on siksi edullista, jos tällaisten henkilöiden työpanosta voidaan käyttää laajemmalti organisaation sisällä.

Tällaisesta uudesta henkilöstöryhmästä käytetään englanninkielistä termiä *research software engineer* (RSE), ja suomalaiset yliopistot ovat osin alkaneet käyttää tällaista henkilöstöryhmää. Kokemukset RSE-henkilöstön työstä ovat mm. Aalto-yliopistossa erittäin positiivisia. Myös juuri käynnistynyt AI Factory sisältää merkittävän investoinnin RSE-henkilöstöön. Tutkimuslaitoksissa RSE-tyyppistä henkilöstöä on ollut pitempäänkin.

Työryhmä katsookin, että korkeakoulujen tulisi tutkimuksen digitalisaation kehittämisen yhteydessä pohtia henkilöstörakenteitaan ja mahdollista tarvetta esim. RSE-tyyppisten ratkaisujen kehittämiseen.

Tekoälyn vaikutus tutkimukseen ja sen tekemiseen

Tekoäly kehittyminen on hyvin merkittävä teknologinen muutos, joka etenee hyvin nopeasti ja vaikuttaa koko yhteiskuntaan: teollisuuteen, oppimiseen, osaamiseen, vapaa-aikaan jne. Kuten jo edellä todettiin, tekoälyn vaikutus tutkimukseen on sekä *välillinen* (tekoäly muuttaa yhteiskuntaa ja sitä kautta tutkimusta) että *suora* (tekoäly muuttaa tutkimuksen tekemisen tapoja).

Tekoälyn välillinen vaikutus tutkimukseen on vielä hankalammin ennustettavissa kuin tekoälyn suora vaikutus. Tässä keskitymmekin muutamaan arvioon suoran vaikutuksen muodoista.

Tekoälyn vaikutusta tutkimukseen käsiteltiin mm. kahdessa vaikutusvaltaisessa raportissa vuonna 2024: Royal Societyn laatimassa ”Science in the age of AI” -dokumentissa² ja USA:ssa Joe Bidenin presidenttikaudella laaditussa raportissa ”Supercharging Research: Harnessing Artificial Intelligence”³..

Näiden kahden varsin samanlaisen raportin havainnot ja toimenpide-ehdotukset liittyvät suureksi osaksi tutkimusprosessin laajoihin osiin ja teemoihin, jotka ovat olleet esillä jo ennen kuin tekoälyn kehityksen nopeus kiihtyi merkittävästi. Kuluvan vuoden 2025 aikana on näiden esimerkkien lisäksi nähty paljon muunkinlaisia esimerkkejä myös pienempiin tutkimusprosessin osiin liittyvistä tekoälyn käyttökohteista.⁴ Mm. tällaiset esimerkit vahvistavat näkemystä, jonka mukaan tekoälyn kehityksen suoratkin vaikutukset tutkimuksen tekemiseen tulevat olemaan merkittävämpiä kuin vielä vuonna 2024 arvioitiin.

Joitakin kehitystrendejä

Tässä jaksossa esittelemme joitakin käynnissä olevia kehityskulkuja tekoälyn käytössä tutkimuksessa.

Data-analyysin mahdollisuudet lisääntyvät. Tutkimuksen digitalisaatio on merkinnyt sitä, että vapaasti käytettävissä on hyvin paljon avoimen lähdekoodin data-analyysiohjelmistoja. Tällä on suuri merkitys tutkimuksen tekemisessä. Tekoälyn kehittyessä erityisesti pienen skaalan ohjelmistojen tekeminen on helpottunut entisestään. Siten tutkimuksessa on aikaisempaa helpompaa laatia ja käyttää monimutkaisia data-analyysimenetelmiä. Tällöin tarve menetelmien perusteiden ja oletusten ymmärtämiselle lisääntyy entisestään.

Teksti-, audio- ja videoaineistojen käyttö tutkimuksessa helpottuu ja siksi niitä käyttävä tutkimus yleistyy. Esimerkiksi laajojen tekstiaineistojen analysointi puhtaasti ihmistyönä on työlästä ja kallista. Kielimallit tarjoavat – ainakin periaatteessa – mahdollisuuden analysoida suuria tekstimassoja nopeasti, ja näin esim. kvalitatiivisen tutkimuksen tekeminen voi olla yksinkertaisempaa kuin ennen. Samalla lailla esim. audio- tai videoaineistojen analyysi tekoälymalleja käyttäen on aikaisempaa helpompaa. Tulosten luotettavuuden arviointi on kuitenkin edelleen suuri ongelma erityisesti näillä alueilla.

Data-analyysin helpottuminen ja uudentyyppisten aineistojen aikaisempaa helppokäyttöisemmät analyysimenetelmät ohjaavat tutkimuksen suuntautumista.

Tutkimuksessa käytettävissä olevat menetelmät vaikuttavat siihen, mitä tutkitaan. Esimerkiksi mikroskoopin kehittyminen johti siihen, että kooltaan pieniä luonnonilmiöitä alettiin tutkia enemmän kuin ennen. Sama ilmiö on havaittavissa nyt: kun data-analyysi on (ainakin periaatteessa) helpompaa kuin ennen ja kun tekstin, audion ja videon käsittely (ainakin periaatteessa) yksinkertaistuu, niin näitä asioita tehdään tutkimuksessa enemmän.

² <https://royalsociety.org/-/media/policy/projects/science-in-the-age-of-ai/science-in-the-age-of-ai-report.pdf>.

³ https://bidenwhitehouse.archives.gov/wp-content/uploads/2024/04/AI-Report_Upload_29APRIL2024_SEND-2.pdf.

⁴ Esimerkkinä ks. <https://mathstodon.xyz/@tao/114173696303072269>, jossa Fieldsin mitalin voittanut matemaatikko Terence Tao kertoo, miten hän käytti kielimallia tietyn kaavan löytämiseen (!) ja kaavan perusteella tapahtuvan numeerisen simulaation koodaamiseen.

Tekoälymenetelmiä käytetään myös tutkimusprosessin ytimessä. Edelliset esimerkit ovat kuvanneet tekoälyn käyttöä tutkimuksen aineistojen analysoinnissa. Mm. materiaalitutkimuksessa tekoälymenetelmiä on kuitenkin alettu käyttää myös tutkimusprosessin ytimessä: esim. etsittäessä sopivia yhdisteitä voidaan koneoppimisen avulla nopeuttaa ja parantaa mahdollisten ehdokasyhdisteiden löytämistä, jolloin varsinaisia kokeita joudutaan tekemään vähemmän.

Tekoälymenetelmien käyttäminen tiedonhaun ja kirjoittamisen apukeinona yleistyy tutkimuksessa. Suurten kielimallien myötä esim. tutkimusartikkelien kirjoittajat käyttävät jatkuvasti enemmän tekoälymenetelmiä kirjoittamisen apuna. Tähän kehitykseen liittyy positiivisiakin piirteitä (esim. artikkelien kieliasu paranee), mutta myös paljon ongelmia, kun kielimallien tuottamaa materiaalia käytetään tarkistamatta.

Tekoälymenetelmien luotettavuus on edelleenkin ongelma. Vaikka tekoälymenetelmien antamien tulosten luotettavuus on parantunut, on tulosten verifiointi ja tutkimuksen toistettavuuden varmistaminen edelleenkin keskeinen kysymys. Erityisesti suuret kielimallit tuottavat usein selkeästi virheellisiä tuloksia. Tekoälyä tutkimukseen (tai muuhun) käytettäessä tulee käyttää menetelmiä, joiden tulokset voidaan tarkistaa. Uusimmat kielimallit tekevät tämän aikaisempaa helpommaksi tarjoamalla myös suoria linkkejä vastauksensa lähteisiin. Mallien muuttuessa analyysitulokset saattavat muuttua, ja tämä tarkoittaa hankaluuksia tutkimuksen toistettavuudessa.

Monitieteisyyden merkityksen kasvu. Tutkimuksen digitalisaation ja tekoälymenetelmien käytön myötä tutkimustyössä tarvitaan entistä laajempaa menetelmätieteiden ja muiden tieteiden yhteistyötä.

Tutkimuksen tekoälymenetelmien kehittyminen tapahtuu sekä yleisten tekoälymenetelmien kehityksen kautta että alakohtaisen työn kautta. Esim. kielimallit ovat ainakin osittain varsin yleiskäyttöisiä menetelmiä hyvine ja huonoine ominaisuuksineen. Toisaalta esim. matematiikan alueella on käynnissä paljon työtä, jossa kehitetään tekoälymenetelmiä nimenomaan matematiikan alueen tutkimuskäyttöön.

Viisi skenaariota tutkimuksen digitalisaation edistymiselle

Ohessa viisi skenaariota siitä, miten tutkimuksen digitalisaatio etenee ja erityisesti siitä, miten tekoälyn nopea kehitys tulee vaikuttamaan tutkimukseen. Nämä skenaariot eivät ole vaihtoehtoisia, vaan pidämme todennäköisenä, että osia useista skenaarioista toteutuu samaan aikaan.

1. Tutkimuksen digitalisaatio etenee ja kattaa aikaisempaa laajemman osan tutkimuksesta. Laskennalliset menetelmät, suuret datajoukot ja tekoäly ovat osa yhä useamman sovellusalan tutkimusta ja niitä käytetään yhä useammassa tutkimusprosessin vaiheessa.
2. Tekoälyn nopea kehitys jatkuu ja siitä kehittyy hyvin merkittävä tutkimuksen apuväline, joka helpottaa globaalien haasteiden ratkaisemista.
3. Tekoälyn nopea kehitys jatkuu, mutta tutkimuksen kannalta tuloksena on suuri määrä huonotasoista tutkimusta, joka ei vie tiedettä eteenpäin eikä ratkaise globaaleja haasteita,

vaan päinvastoin johtaa tutkimuksen tason laskemiseen mm. vertaisarvointiin liittyvien ongelmien takia.

4. Geopoliittiset ja muut tekijät johtavat siihen, että edistyneimmät tekoälymenetelmät ovat vain tiettyjen maiden tai tiettyjen tutkijoiden käytettävissä; tiede eriytyy ja julkaiseminen polarisoituu.
5. Tekoälyn nopea kehitys mullistaa kansainvälisen talouden ja merkittävä osa Euroopan asiantuntijatyöpaikoista katoaa tai siirtyy muualle. Tämän seurauksena länsimaat vajoavat syvään lamaan, levottomuudet leviävät ja tutkimuksen rahoittaminen loppuu.

Tällä hetkellä (heinä-elokuussa 2025) näyttää siltä, että sekä skenaario 1 (tutkimuksen digitalisaation jatkuminen) on selvästi toteutumassa.

Tekoälyyn kytkeytyvistä skenaarioista sekä tutkimuksen kannalta positiivinen skenaario 2 että negatiivinen skenaario 3 ovat toteutumassa ainakin osittain jo nyt. Näemme esimerkkejä erittäin laadukkaasta tutkimuksesta, jossa tekoälymenetelmien oivaltavalla käytöllä on suuri merkitys, mutta näemme myös esimerkkejä heikotasoisesta tutkimuksesta, jonka tekemisessä tekoälymenetelmiä on todennäköisesti käytetty runsaasti.

Skenaarioiden 4 ja 5 toteutumistodennäköisyyttä on vaikea ennustaa. Tekoälyn taloudellisia ja yhteiskunnallisia vaikutuksia luultavasti sekä yli- että aliarvioidaan, mutta selkeästi kyseessä on hyvin merkittävä muutos, joka tulee monella tavalla vaikuttamaan myös tutkimuksen tekemiseen, sekä suoraan että välillisesti.

Suositukseset

1. Tutkimuksen digitalisaatio, mm. laskennalliset menetelmät, data-analyysi ja tekoäly, tulee jatkossakin muuttamaan tutkimuksen tekemistä, ja nämä muutokset tulee ottaa huomioon tutkimuksessa ja tutkijakoulutuksessa.
2. Tutkimuksen digitalisaation edetessä korkeakoulut tarvitsevat entistä enemmän laskenta- ja tekoälymenetelmiä osaavaa henkilökuntaa, ja korkeakoulujen tulisikin pohtia urajärjestelmiään tästä näkökulmasta.
3. Tutkimuksen digitalisaation ja tekoälymenetelmien tarvitsemasta laskentainfrastruktuurista tulee pitää jatkossakin huolta.
4. Alojen välistä monitieteistä yhteistyötä tekoälymenetelmien kehityksessä ja soveltamisessa tulee kannustaa ja tukea johdonmukaisesti ja pitkäaikaisesti.
5. Tekoälymenetelmien kehittyminen vaikuttaa tutkimuksen tekemisen lisäksi myös opetukseen ja opetusmenetelmiin monimutkaisilla tavoilla, jotka oletettavasti ovat eri aloilla erilaisia. Alojen välinen yhteistyö on tärkeätä myös opetuksen näkökulmasta. Korkeakoulujen yhteistyö voisi olla hyödyllinen pohdittaessa esim. tekoälyn vaikutuksia jatkokoulutukseen ja myös alakohtaisia vaikutuksia selvittäessä. Tekoälymenetelmät vaikuttavat myös eri osaamisten yhteiskunnalliseen kysyntään ja koulutustarpeisiin.
6. Tekoälymenetelmien vastuullinen käyttö (sisältäen myös ilmasto- ja muut kestävyysnäkökulmat) tulee varmistaa.
7. Tutkimuksen digitalisaation ja tekoälymenetelmien kehittyessä korkeakouluissa tarvitaan alakohtaisen kehittämisen ja strategisen lähestymistavan yhdistämistä.