



Manna Satama

# Tekoälyllä parempaa tutkimusdatanhallintaa?

Tekoälyä käytetään tutkimuksessa, opetuksessa ja muussa asiantuntijatyössä tietoisesti - ja välillä tiedostamattakin. Tutkimusdatanhallinnan asiantuntijat pohtivat tekoälyn mahdollisuuksia ja rajoituksia eri näkökulmista: Mitä tutkijat tekevät tekoälyn avulla ja miten tekoäly voisi hyödyntää omassa tukipalvelussa?

**T**ekoäly on osa kirjaston asiantuntijoiden arkista työtä muun muassa tekstin tuottamisen ja kääntämisen apuna sekä tiedonhaussa, kuten STKS:n tekoälykyselyä käsittelevässä Signumin artikkelissa todetaan (Lahikainen et al. 2024). Mutta kuinka tekoäly voisi soveltaa tutkimusdatanhallinnan tukipalvelun asiantuntijatyöhön? Tätä pohdittiin 31. 10. 2024 Tampereella järjestetyssä DMP-konsortion verkostotapaamisessa.

## Datanhallinnan tuki ja tutkijan työ

Tutkijan työn ymmärtäminen on tutkimuksen tukipalvelussa tärkeää, ja tämä pitää paikkansa erityisesti datanhallinnan tukitehtävissä. Aineistohallintaa ja tutkimukseen liittyvää aineiston niin sanottua tutkimuksellista tai tutkimusmenetelmiin liittyvää käsittelyä ei ole aina helppo eikä syytäkään erottaa. Nämä kaksi kulkevat käsi kädessä, minkä huomion ympärille me rakensimme

Itä-Suomen yliopiston kirjastossa *Mastering the research data!* -opintojakson tohtoriopiskelijoille elokuussa 2024 (Karhapää et al. 2024).

Opintojaksolla ei erityisemmin kiinnitetty huomiota tekoölyyn, vaikka se toki oli mukana joidenkin tutkijoiden puheenvuoroissa. Sen sijaan heinäkuussa 2024 LIBERIN konferenssissa järjestettiin työpaja, jossa oli tarkoitus sukeltaa tekoölyyn tutkijan työssä. *Exploring AI Hands-On: Shaping the Future of Research Library Services* -työpajan kohdeyleisö oli tutkimusdatanhallinnan tukitehtävissä työskentelevät asiantuntijat. Työpajan idea oli, että jos halutaan tukea tekoölyä hyödyntäviä tutkijoita, on tukihenkilöidenkin pysyttävä mukana tekoölyn kehityksessä. Työpajassa tutustuttiin tekoölyyn data-tieteen näkökulmasta ja varsinkin tiiviissä aikataulussa, eikä siinä ollut mahdollista syventyä datanhallinnan tukihenkilöiden rooliin tai tekoölyosaamisen vaatimuksiin sen tarkemmin.

### **Fokus datanhallinnan tukeen ja tekoölyyn**

Tampereen marraskuisessa DMP-konsortion työpajassa sen sijaan fokus oli käytännönläheisempi. Ennen työpajaa päätin orientoitua teemaan. Kysyin tekoölyltä (Copilot), miten tekoöly voi auttaa tutkimusaineistohallinnassa. Vastauksena tuli kymmenen kohdan lista asioista, joista osa liittyi vahvasti tutkimusmenetelmiin ja analyysiin (datan kerääminen ja esikäsittely, tietojen analysointi ja kuvantunnistus,

### **Jos halutaan tukea tekoölyä hyödyntäviä tutkijoita, on tukihenkilöidenkin pysyttävä mukana tekoölyn kehityksessä.**

ennakoiva analytiikka, tietojen visualisointi) ja muutama linkittyi selvemmin aineistohallintaan (tietoturva, automaattinen metadatan luonti). Vastaus tuki aineistohallinnan ja tutkimuksen tekemisen tiivistä liitosta.

Halusin myös vastauksia siihen, miten tekoöly voisi auttaa meitä aineistohallinnan tukipalveluissa työskenteleviä omassa työssämme ja miten voisimme auttaa tutkijoita tekoölyn hyödyntämisessä juuri aineistohallinnan suunnittelussa. Tämän kysymyksen esitin Copilotille englanniksi, sillä englanniksi datanhallinnan asiantuntijalle on jossain määrin vakiintuneempia nimikkeitä (esim. data steward). Vastaukset olivat osin samankaltaisia kuin ensimmäisessä kysymyksessäni, mutta nyt mukaan tulivat chatbotit ja virtuaaliassistentit usein kysytyjen kysymysten taklaajina sekä datalinjauksen noudattamisen tarkkailijoina. Copilotin mukaan sen listaamat asiat tukevat datanhallinnan asiantuntijan työn tehokkuutta, jolloin asiantuntijat voisivat keskittyä strategisempiin tehtäviin. Näillä eväillä oli hyvä lähteä kokeilemaan tekoölyä ja pohtimaan tuloksia kollegoiden kanssa.

## Tutkimusdatan laadusta, avoimuudesta ja vähän muustakin

Työpaja alkoi Jukka Rantasaaren (Turun yliopiston kirjasto) alustuksella generatiivisen tekoölyn hyödyntämisestä datanhallinnassa ja aineistohallintasuunnitelmissa. Hän oli myös pitkälti vastannut ryhmätyöskentelyn erinomaisista valmisteluista tekoöllylle syötettävistä taustatiedoista ja esimerkiksi kehoitteista alkaen. Tausta-aineistot olivat täysin keinotekoisia eivätkä sisältäneet henkilötietoa tai salassa pidettävää tietoa - emme siis syöttäneet tekoölysovelluksiin vaikkapa tutkimus- tai aineistohallintasuunnitelmia, jotka ovat salassa pidettäviä dokumentteja. On toki muistettava, että tekoölysovellusten ja -työkalujen kohdalla noudatetaan oman organisaation ohjeita, sillä käytössä saattaa olla myös suljettuja ja tietoturvallisia ratkaisuja.

Pienryhmissä kokeiltiin muun muassa tiivistelmien luomista kirjallisuuskatsausta varten, raakadatan siivoamista, henkilötietojen anonymisointia ja metatiedon luomista. Nämä teemat liittyvät tutkimuksen tekemiseen ja sellaisiin toimenpiteisiin, joiden syvällinen osaaminen ei useinkaan kuulu datanhallinnan yleisiantuntijoiden toimenkuviin mutta joiden tuntemus on kuitenkin tärkeää. Teemat liittyvät datan laatuun, tietosuojaan ja ku-

vailuun, jotka kaikki ovat keskeisiä datanhallinnassa ja joihin datanhallinnan asiantuntijat ottavat kantaa kommentoidessaan tutkijoiden laatimia aineistohallintasuunnitelmia. Jos tutkija kertoo käyttävänsä tekoölyä esimerkiksi metadatan luomiseen, datanhallinnan asiantuntijan on osattava kertoa mahdollisista puutteista, joihin tutkijan olisi kiinnitettävä huomiota. Tekoöly suoriutui metatiedon luomisesta ja raakadatan siivoamisesta vaihtelevasti ja osin myös hyvin.

Entä miten tekoöly hoitelee tietosuoja-asiat ja aineistojen avaamisen suunnittelun? Tekoölyn kanssa tehtävien palastelu ja kehoitteiden tarkkuus on tärkeää. Esimerkiksi anonymisointitehtävästä pienen esimerkkiaineiston kanssa tekoöly suoriutui askel askelelta edettäessä ihan mallikelpoisesti. Datana avaamisen yhteydessä tutkijat joutuvat miettimään omalle datalleen sopivan repositorion, mihin me data-tukipalveluissa tarjoamme tukea. Joillain tutkimusaloilla saattaa olla selkeät käytänteet ja vakiintuneet repositoriot, mutta toisinaan tutkija saattaa joutua aikamoisen selvitystyön eteen. Tekoöly suoriutui etsinnässä vaihtelevasti, ja tarkkana noissakin vastauksissa saa olla.

## Aineistohallintasuunnitelman apuri ja kouluttajien tuki

Muutama harjoitus liittyi aineiston-

**Tekoöly suoriutui metatiedon luomisesta ja raakadatan siivoamisesta vaihtelevasti ja osin myös hyvin.**



hallintasuunnitelman laatimiseen tutkimussuunnitelman perustella ja aineistohallintasuunnitelman arviointiin. Aineistohallintasuunnitelma-annettiin pohjaksi suomalainen DMP-mallipohja ja arviointiin avuksi muun muassa DMP-arviointiohjeistus (Tuuli working group 2021). Näissä molemmissa tekoäly oli tietenkin nopea mutta tuotti hyvin yleistä tekstiä.

Erityisesti aineistohallintasuunnitelman laatimisessa tekoäly pääsi parhaimmillaankin köykäisen suunnitelman tasolle, juuri ja juuri hyväksytyyn suoritukseen vaikkapa opintojakson osana. Aineistohallintasuunnitelman tulee olla mahdollisimman konkreettinen, ja tekoällylle tyypilliset liian yleisluonteiset lauseet saavat datanhallinnan asiantuntijan punakynän viuhumaan. Tekoällyn tulosta pystyisi varmasti parantamaan, jos tausta-aineistoksi syöttäisi organisaation tarkemmat ohjeet ja vaikkapa tarkkaa tietoa tarjolla olevista levytiloista.

Arvioinnista tekoäly suoriutui ehkäpä paremmin, mutta tässäkin tarvittiin ihmisasiantuntijan kriittistä lukua: tekoäly antoi aineistohallintasuunnitelmasta erinomaisen yleisen arvion, vaikka sillä oli liuta parannusehdotuksia. Tämä saattaa toki olla tekoällyn koodatun kohteliaisuuden tai suoma-

## **Tekoällylle tyypilliset liian yleisluonteiset lauseet saavat datanhallinnan asiantuntijan punakynän viuhumaan.**

laisen korvaan särähtävän turhan ylistyksen syytä.

Koulutusmoduulin luomisessa tekoäly tuotti toimivia kurssirakenteita ja päteviä perussisältöjä. Se tuntui kuitenkin myös jäävän usein hyvin yleiselle tasolle ja antavansa epäolennaisia vastauksia vaikkapa koulutuksen kestosta.

### **Monimutkaisia kysymyksiä, yksinkertaisia vastauksia tai toisin päin?**

Tekoäly tuottaa tekstiä ja koodia nopeasti. Tekstiä on runsaasti, mutta se on paikoitellen toisteista ja tylsääkin. Teksti voi olla sisällöllisesti laadukasta tai virheellistä. Nämä jo ennestään tutut havainnot tulivat selkeästi esille myös tutkimusaineistohallinnan tekoälytestaamisessa. Aineistohallinnan kohdalla tarvitaan usein räätälöityjä ja jo suunnitelmaa varten riittävän konkreettisia ratkaisuja, mihin tekoäly ei välttämättä kykene. Sen vastaukset saattavat olla niin yleisiä, että ne eivät läpäise kriittistä arviointia. Tekoällyn tuotosta on luettava suurenuslasin kanssa. Tekoäly saattaa antaa monimutkaiseen aiheeseen tai kysymykseen liian yksinkertaisen ja sisällöllisesti tyhjähkön vastauksen. Toisaalta yksinkertaiseen kysymykseen voi tulla



pahimmillaan turhalla sanahelinällä kuorrutettu vastaus.

Tyhjän paperin kammon selättämiseen tekoöly on erinomainen, ja tämä pitää paikkansa myös aineistohallinnassa. Se luo rakennetta, antaa malliksi sopivia esimerkkejä ja ohjaa lisätietoihin väsymättömän assistentin tavoin. Se ei myöskään ole turhan niuho oikeakielisyyden vaatimuksessa, joka joskus hankaloittaa erilaisten hakupalveluiden tai portaalien käyttöä.

Kehotteiden ja sopivien tausta-aineistojen merkitys on aivan keskeinen. Tämä tarkoittaa sitä, että asiasta on jo jotain tiedettävä ennen kuin sitä voi työstää tekoölyn avustuksella. Ehkäpä datanhallinnan asiantuntijat voi-

vat tukea tutkijoita juuri kehotteiden ja tausta-aineistojen käytössä. Samoin tietoturvanäkökulma on tukipalveluisa keskeinen: tukipalvelu voi tarjota tietoa siitä, mitä tekoölytyökalua saa käyttää salassa pidettävään tai erilaisten tekijänoikeuksien alaiseen aineistoon.

Tekoölytyökalut muuttuvat nopeasti. Tämän artikkelin havainnot saattavat olla vanhentuneita jo muutamassa kuukaudessa. Tekoölyä koulutetaan jatkuvasti, mutta se ei tapahdu itsensä eikä ilmaiseksi. Kouluttamiseen vaaditaan aikaa, rahaa, tallennustilaa ja ihmisiä. Tähän kokonaisuuteen kuuluu myös kokonaisvaltainen eettinen pohdinta. ❖

## Lisälukemista ja lähteet

KARHAPÄÄ, A., NISKANEN, N., RAHNASTO-RILLA, M. & SATAMA, M. (2024).

Tutkimusaineiston haltuun! | Mastering the research data! Blogiteksti, Itä-Suomen yliopiston kirjasto 2. 10. 2024. <https://blogs.uef.fi/ueflibrary/tutkimusaineisto-haltuun-mastering-the-research-data/>

LAHIKAINEN, J., LAUHAKARI, A.-P., LUOKKANEN, S., & TIITTANEN, L. (2024). Hyvä apu työssä – STKS:n tekoölykyselyn satoa. *Signum*, 57(3), 4–10. <https://doi.org/10.25033/sig.148541>

Tuuli working group (2021). Finnish DMP evaluation guidance. Zenodo. <https://doi.org/10.5281/zenodo.4762326>

## Kirjoittaja

MANNA SATAMA

Itä-Suomen yliopiston kirjasto

manna.satama@uef.fi

<https://orcid.org/0000-0003-3775-9363>