



# Sprogteknologi.dk – Visioner, udfordringer og fremgang

November 2021



DIGITALISERINGSSTYRELSEN



# Agenda



Sprogteknologi.dk

---



Fremadrettet arbejde for sprogteknologi

---



Samarbejde med EU

---



Hvordan løfter vi dansk sprogteknologi

# Initiativet sprogteknologi.dk



- Fællesoffentligt samarbejde
- **Formål – at styrke udviklingen af løsninger inden for dansksproget kunstig intelligens**
- Målgrupper – sprogteknologiske virksomheder samt slutbrugerne i form af myndigheder, forskere, virksomheder og borgere
- Samarbejde med private og offentlige aktører

## Sprogressourcer

*I scope af sprogteknologi.dk*

### Sprogdata

Talekorpora  
Tekstkorpora  
Ordbøger  
Termbaser

### Sprogteknologisk supportværktøj

Dataindsamlingsværktøjer  
Anonymiseringsværktøjer  
Værktøjer automatisk opmærkning  
Værktøjer manuel opmærkning  
Modeltræningsværktøjer (...)

### Sprogteknologisk infrastrukturkomponent

Talegenkendelseskomponent  
Talesyntesekomponent  
Sprogforståelseskomponent  
Maskinoversættelseskomponent (...)

### Softwareløsning

Tale-til-tekst  
Tekst-til-tale  
Dialogsystem  
Maskinoversættelsesystem  
Læse-/skriveværktøj (...)

Dansk sprogteknologi er i udvikling

Analysebureau: Anerkendelse for angreb på Facebook

SPROGMODELLEN ÆLÆCTRA VIL FORBEDRE DANSK SPROGTEKNOLOGI PÅ EN KLIMAVENLIG MÅDE

Superalgoritme kortlægger og afslører yndlingsforetrukne  
Kommune: Dorte gennemgår samtaler og træner Al'en manuelt

et mere end

Her er robotten Bridget. Hun skal læse tv-nyheder op for bornholmerne  
Danish Gigaword

Danske hadbook



Ny skandinavisk NER model med 50...



DIGITALISERINGSSTYRELSEN

# »» Hvor ser vi udfordringerne for dansk sprogteknologi?



Dansk vil altid være et mindre udbredt sprog på verdensplan



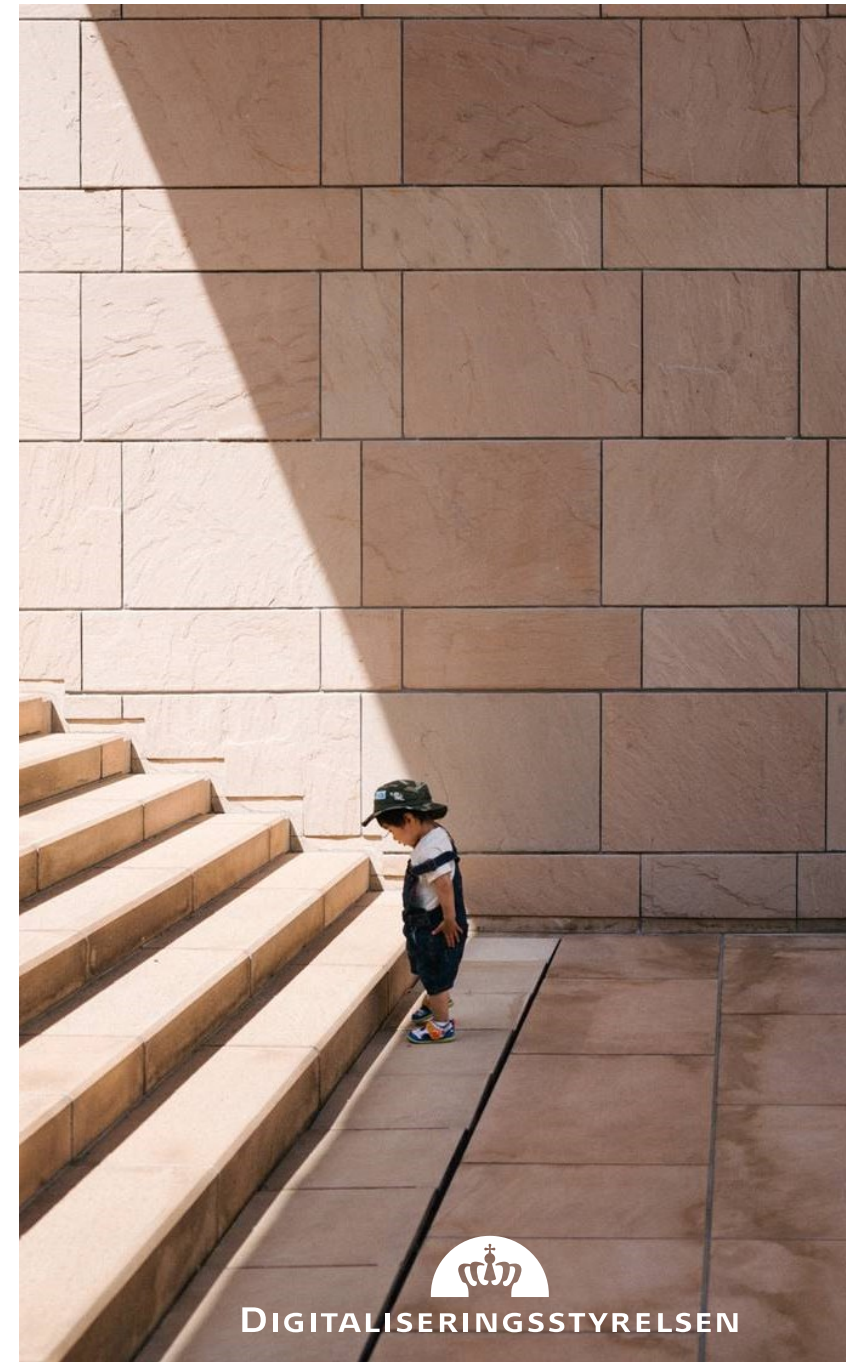
Behov for flere sproressourcer, særligt på lyd



Ressourcekrævende at udvikle eller ibrugtage dansk sprogteknologi

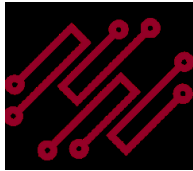


Vedvarende og løbende kvalitetssikring af eksisterende løsninger





# Arbejdsprogram 3.0



## Sprogteknologi.dk

Portal for den fælles danske sproressource

Over 120 ressourcer på portalen

- Datasæt
- NLP værktøjer
- NLP modeller



## Centralt Ord Register (COR)

Samarbejde med Dansk Sprognævn, Center for Sprogteknologi ved KU, Det Danske Sprog- og Litteraturselskab



## Taleteknologi

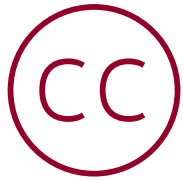
Et større datasæt (talekorpus med lyd, tekst og annotationer i høj kvalitet og frit tilgængeligt).

Analyse af mulighederne for samarbejde med interessenter om at udarbejde datasæt baseret på lydbogsmateriale



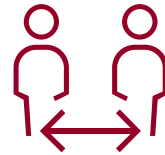


# Arbejdsprogram 3.0



## Standarder

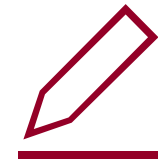
- For licensvalg
- Opmærkning
- Strukturering af data
  - Metadata
  - Annotation



## Forbinde miljøet

Flere events og begivenheder

Deltagelse i følgegrupper på  
sprogteknologiske projekter



## Guide til samtykke

Gøre det nemmere for udviklere  
og anvendere at indsamle den  
fornødne data in-house





# Europæiske arbejde

## Deltagelse i forskellige initiativer

- European Language Equality (ELE)
- European Language Resource Coordination (ELRC)
- European Language Grid (ELG)

## Arbejder for at dansk bliver tilgodeset i EU

- Digital Europe Programme 2022 - 2023 → Language Data Spaces
- Løbende inputs til CEF platformen, særligt eTranslation
- Følger med i EU projekter, der indeholder AI relateret sprogteknologi



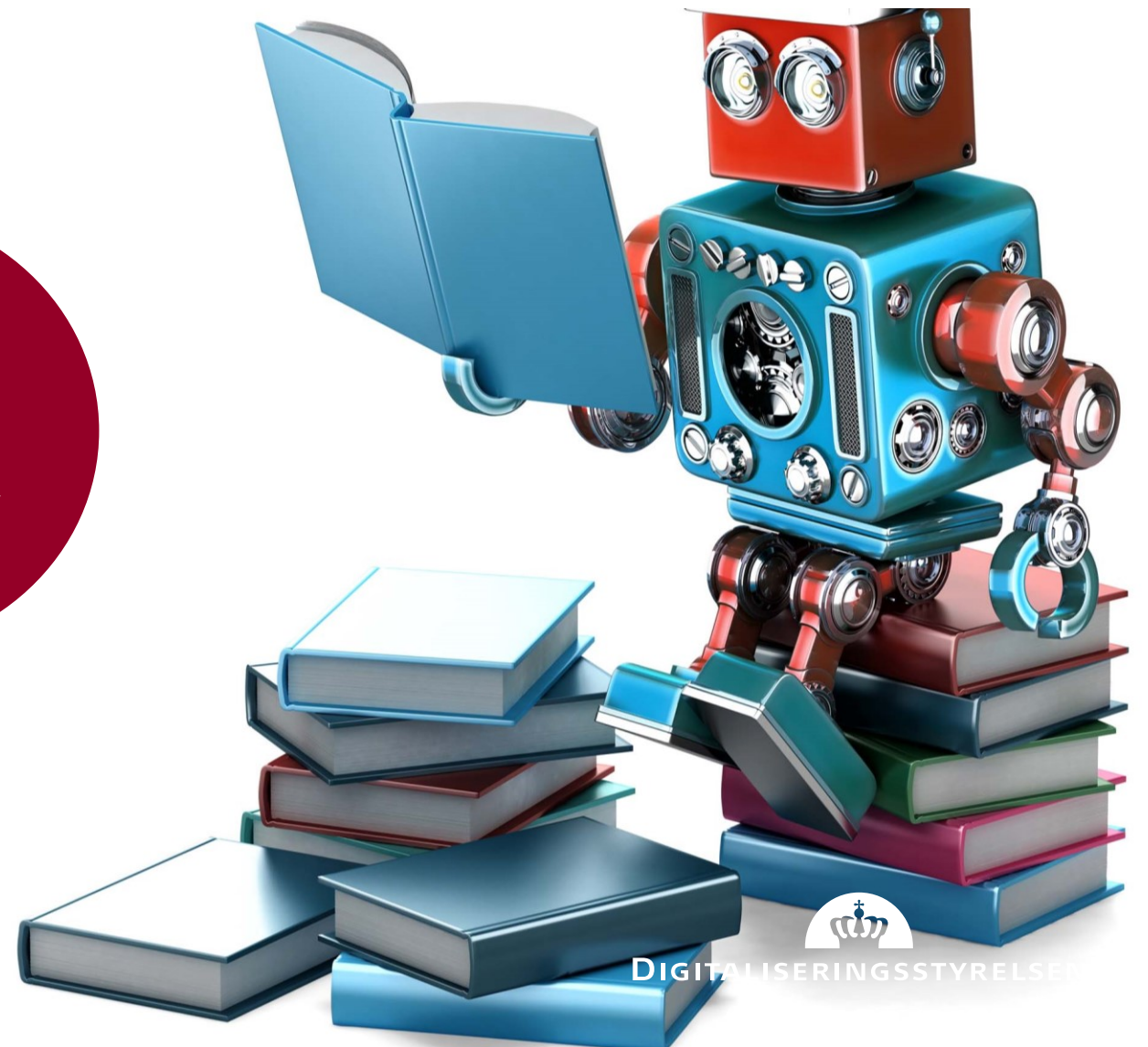


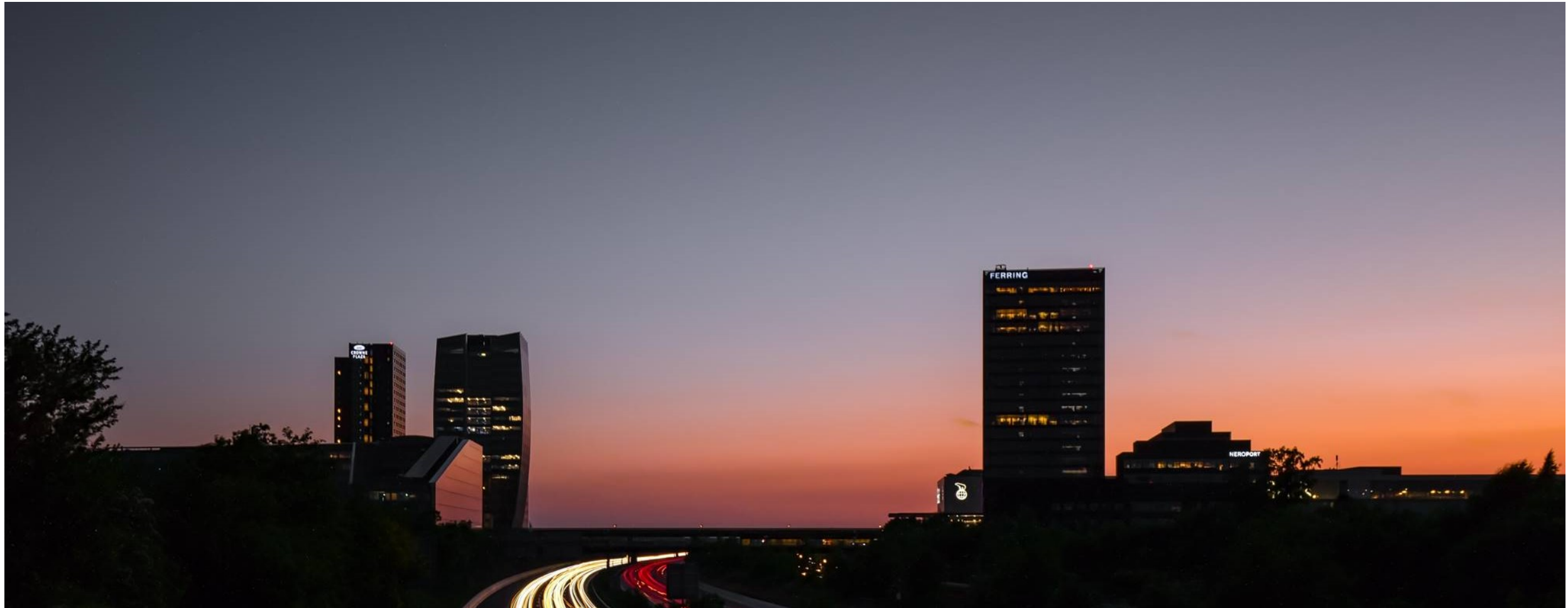
# Hvordan løfter vi dansk sprogteknologi?

Del jeres  
sprogressourcer

Tænk over  
fremtidig  
indsamling af  
sprogressourcer

Open-source  
så meget som  
muligt





Christian Plaschke, kontorchef  
[info@sprogteknologi.dk](mailto:info@sprogteknologi.dk)



DIGITALISERINGSSTYRELSEN