

Danish Conversational and Read-aloud speech dataset

WEBINAR 14. APRIL 2026

CQRa1

Hele Danmarks tale-til-tekst model



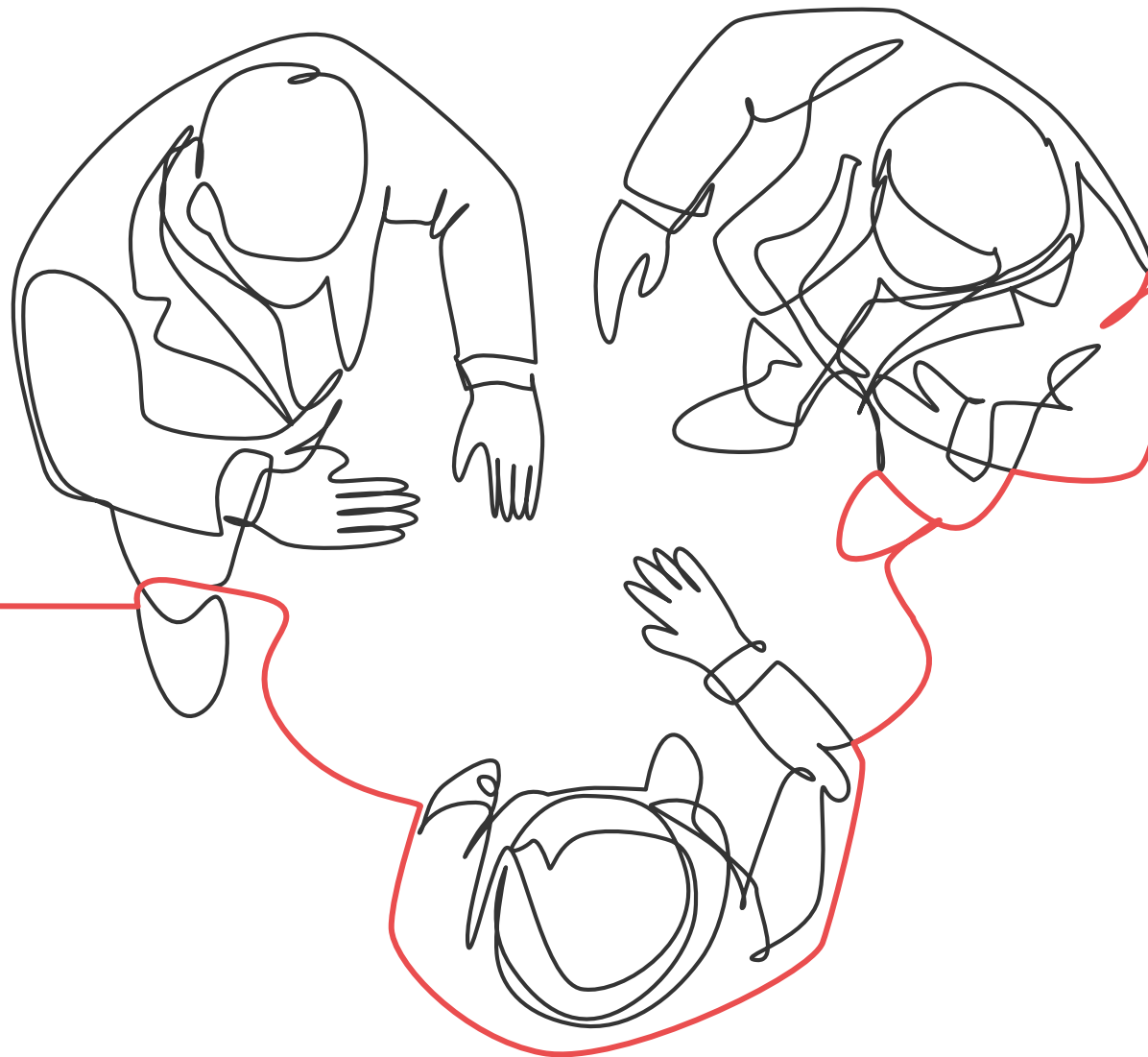
PROJEKTVISION

At opbygge et dansk **taledatasæt**, der løfter dansk tale teknologi og hjælper med at skabe nye virksomheder, arbejdspladser og vækst i taleteknologibranchen og i samfundet generelt



Hvilke udfordringer adresserer **CoRal**

- Digital ligestilling
- Demokrati
- Digital suverænitet



Hvor svært
kan det
være



Dataindsamling



Sådan kan en præst på et bibliotek forbedre din Alexa

Talegenkendelse på dansk halter efter andre sprog. Et nyt projekt til 22 millioner kroner skal skabe et datagrundlag, der stiller gratis til rådighed for virksomheder.



Den intelligente talestemme

Dansk og svensk mange dialekter udfordrer computerens talegenkendelse



Brit og Kennet får AI til at forstå bornholmsk



05.53

TVSYD.DK

tvsv D-PLAY

SE MERE

FORSIDE NYHEDER LIVE UDFORSK PROGRAMOVERSIGT

Seneste 19.30-indslag

Dialekter bliver gemt og skaber grobund for fremtidens teknologi

CoRal Datasæt



Data volume
> 710 timer
(heraf ~ 167 timer samtale)



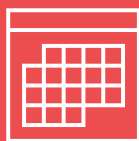
Antal samples
> 468.000 ytringer
med tilhørende
transskribering



Antal indtalere
> 1000 personer
mellem 11-97 år



Lokationer
> 70 forskellige

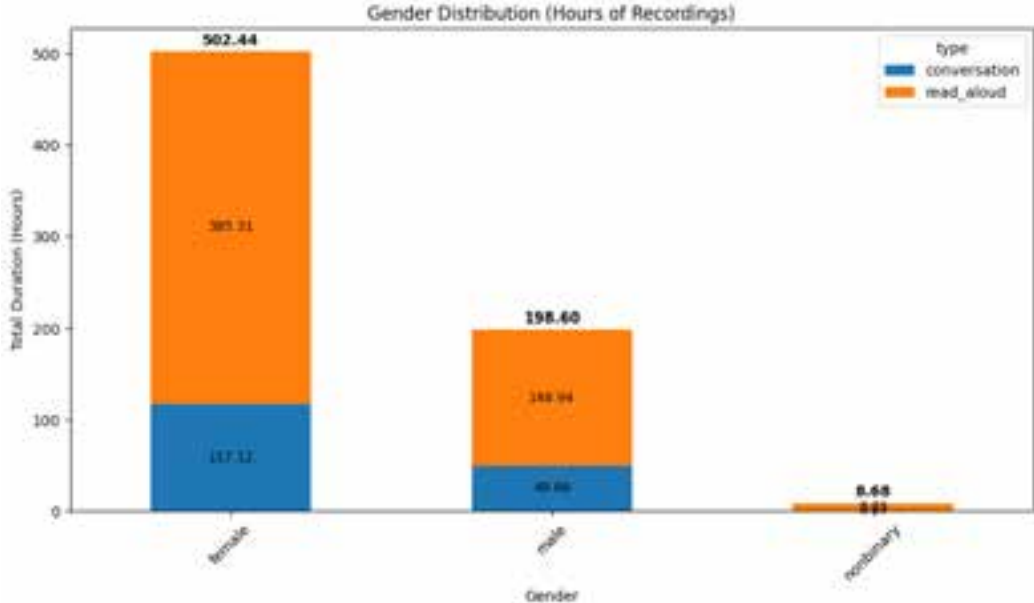
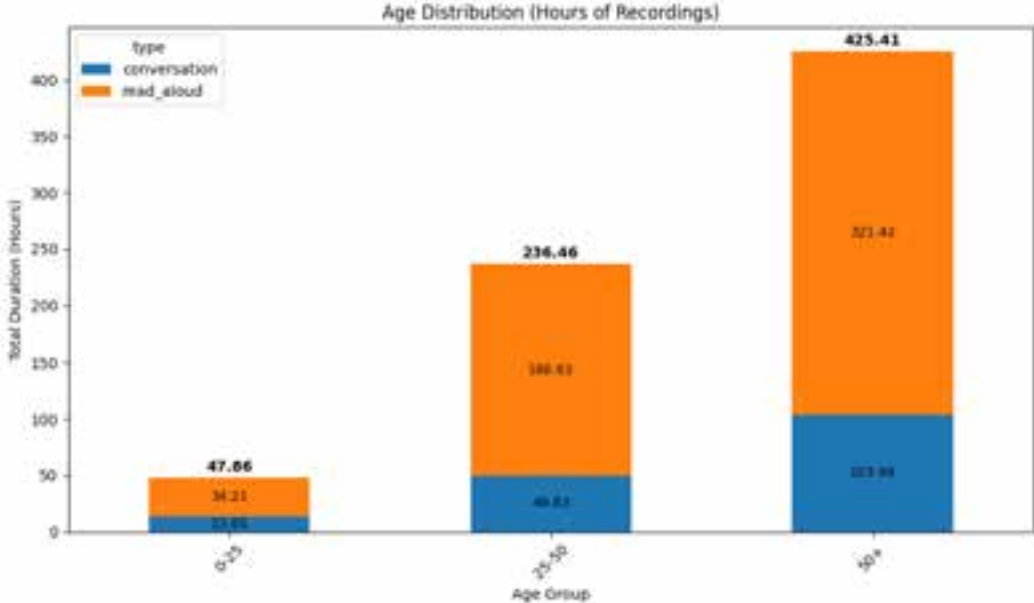
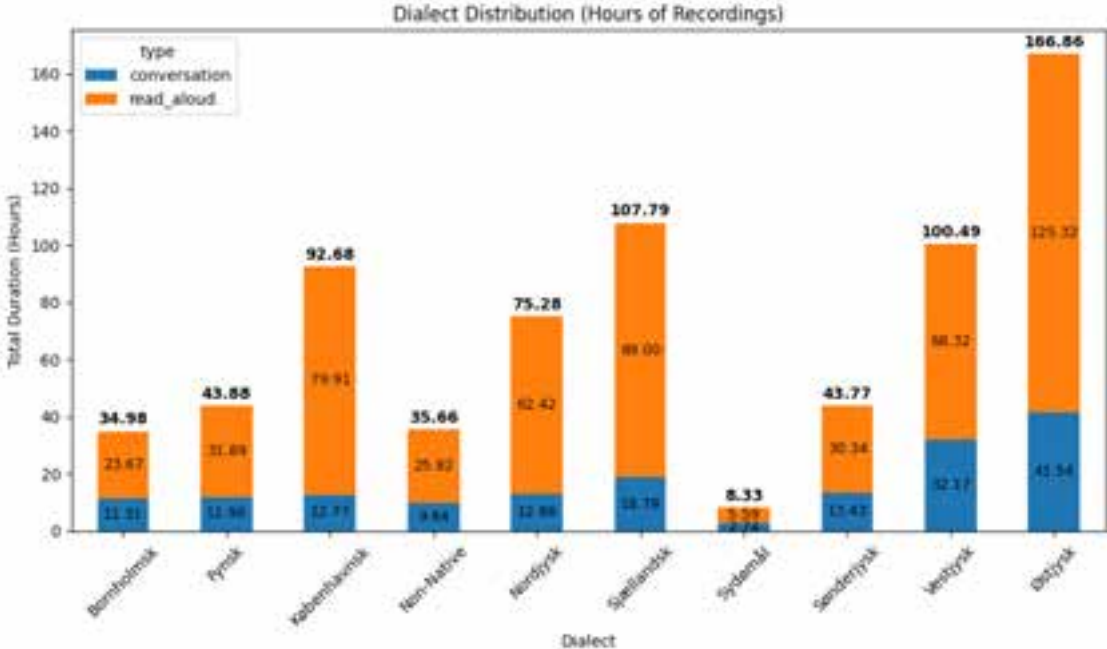


Optageperiode
august 2023
- maj 2025



100% dækning
af alle dialektområder

CoRaI Datasæt 3.0



CoRal Datasæt 3.0

Datasæt

<https://huggingface.co/datasets/CoRal-project/coral-v3>

<https://huggingface.co/datasets/CoRal-project/coral-tts>

Øvrige informationer

Transskriberingsguide, rapporter, værktøjer etc.

<https://alexandra.dk/coral>

Udgivet marts 2026

Licensinfo

liability for the Licensor whatsoever.

4. Supplementing Attachment A to the Conditions, You additionally agree not to use the licensed Material or the Adapted Material:

- (a) To impersonate any person or entity or otherwise misrepresent its affiliation with a person or entity, including (but not limited to) the creation of synthetic speech (i) emulating a specific natural person or (ii) which is found to be similar to a specific natural person's speech (regardless of whether the identity of such natural person is known at the time of the creation); or
- (b) To detect or infer aspects and/or features of an identity of any natural persons, such as name, family name, address, gender, sexual orientation, race, religion, age, location (at any geographical level), skin color, society or political affiliations, employment status and/or employment history, and health and medical conditions.



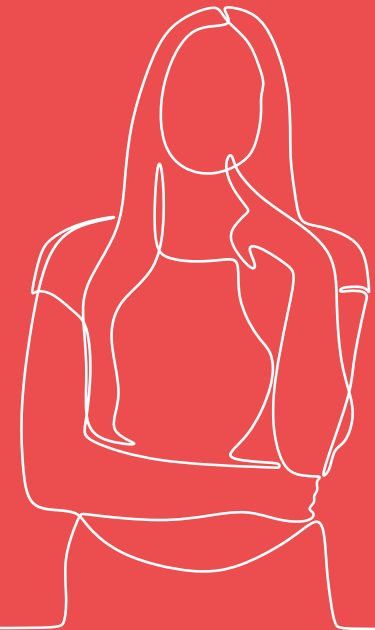
mmmm (noget er lækkert, smager godt)	mm
mhm mhm (nej)	mhmh
mhm (aha/jaså)	mhm
/hæ/ bruges i f.eks. /o hæ/ (å-hå)	hæ
/hu/	hu
/aha/	aha
/au/ (pirat-agtig lyd)	aim
/ær/ (frastødeslyd)	ad
/ærr/ /ærr/ (frastødeslyd med r sidst)	adr
/o/	oh
/o x/ /o y/ /o h/ eller lignende	ohh
/o k/ som i 'ork ja'	ork
/uf/	uf
/uhæ/	uha
/æ/	schy
no (når noget er nuttet)	Nåå
no (når man forstår; aha)	Nårh
no (som i 'nå, for Seren')	Nåh
/rja æ/ /njæ æ/ (udtryk for tvivl og/eller forbehold)	Nja
wu (overraskelse/overvældelse - a la 'uh')	Wuh

Erfaringer fra dataindsamling

- Jura og GDPR
- Forbered de rigtige værktøjer og test dem med brugere
- Rekruttering er svært !
- Visse demografiske grupper er specielt svære (unge og accenter)
- Fokuser på kvalitet og kontrollér den løbende

Nøgletal

- Indsamling 2-3 time forbrugt pr indtalt time ekskl. koordinering
- Transskribering 8-10 timer pr. indtalt time
- Tab på ca. 20% pga. kvalitet og personlig information (PI)





Fra: [REDACTED]

Dato: lørdag, 21. oktober 2023 kl. 09.38

Til: Alexandra Institutet <ALEXANDRA@alexandra.dk>

Emne: synnejysk

æ ka snak synnejysk å det ka min mand å.

æ er 76 år å min mand æ 81 år.

vi ka deltæ hele wonsda å torsda ærremidda

hjelsen fro

Ingerlise og Frede Jensen

Talegenkendelse

- Speech to Text (STT)



Modeltræning



CQRA1

Hele Danmarks tale-til-tekst model

Modeller er
ikke ens

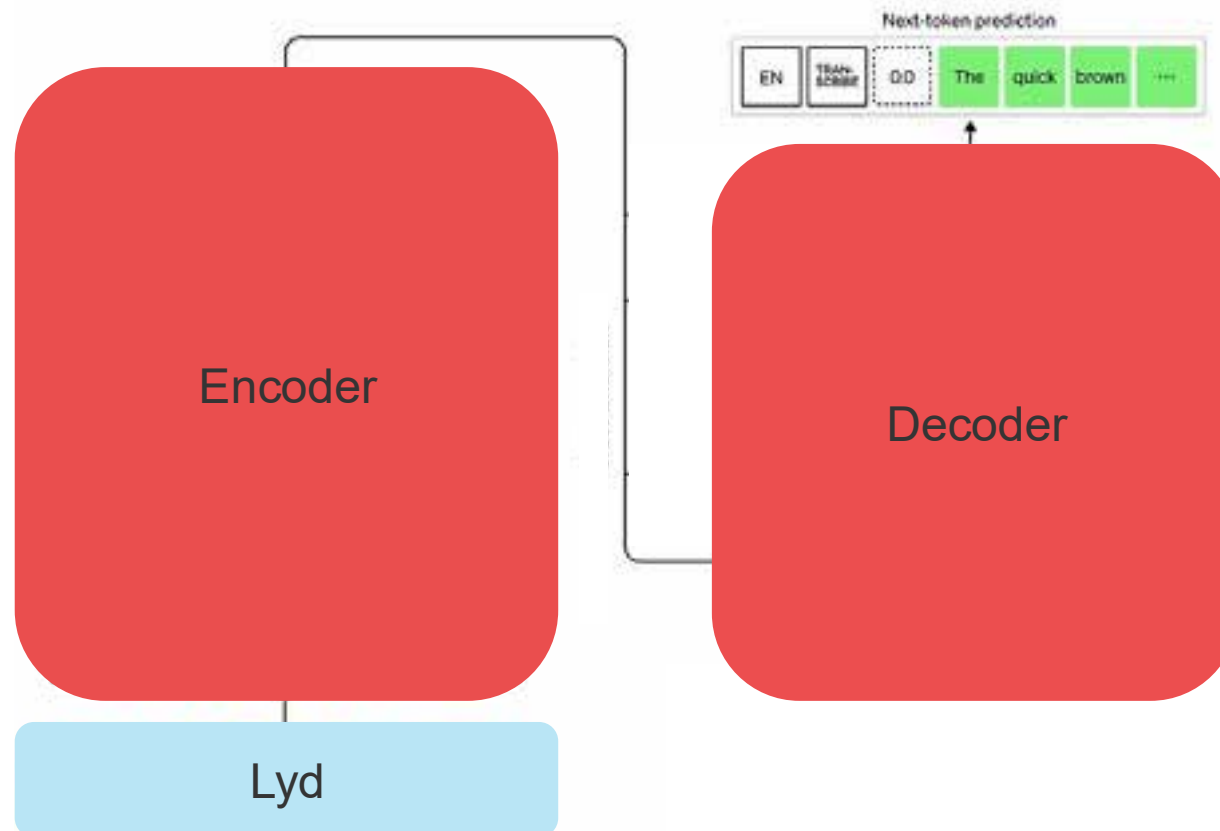


Whisper (2023)



Robust Speech Recognition via Large-Scale Weak Supervision

Whisper (2023)



Whisper (2023)



Robust Speech Recognition via Large-Scale Weak Supervision

Pros

- Fungerer supergodt
- God support i mange applikationer

Cons

- Hallucinerer ofte
- Ret stor model, så kan være langsom

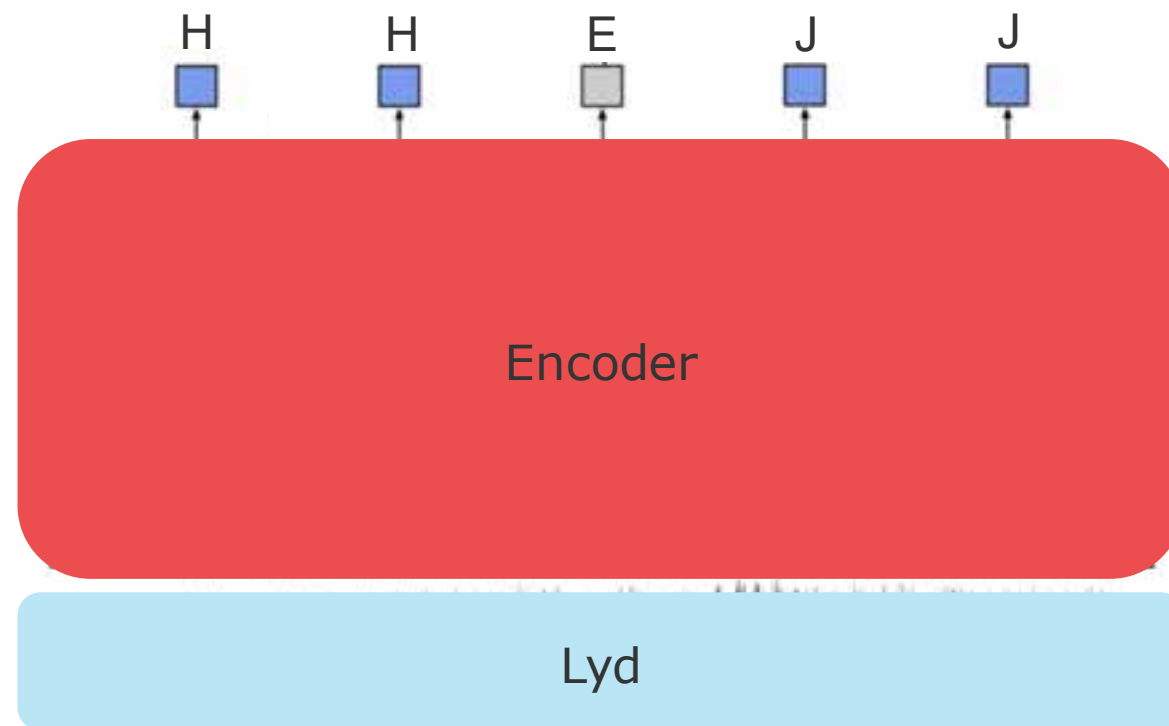
Wav2vec 2.0 (2020)



**wav2vec 2.0: A Framework for Self-Supervised
Learning of Speech Representations**

Wav2vec 2.0 (2020)

**wav2vec 2.0: A Framework for Self-Supervised
Learning of Speech Representations**



Whisper (2023)



wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations

Pros

- Kan ikke hallucinere
- Lille model (5x mindre end Whisper), så kan køre hurtigt

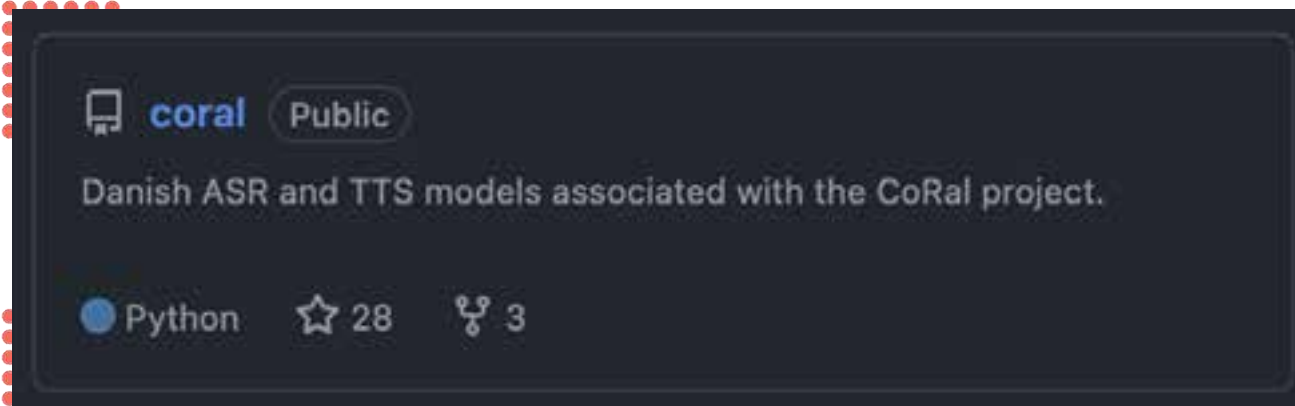
Cons

- Fungerer typisk ikke ligeså godt som Whisper
- Mindre support i applikationer

Model- trænings- framework

CoRa1

Hele Danmarks tale-til-tekst model



A screenshot of a GitHub repository page for 'coral'. The repository is public and contains Danish ASR and TTS models associated with the CoRa1 project. It is written in Python and has 28 stars and 3 forks.

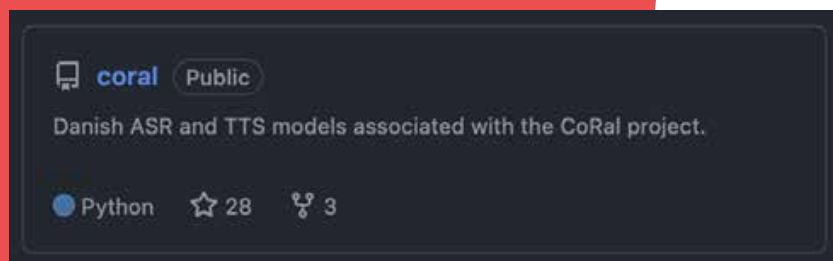
coral Public

Danish ASR and TTS models associated with the CoRa1 project.

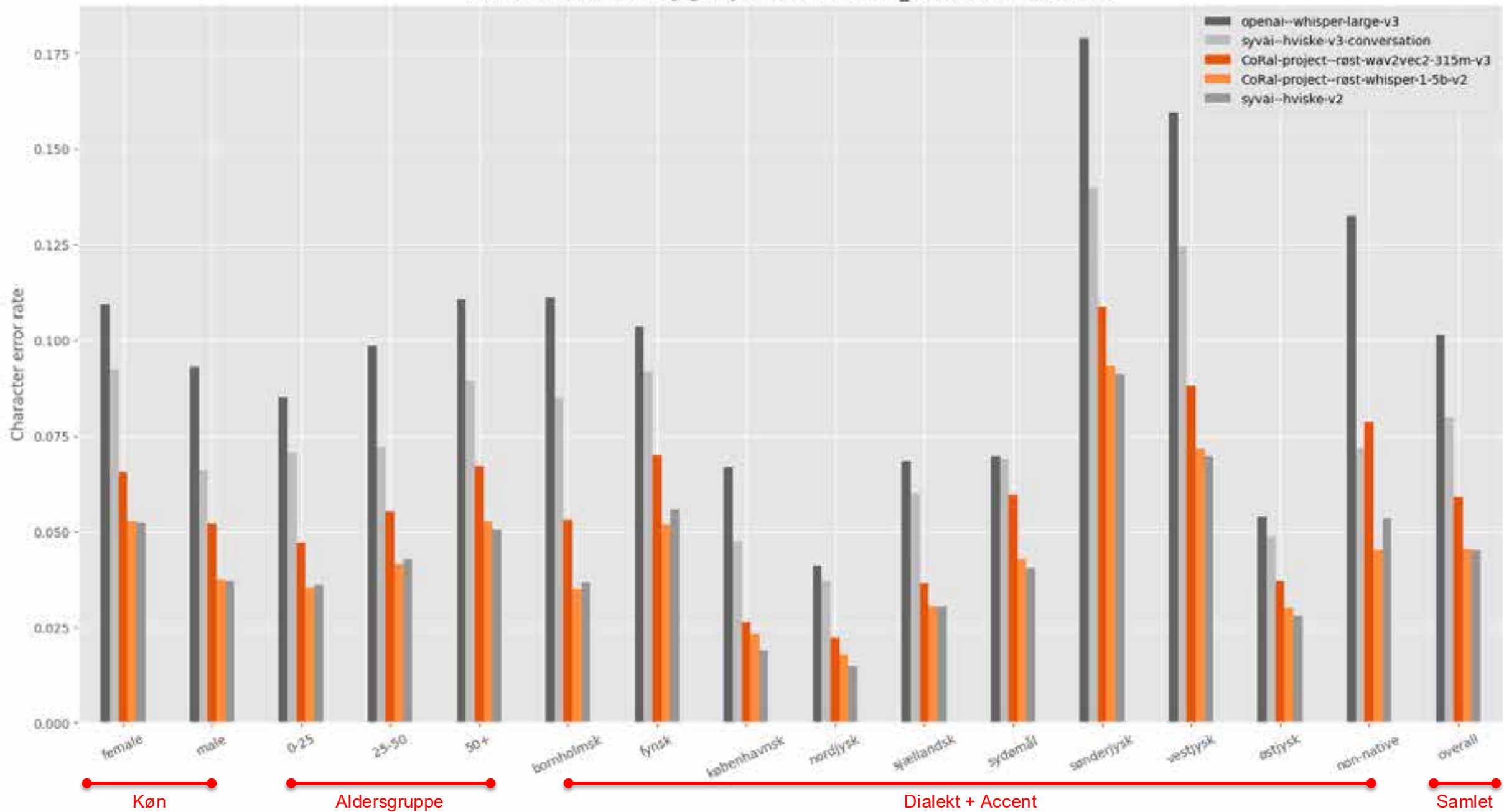
Python 28 3

Model- trænings- framework

- Kan bruges til at (videre-)træne nye modeller
- Gode gennemtestede standardopsætninger, men alt kan konfigureres
- Understøtter multi-GPU træning
- Tilføjer automatisk baggrundsstøj, og forvrænger lyden nogle gange, for at gøre modellerne mere robuste
- Anekdotisk har det vist, at dette har hjulpet en del med robusthed i virkelige støjende scenarier

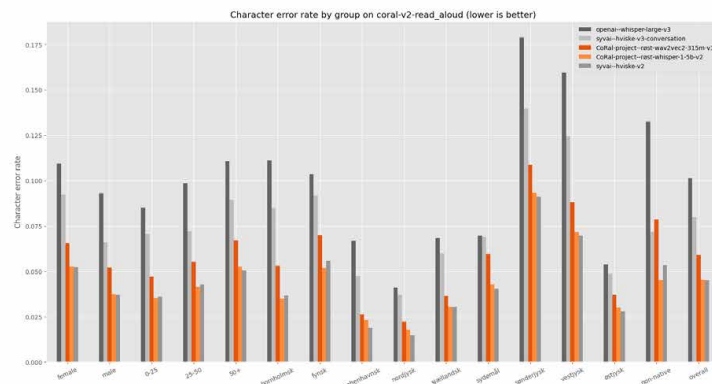


Character error rate by group on coral-v2-read_aloud (lower is better)

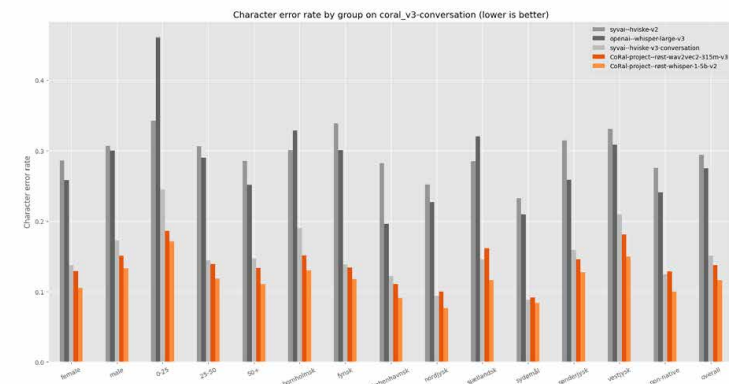


Model- resultater

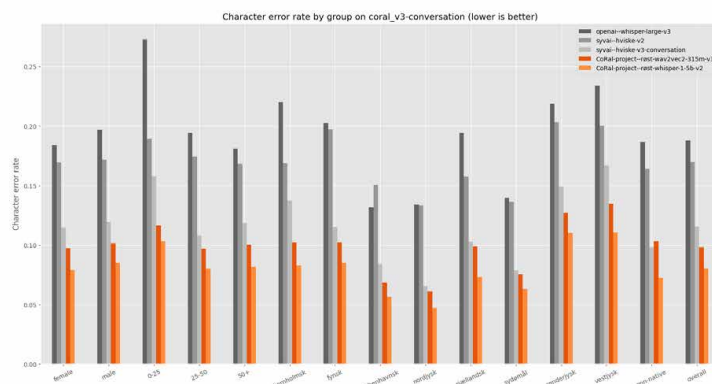
Samtaleperformance



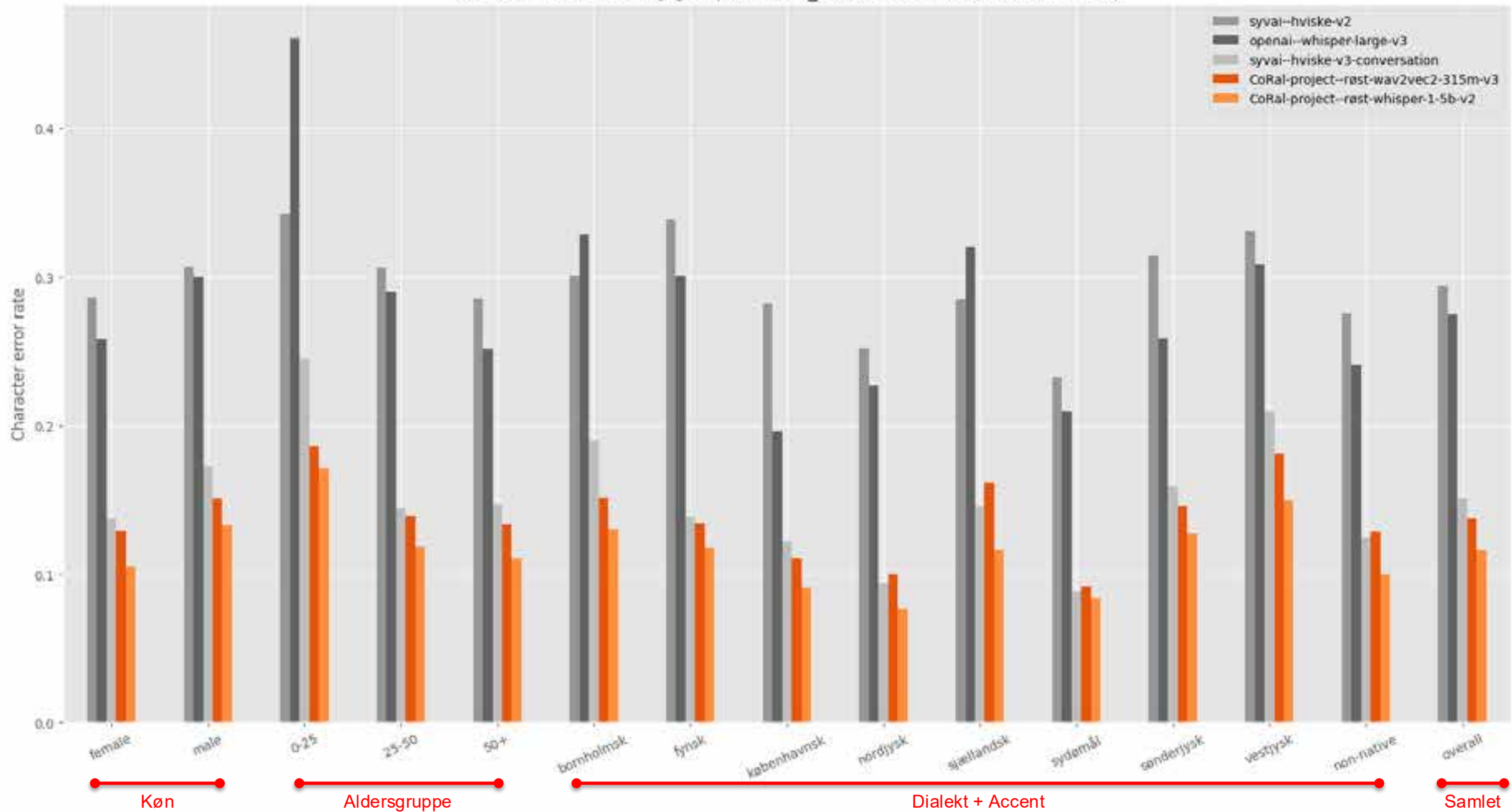
Oplæsningsperformance



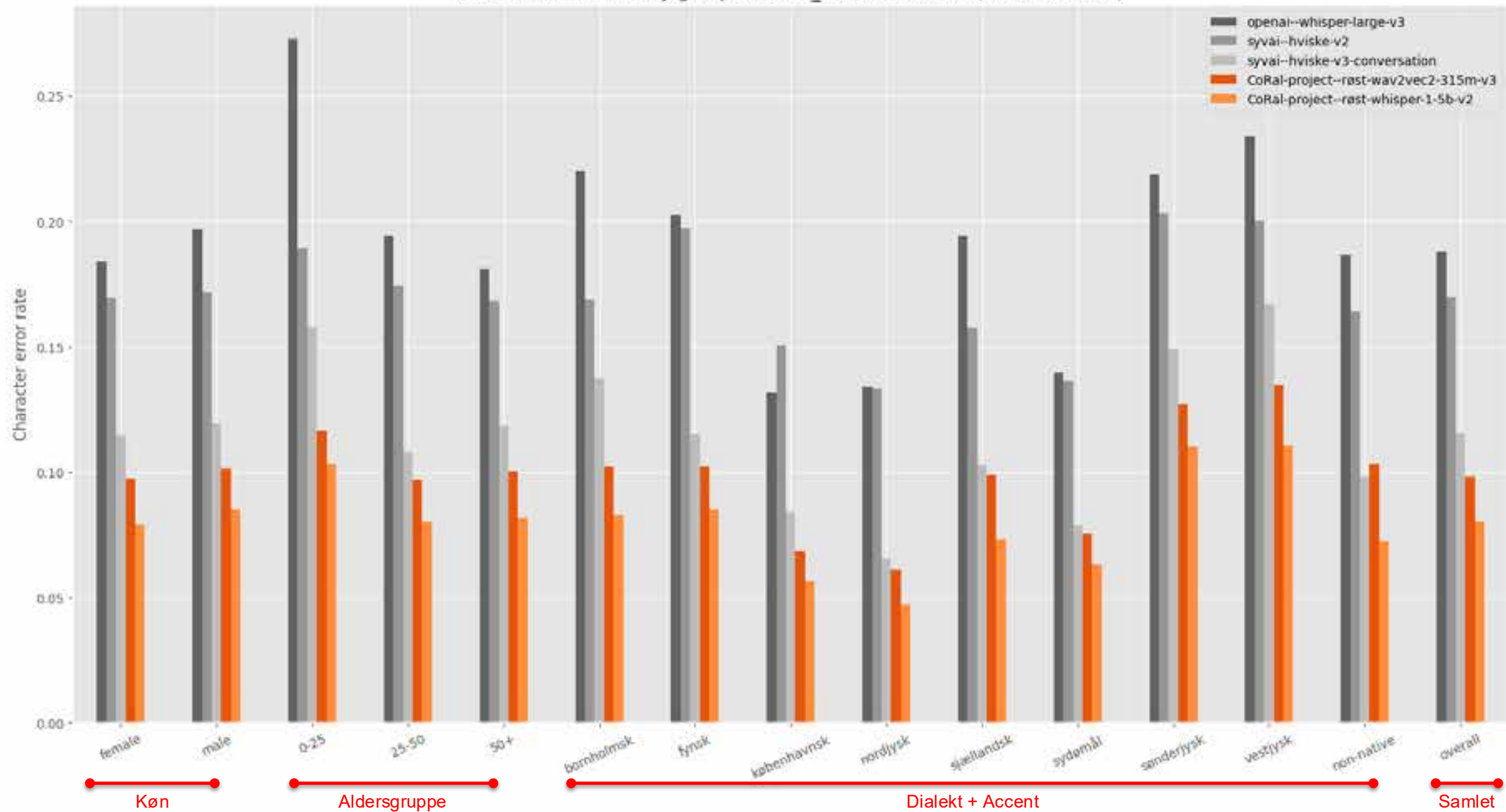
Samlet performance



Character error rate by group on coral_v3-conversation (lower is better)

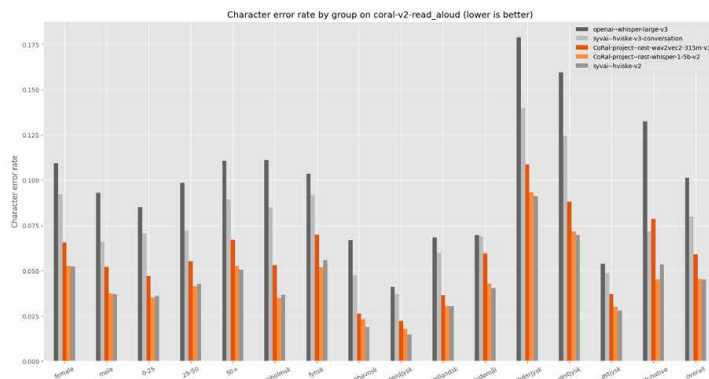


Character error rate by group on coral_v3-conversation (lower is better)

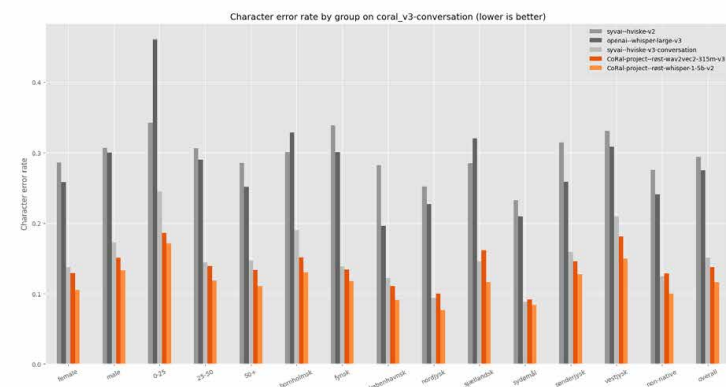


Model- resultater

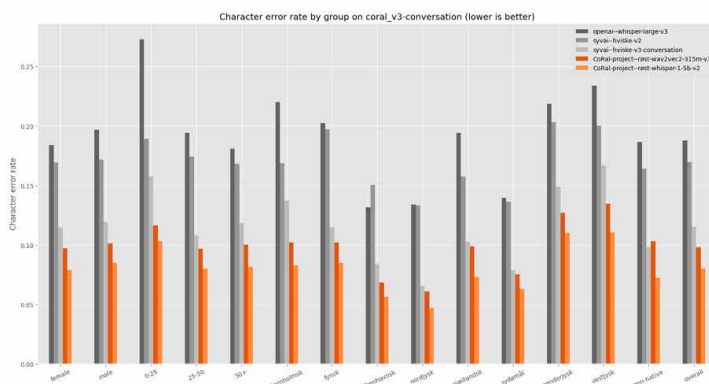
Samtaleperformance



Oplæsningsperformance



Samlet performance



Talesyntese

- Text to Speech (TTS)



Udfordringer - TTS

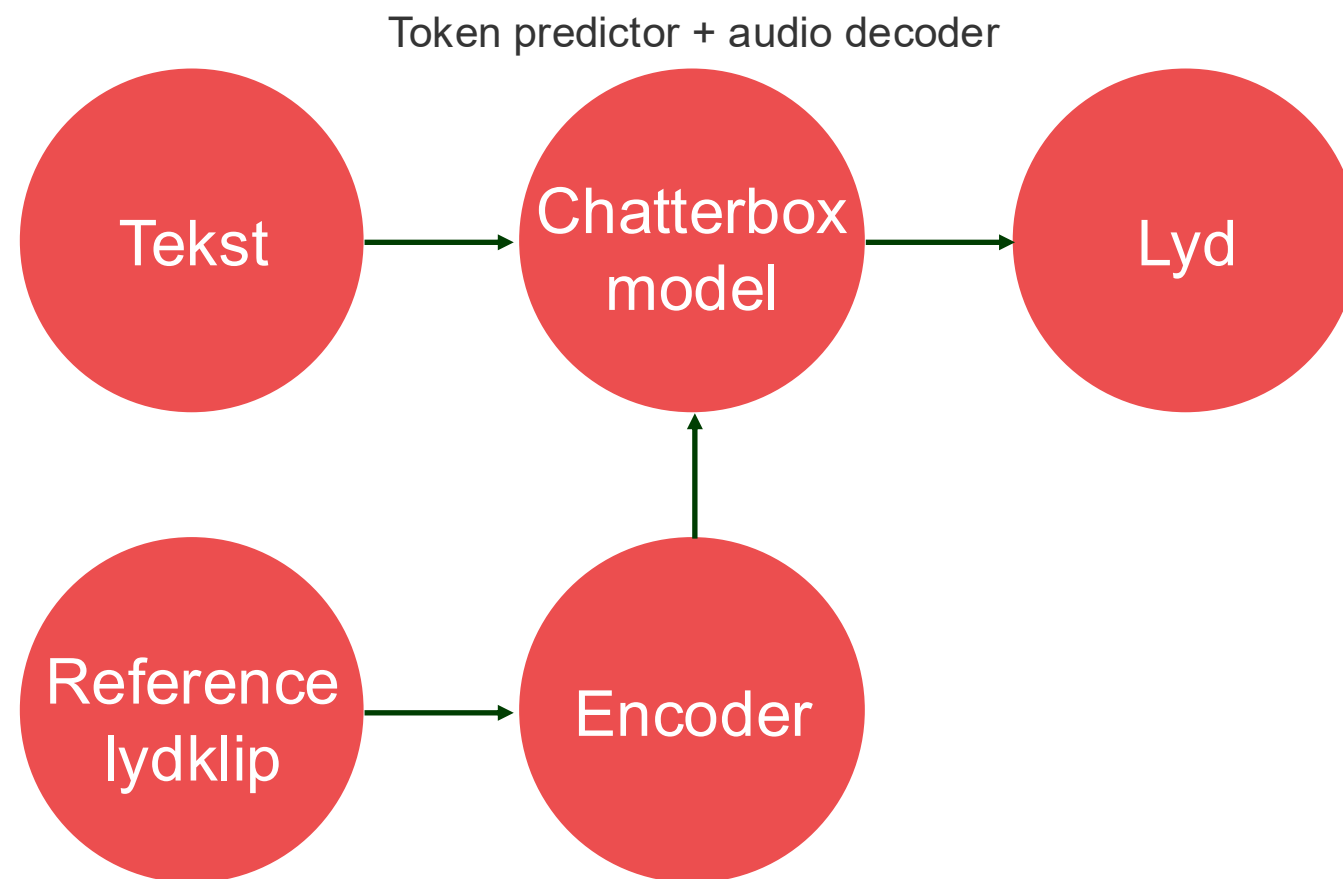
- Mange nuancer som er rigtig svære at efterligne
 - Tone, lyde, rytmer, udtale, fremmedord, forkortelser, osv.
 - Kræver store mængder data
- Mangel på eksisterende, moderne modeller som tilbyder træningsframework

Chatterbox (2025)



- Findes i 3 versioner
 - Original
 - Multilingual
 - Turbo

Chatterbox (2025)



Røst 500 M

Pros

- Præ-trænet på dansk
- Større model giver naturligt bedre performance

Cons

- Langsom inferens

Røst 350 M

Pros

- Meget hurtigere inferens
- Har god base-performance

Cons

- Ikke præ-trænet på dansk
- Kan hallucinere/have svært ved ukendte ord

Træningsnoter

- Trænet på over 2000 timers dansk tale
 - Coral-TTS, nst-da, ft-speech og nota
- Voice cloning:
 - Reference embedding funde ved anden sample fra samme speaker ID
- Træningsframework bliver udgivet snarest

Træningsnoter

- Trænet på over 2000 timers dansk tale
 - Coral-TTS, nst-da, ft-speech og nota
- Voice cloning:
 - Reference embedding funde ved anden sample fra samme speaker ID
- Træningsframework ~~bliver udgivet snarest~~
ER UDGIVET

Evaluering

- Evalueret ved brug af Mean Opinion Score (MOS)

Resultater

Model	Score
Røst 350M	4.01
Røst 500M	4.28
Azure	2.91

MOS	Beskrivelse
1.00 – Dårlig	Talen lyder helt unaturlig og kunstig. Den syntetiske stemme er meget tydelig og så generende, at den kan være svær at lytte til.
2.00 – Ringe	Talen lyder for det meste unaturlig. Den syntetiske kvalitet er tydelig og opleves som generende, men den er stadig til at holde ud.
3.00 – Middel	Talen fremstår både naturlig og unaturlig i nogenlunde lige grad. Det er tydeligt, at stemmen er syntetisk, og det kan virke en smule forstyrrende.
4.00 – God	Talen lyder overvejende naturlig. Man kan godt høre, at det er syntetisk tale, men det er kun svagt mærkbart og ikke generende.
5.00 – Fremragende	Talen lyder helt naturlig og kan ikke skelnes fra en rigtig menneskestemme. Der er ingen hørbare tegn på, at det er syntetisk tale.

Hvad bruger
virksomhederne
CoRal til?



CoRal relevans og bidrag

- Virksomheder ser et stort potentiale i projektets anvendelsesmuligheder og bruger det som et reelt alternativ til engelsksprogede løsninger.
- Udfylder et vigtigt hul i det danske tale teknologiske landskab med åbne og mangfoldige ressourcer.

Det er en unik ressource, som jeg tror, at f.eks. Microsoft måske ikke ligefrem kan ramme på kvaliteten, bare for det danske sprog isoleret set.

NRGI

Mere diversitet og alle de her ting, som CoRal prøver at hjælpe med ved at give et større datasæt, det giver mere til at træne med, og det er bare fantastisk.

Mediacatch

Virksomhedernes erfaringer med CoRals ressourcer

Unik ressource der forbedrer både egne og andres modeller

Diversiteten i data åbner op for flere muligheder:

Jeg ser en stor værdi og et kæmpe potentiale i det, fordi uanset hvordan man vender og drejer det, så har vi mange dialekter og accenter i Danmark. Så det betyder jo, at forskellige løsninger kan blive endnu mere udbredt på dansk.

N1

Vi har talt med vores leverandører og hvis deres [data] ikke er god nok til at forstå dansk, så har vi CoRal. Og de siger, at man nemt kan plugge det ind.

Børns Vilkår

Nice and clean på godt og ondt

God lyd kvalitet er vigtigt:

Det er nice and clean, hvor jeg før har brugt en masse tid på at rense data for støj eller justeret lyden, har jeg her ikke oplevet problemer med dataen.

Nvidia, oversat

Lydkvaliteten passer ikke altid med usecasen:

Data er urealistisk godt. Vi kunne godt bruge data med en dårlig mikrofon og folk der sidder på kontor.

syv.ai

Virksomhedernes ønsker til videreudvikling

Mere samtaledata – where the fun begins

- Hverdagssamtaler
- Samtaler mellem mennesker, der kender hinanden

God følelsesbetonet dialog, f.eks. en diskussion mellem venner, der planlægger en rejse eller et mere teknisk arbejds møde. Den type data, hvor ideer udveksles, hvor man kan høre forvirring, livlige diskussioner simpelthen, og ideelt set optaget med flere mikrofoner og forskellige lydkanaler.

Nvidia, oversat

Domænespecifikke datasæt

- Ønske om målrettet dataindsamling, herunder nichebegreber, fagsprog og branche-specifik terminologi

Vi ser en øget brug af [løsningen] i sundhedssektoren, specielt fra kommunernes side af, og der er nogle af de her tekniske formuleringer, der godt kan give problemer. Sådan specifikke nicher, fagtekniske nicher, der kunne det være interessant at have noget i en højere kvalitet.

Vitec MV

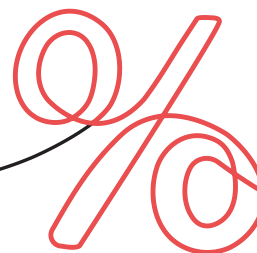


Hvordan kan I
bruge CoRal?



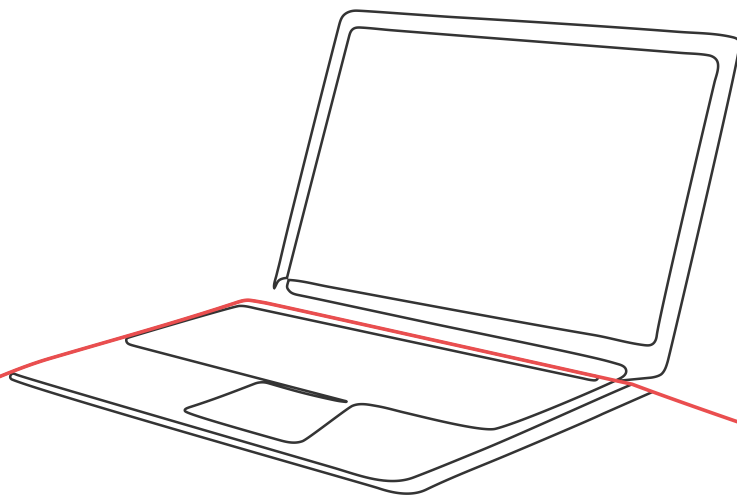
Brug CoRal datasættet til:

- at træne tale-til-tekst modeller til at forstå dansk
- at evaluere jeres eksisterende modeller og identificere forbedringsområder
 - Opfordre jeres leverandør til at anvende CoRal datasættet til netop at forbedre områderne indenfor dialekt, køn og alder
 - Når I skal udvælge en ny leverandør eller model



Brug RØST som model

- Open source modeller er i dag på niveau med kommercielle løsninger
- Bidrager til at sikre digital suverænitet og bred repræsentativitet på alder, køn og dialekt
- Vigtigt at I tænker hele arkitekturen i jeres løsning, hvor modellen kun er én ting



Kickstart jeres
taleteknologiske udvikling

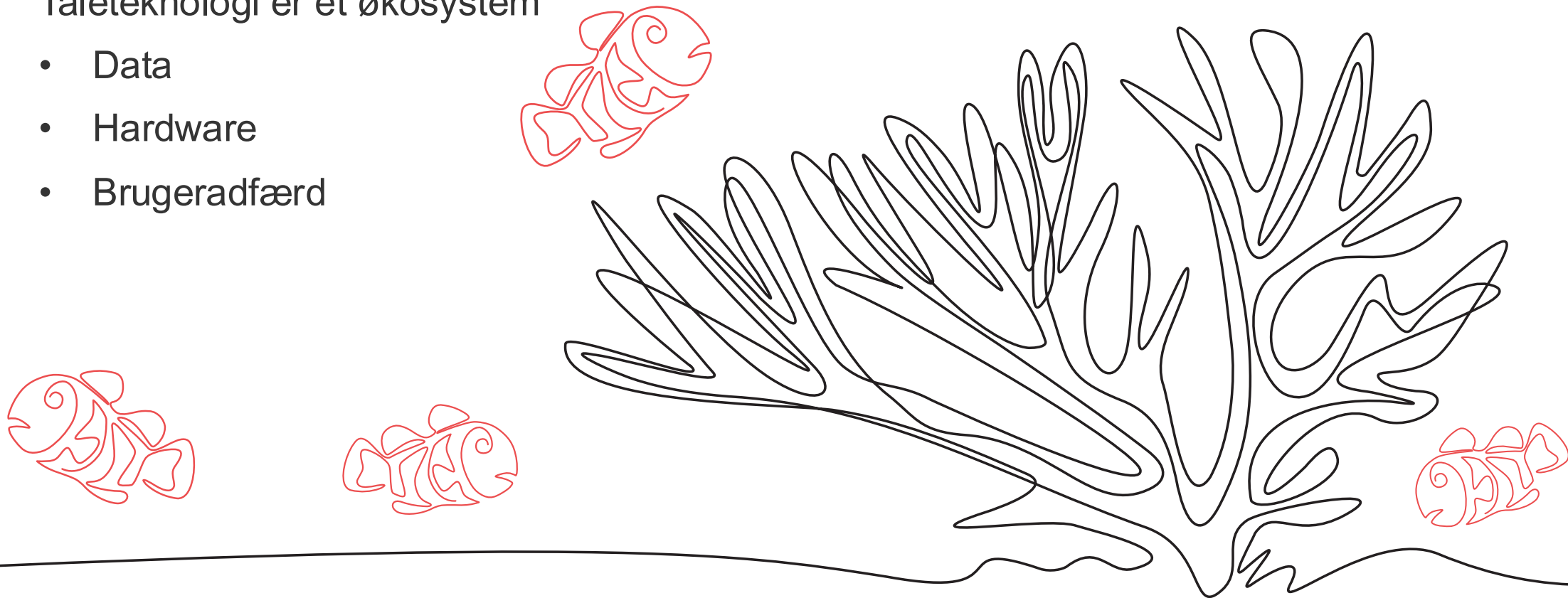
CoRa1

Hele Danmarks tale-til-tekst model

Taleteknologi er ligesom koralrevet

Taleteknologi er et økosystem

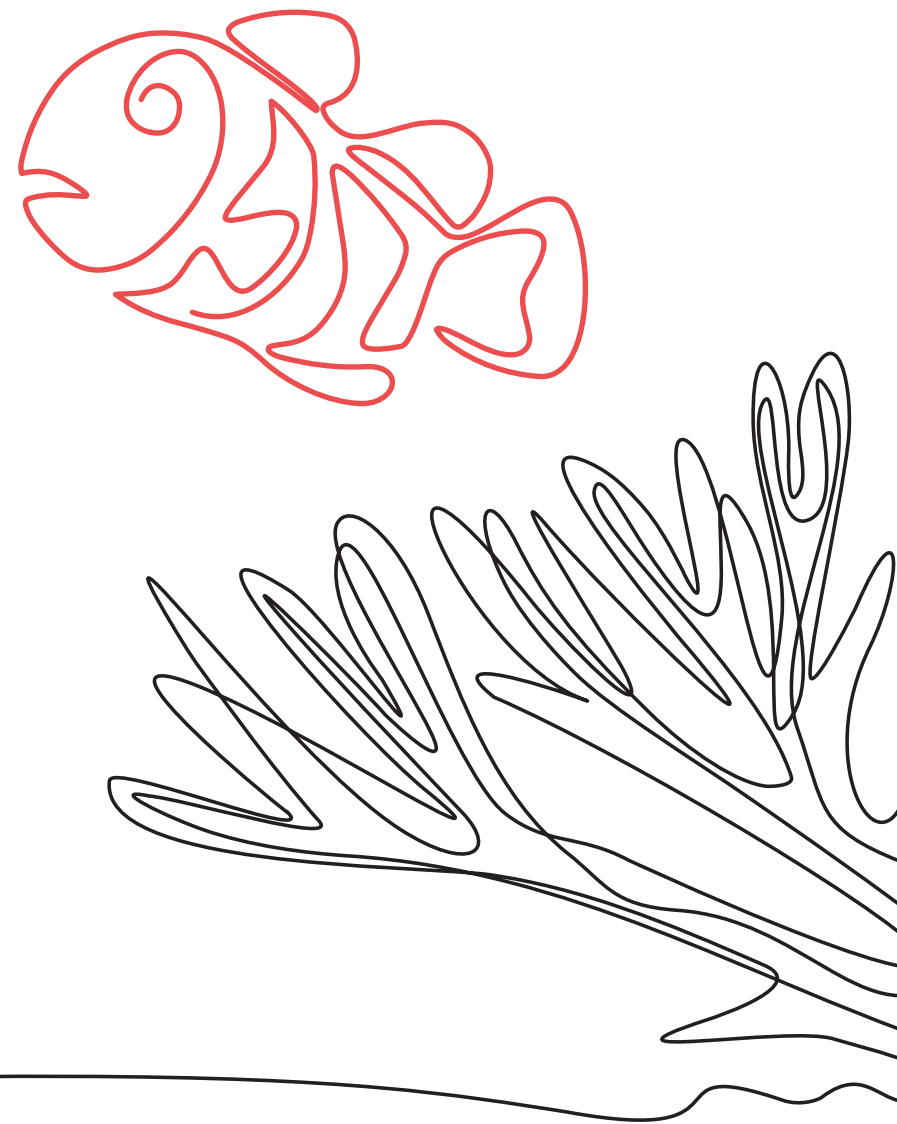
- Data
- Hardware
- Brugerafdærd



(Bio)diversitet skaber robusthed

Taleteknologi er robust, når det er diverst:

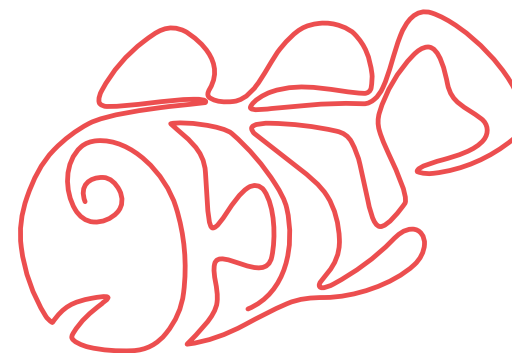
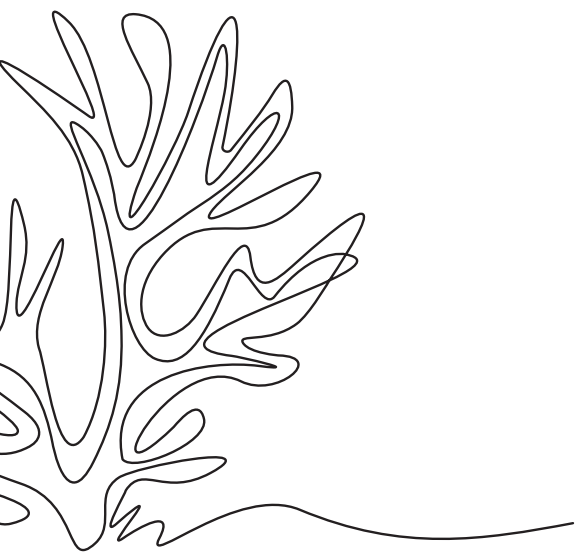
- mange dialekter og accenter
- forskellige stemmetyper, aldre og tale tempo
- flere måder at formulere sig på



Koralrev lever i symbiose

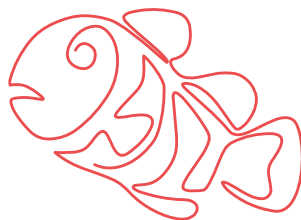
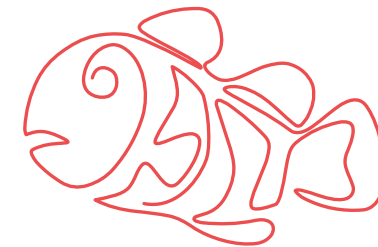
Teknologien skal:

- Aflaste mennesker
- Give mening i deres arbejde
- Ikke føles som kontrol eller overvågning



Miljøbetingelserne skal være på plads

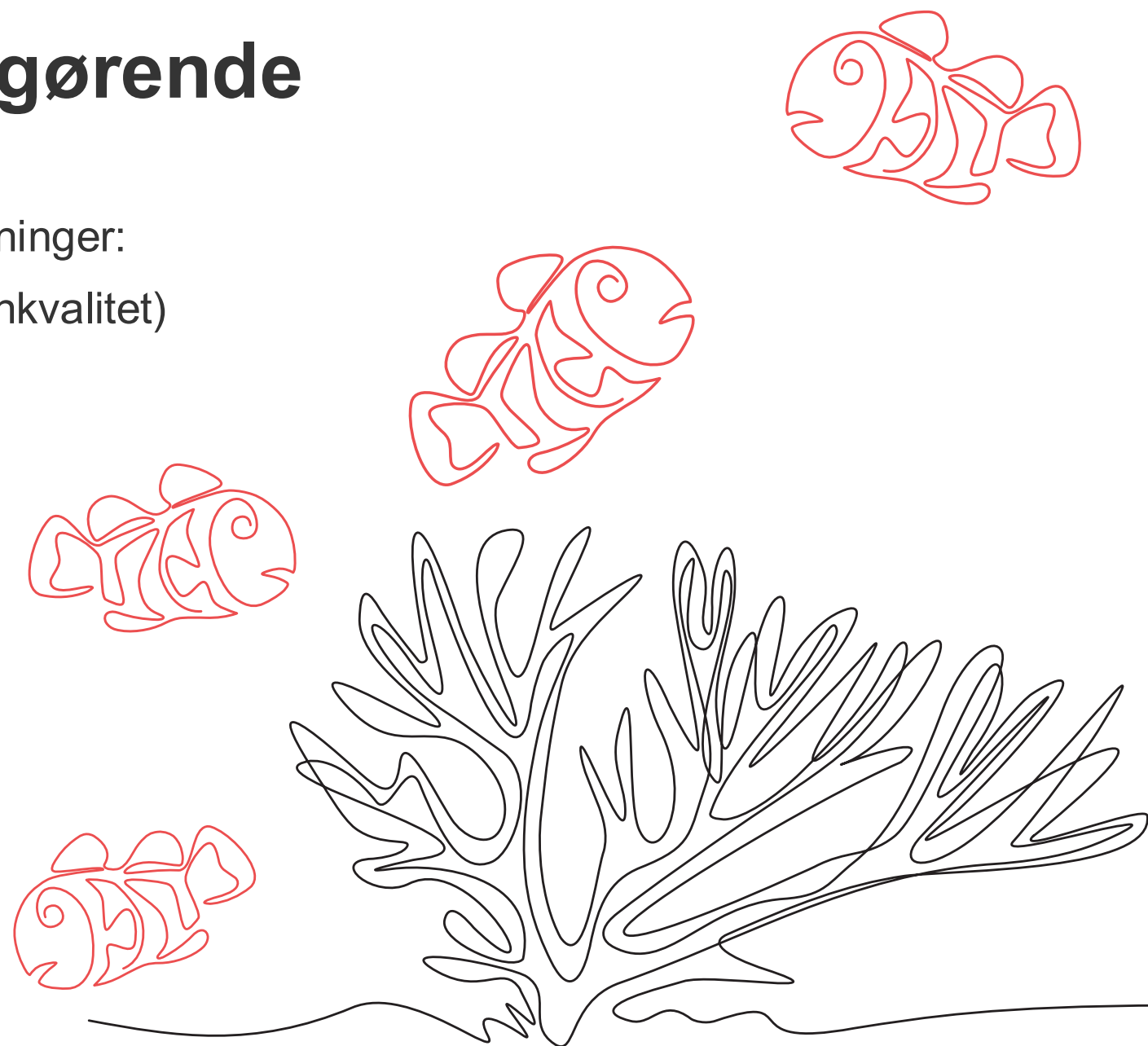
- Dataprivacy og compliance
- Kvalitet (WER, stabilitet, robusthed)
- Driftsmodel (on-prem, cloud, hybrid)



Lokale forhold er afgørende

God taleteknologi kræver lokale tilpasninger:

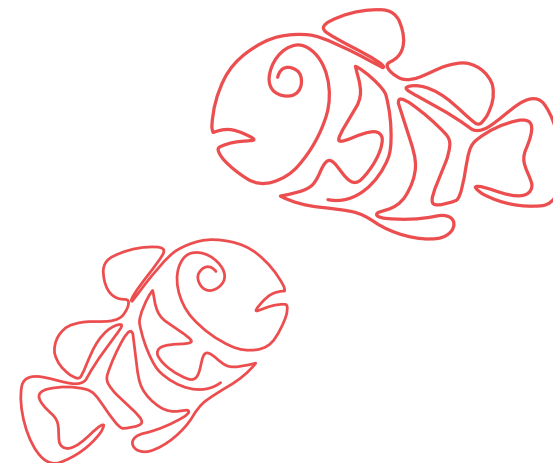
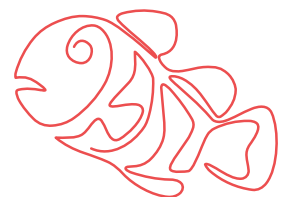
- Miljøspecifikke hensyn (støj, telefonkvalitet)
- Domænespecifik data



Et sundt rev er dynamisk, vokser, skifter farver og tilpasser sig

Taleteknologi skal kultiveres og det kræver:

- Løbende monitorering
- Opdatering af modeller, pipelines og processer
- Håndtering af ændrede behov og nye use cases



TaleTek – et dansk taleteknologisk forum



TaleTek – et dansk taleteknologisk forum

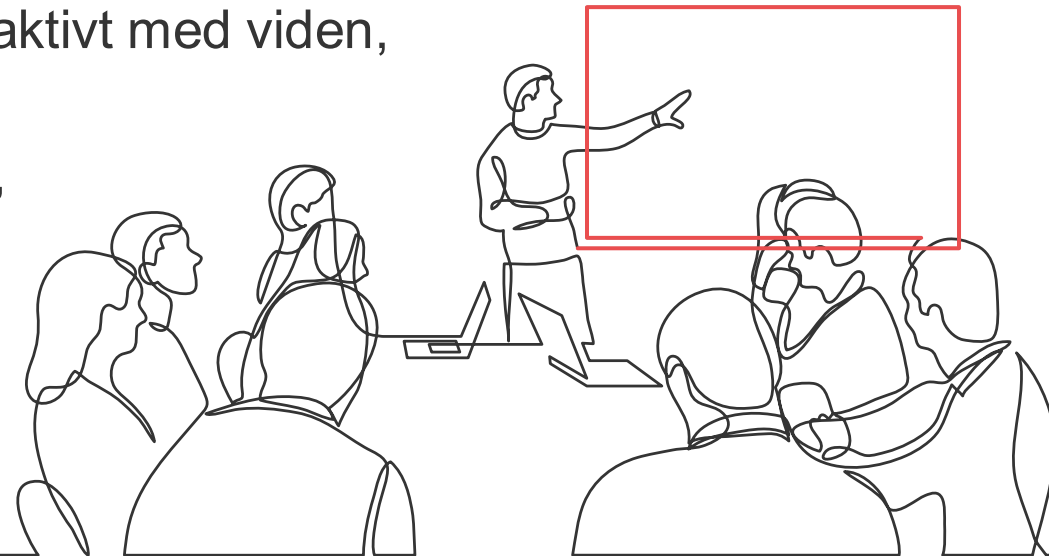
Formål

Netværket skal være et forum for videndeling, inspiration og samarbejde, med plads til at dele erfaringer, præsentere projekter og udforske nye anvendelser – fra forskning til kommercielle løsninger.

Hvem kan deltage

Netværket er for personer, der kan og vil bidrage aktivt med viden, perspektiver og idéer eks.

- Virksomheder, myndigheder og organisationer, der anvender eller arbejder med taleteknologi.
- Forskere og udviklere med specialviden inden for taleteknologi.



TaleTek – et dansk taleteknologisk forum

- To årlige møder med oplæg, workshops og diskussioner. Fokus på konkrete problemstillinger og løsninger med mulighed for deep dives i teknologierne.
- Første møde den 1. september kl. 13-15 i København

Hvorfor skal du være med?

- Bliv en del af et stærkt fagligt fællesskab.
- Få indsigt i de nyeste trends og teknologier inden for dansk taleteknologi.
- Skab samarbejder, der kan føre til nye projekter og innovation.



Kontakt

Sif Bernstorff Lehmann
Senior anthropologist

sif.lehmann@alexandra.dk
+45 6018 5869

Simon Rands Leminen
Senior AI Specialist

simon.leminen@alexandra.dk
+45 9352 2354