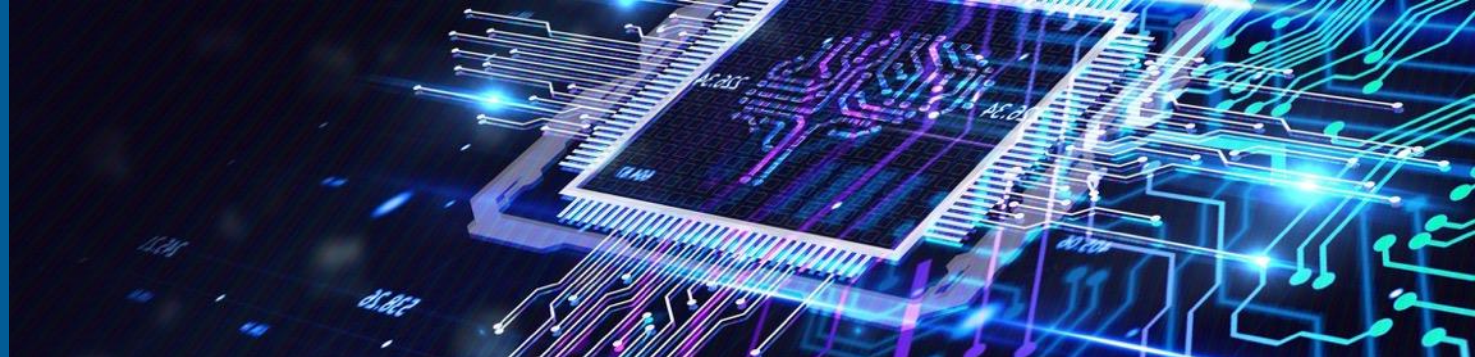




CSC

ICT Solutions for  
Brilliant Minds

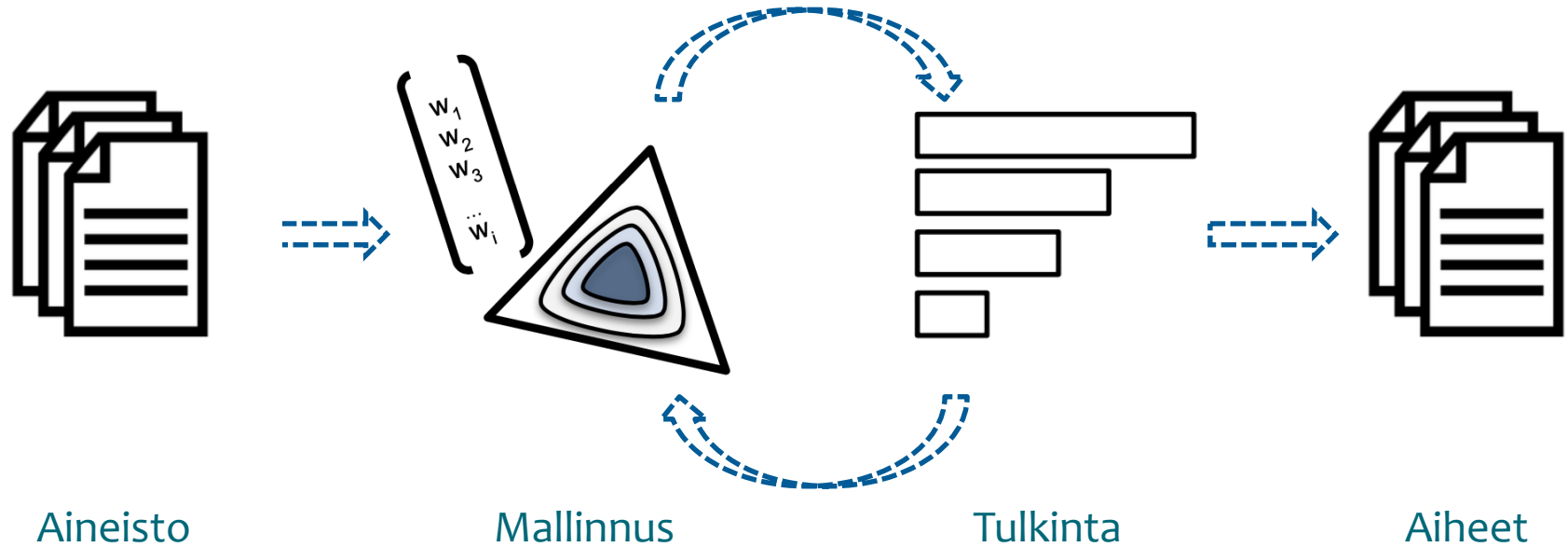


# Hankkeiden aihemallinnus tutkimustietovarannossa

KOTA-seminaari 2023 – Katja Mankinen & Joonas Kesäniemi

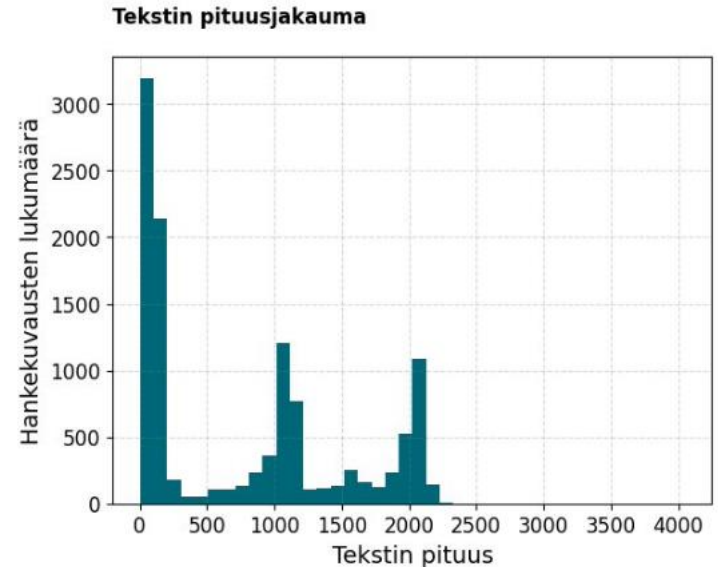


# Aihemallinnus menetelmänä



# Hankedatan aihemallinnus 2023

- 11431 rahoituspäätöstä vuosilta 2007-2024 (2021 ~7000) TTV:n rajapinnan kautta
- BERTopic-aihemallinnus menetelmä
  - Monikielisyys sisäänrakennettuna
  - 8002 koulutukseen ja 3429 testaukseen
- Mallin päivitys, eli mallinnuksen tulosten tulkinta vielä kesken
  - Edellisessä mallissa 94 aihetta
  - Portaalissa julkaistavien aiheiden ylläpitomalli



# Demo

- [Tiedejatutkimus.fi](https://tiedejatutkimus.fi)
  - Vuoden 2021 mukaiseen mallinnukseen perustuvat aiheet
  - Uuden sisällön automaattinen luokittelu mallinnettuihin aiheisiin
- [KOTA 2023 demo](#)

# Jatkokehitysideoita

- Aiheiden evoluution analysointi
- Mallinnukseen käytettävän tietopohjan automaattinen laajentaminen muilla koneoppimismenetelmillä
  - Noin 50% rahoituspäätöksistä sisältää tieteenalan
  - Noin 25% dokumenteista löytyy avainsanoja
  - Automaattinen asiasanoitus ja ontologian pohjalta päätellyt tieteenalat

# Lopuksi – eli kotinviemiset

- Aihemallinnuksen avulla on mahdollista tuottaa uutta dataa ilmiöpohjaisen analytiikan pohjaksi
  - Aihe/ilmiö on joustavasti määriteltävissä niin hyvässä kuin pahassa
  - Aihe generoitavissa myös esimerkiksi rahoituspäätöksille, joille ei ole asetettu tieteenalaa tai asiasanoja
  - Menetelmän hyödyntämiseen liittyvät rajoitteet tiedostettava ja kommunikoitava
- Menetelmään liittyvää osaamista kehitetään myös jatkossa osana tutkimustietovarantoa
  - Portaalin aihemallin päivitykset ja käyttöönoton laajentaminen
  - Uusien menetelmien evaluointi



Joonas Kesäniemi



[facebook.com/CSCfi](https://facebook.com/CSCfi)



[twitter.com/CSCfi](https://twitter.com/CSCfi)



[linkedin.com/company/csc--it-center-for-science](https://linkedin.com/company/csc--it-center-for-science)



[github.com/CSCfi](https://github.com/CSCfi)