



KANSALLISARKISTO

Kansalliset tietoaineistot tutkimuskäyttöön tekoälyn aikana

12.2.2025

Tanja Välisalo

Tutkimuspäällikkö





KANSALLISARKISTO

Kansalliset tietoaineistot tutkimuskäyttöön tekoälyn aikana

12.2.2025

Tanja Välisalo

Tutkimuspäällikkö



Aineistot infrastruktuureina



Aineistot infrastruktuureina



KANSALLISARKISTO

Kansallisarkiston aineistot



450 milj. teratavua aineistoa
keskiajalta nykypäivään



240 hyllykilometriä
analogista aineistoa



Julkishallinnon, yhteisöjen ja
yksityishenkilöiden arkistoja

Julkishallinnon aineistot Kansallisarkistossa

Valtionhallinto

Väestötieto

Paikallishallinto

Patentti- ja
rekisterihallinto

Tuomioistuimet,
poliisi



Kansallisarkiston aineistot tukevat laajoja tutkimusteemoja



Kestävä
kehitys

Muuttoliikkeet



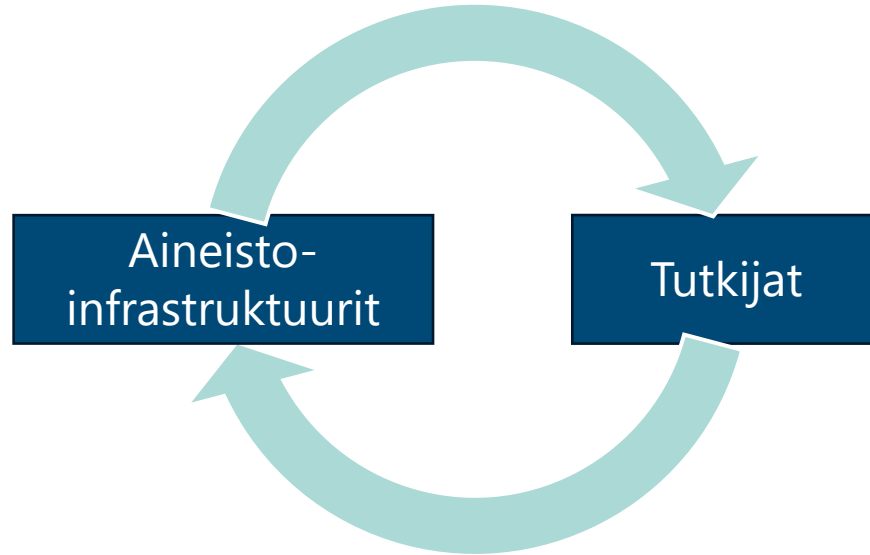
Hyvinvointi

Koulutus



Talous

Tutkimusaineistot tehokkaampaan käyttöön

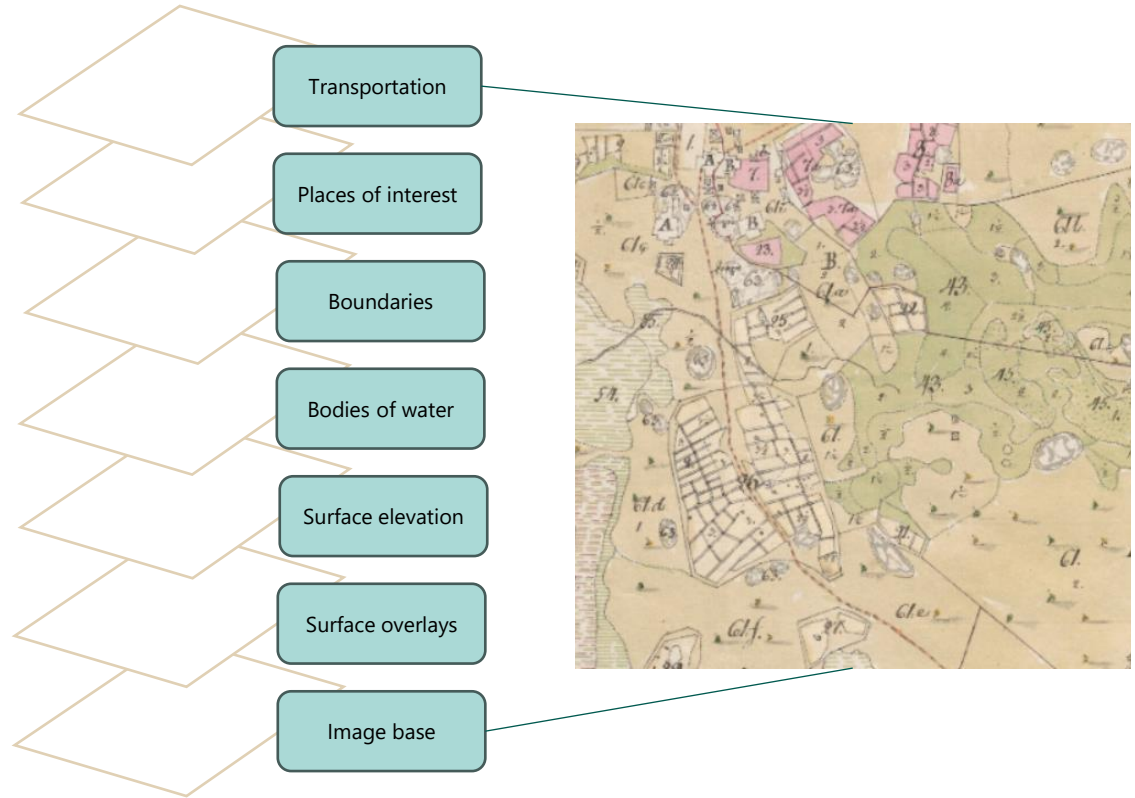


Joustavat tietoaineistot

Joustavat tietoaineistot

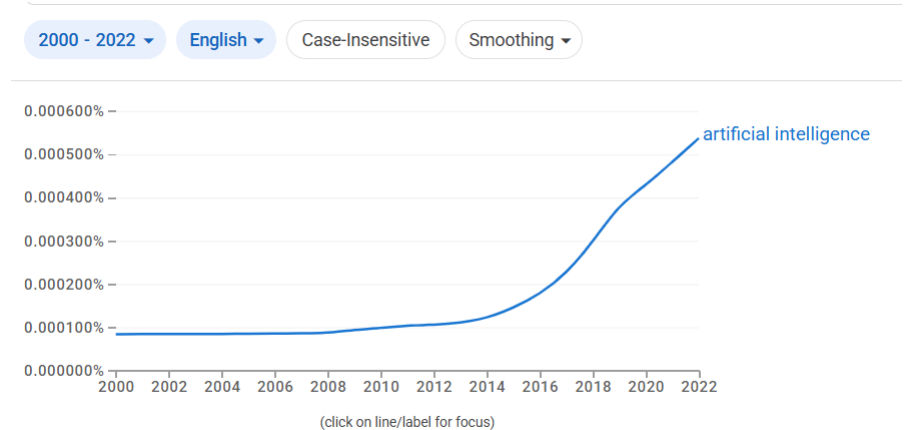
Digitaaliset menetelmät ajureina

Uudet tutkimusmenetelmät toisaalta ajavat muutosta, toisaalta tekninen kehittäminen mahdollistaa uusia menetelmiä



Tekoäly vastauksena?

- Suuret kielimallit ja generatiivinen tekoäly kehittyvät kiihtyvää tahtia
- Useimmat tekoälyratkaisut eivät yksinään tarjoa suoria hyötyjä tutkimukselle, vaan vaativat versioita ja sovelluksia, jotka on kehitetty tieteellisen tutkimuksen tarpeiden mukaan
- Ratkaisujen kehittäminen vaatii usein opetusaineistoja eli ihmisen annotoimaa aineistoa





Tekoälysovelluksia aineistoinfrastruktuureille

Digitoinnin tehostaminen ja laaduntarkastus

Keskusteleva hakubotti

Tehostetut tekstihaut suurista aineistomassoista

Automaattinen metatietojen tuotanto

Aineistojen automaattinen yhdistäminen

KANSALLISARKISTO

Kuva: Ilkka Jokipii

Esimerkkejä tekoälysovelluksista

Kansallisarkiston Tutkimus ja innovaatiot -yksikön tuotantoa:

- 1600-1900-luvut kattava HTR-malli **käsin kirjoitetun tekstin tunnistukseen**
- **Nimientieettien** tunnistukseen NER-malli 1900-luvun aineistoille
→ kestää skannatun tekstin virheet ja ymmärtää byrokratian kielen (yhteistyössä Jyväskylän yliopisto)

Kansallisarkiston tekoälyratkaisuja avoimesti jaossa:
<https://huggingface.co/Kansallisarkisto>

191# Predicted: at tā han, uppå ärhällen kal-
Original: at tā han, uppå ärhällen kal-

at tā han, uppå ärhällen kal-

192# Predicted: telse af Inspektoren Livijn, wa-
Original: telse af Inspektoren Livijn, wa-

telse af Inspektoren Livijn, wa-

193# Predicted: rit til Spinhuset at underrätta
Original: rit til Spinhuset at underrätta

rit til Spinhuset at underrätta

44321# Predicted: bestridt alla honom äliggande Presteliga
Original: bestridt alla honom äliggande Presterliga

bestridt alla honom äliggande Presterliga

44322# Predicted: görenal till Consistorii och Församlingens
Original: göromål till Consistorii och Församlingens

göromål till Consistorii och Församlingens

44323# Predicted: ofullkomliga nöje, under iakttagande
Original: fullkomliga nöje, under iakttagande

fullkomliga nöje, under iakttagande

4509# Predicted: van velkomušen perusteena oleva kont-
Original: van velkomušen perusteena oleva kont-

van velkomušen perusteena oleva kont-

4510# Predicted: tokuranti oli oikein uloskirjoitettu Fredriks
Original: tokuranti oli oikein uloskirjoitettu Fredriks-

tokuranti oli oikein uloskirjoitettu Fredriks-

4511# Predicted: sonin reskontrista vuosilta 1887, 1888, 1889 ja
Original: sonin reskontrista vuosilta 1887, 1888, 1889 ja

sonin reskontrista vuosilta 1887, 1888, 1889 ja

Kuva: Ilkka Jokipii

Aineisto- infrastruktuurin vahvistaminen

- Infrastruktuurien kehittämishankkeet tärkeitä → kehittämistä ei kuitenkaan voi tehdä tyhjiössä
- Infrastruktuurikehitys tarvitsee **tiiviiin kytköksen tutkimukseen**
- Aineistoinfrastruktuurin vahvistaminen voi olla yksi tutkimushankkeen tavoitteista



**KOHTALONA
NEUVOSTOLIITTO**

6 000
uutta
nimeä

31 000
henkilöä

Uusi kortistolöytö

KANSALLISARKISTO

Aineistoinfrastruktuurien kehitys edellyttää yhteistyötä

Yhteistyön tavoitteena aineiston käytettävyyden parantaminen ja tutkimuksen edellytysten kehittäminen pitkällä aikavälillä





KANSALLISARKISTO

www.kansallisarkisto.fi



@kansallisarkisto



@kansallisarkist