

# Vitlíki til Tala-til-Tekst máltøkni til føroykst

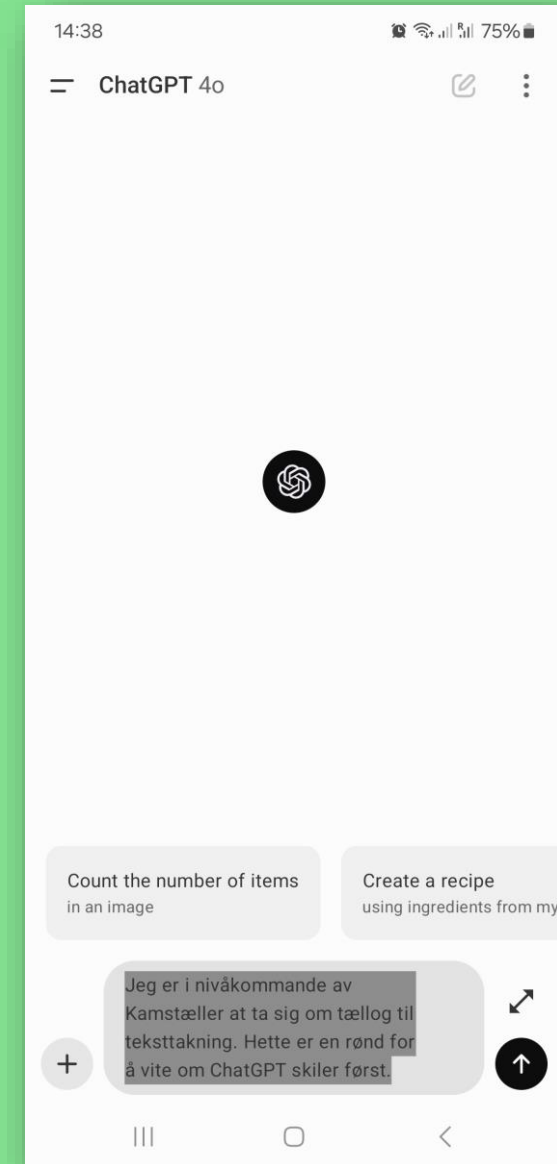
*Dávid í Lág, M.Sc. Teldufrøði, Ph.D.-lesandi*



FRÓÐSKAPARSETUR FØROYA  
UNIVERSITY OF THE FAROE ISLANDS

# Tala-til-Tekst

- Tosa við ChatGPT á føroyskum?
- Hví granski eg í Tala-til-Tekst tøkni
  - Arbeiða við tí nýggjastu tøknina innan vitlíki
  - Áhuga innan mál, serliga í Norðurlendskum
  - Ljóð
- Raðfesting og íløga frá tí almenna
- Endamálið við granskingini
  - Gera tað gjørligt hjá teldum at skilja og skriva føroyska talu



# Hvat er Tala-til-Tekst?

- Umskapa talað mál til skrivaðan tekst
  - Lætt hjá menniskjum, trupult hjá teldum
- Tøknin kann t.d. brúkast til
  - Tekst frá tímanum ella fundum
  - Undirtekstir á YouTube
  - Voice commands/search
- Avbjóðingin
  - Riggar væl á enskum
  - Riggar ikki væl á føroyskum

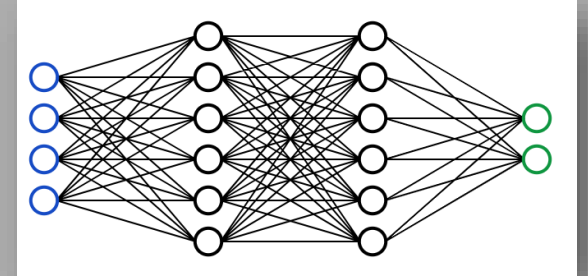


# Hví er føroyskt ein avbjóðing?

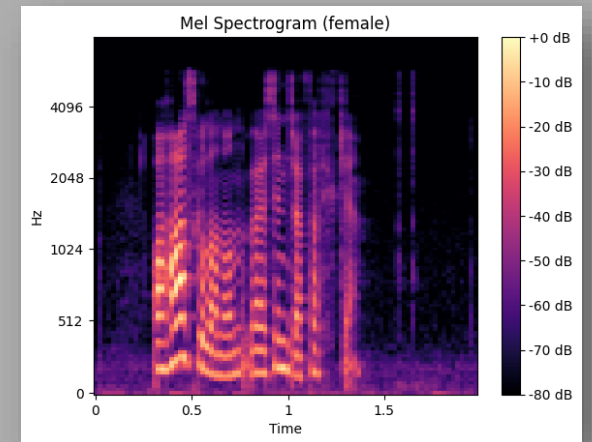
- Lítið mál
  - Bara 50.000+ fólk tosa tað
  - Ikki nógvar dátur
- Kompliksiteturin av talu
  - Dialektir, aldur, kyn, talumátar, larmur, v.m.

# Tala-til-Tekst tøkni

- Maskinlæring og djúplæring við Neuralum netverkum
- Mappa ljóð til represantióinir av ljóðið
- Brúka dátur at trena modellir at læra tað føroyska málið
- Málmodellið ger forsøgn um orð ella setningar

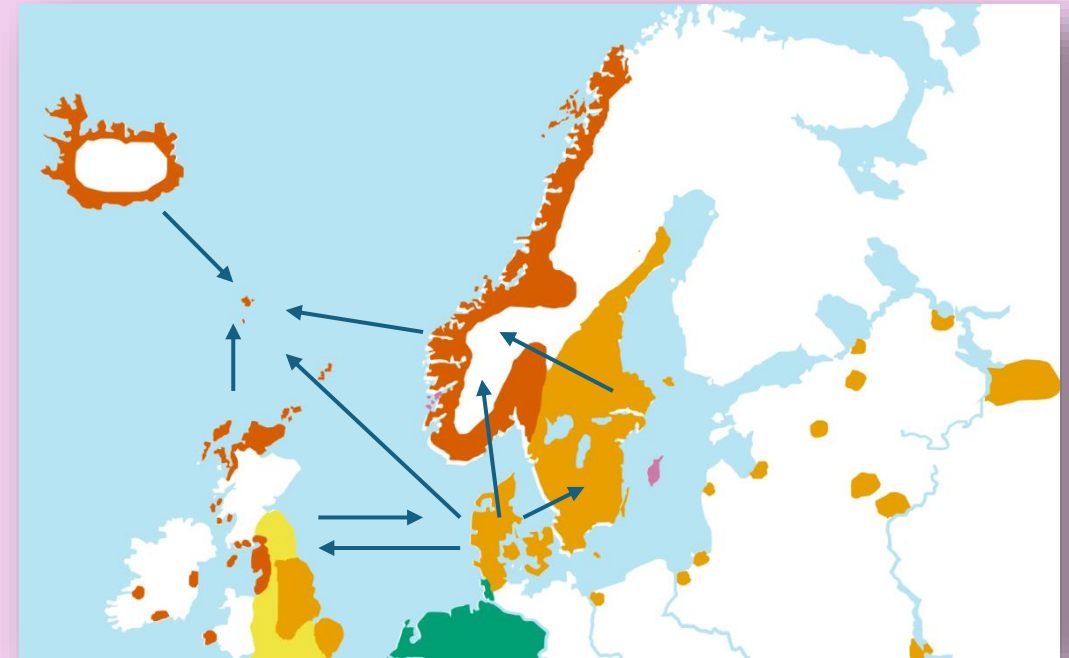


*”Á háskúlanum var eitt sindur av fólki”*



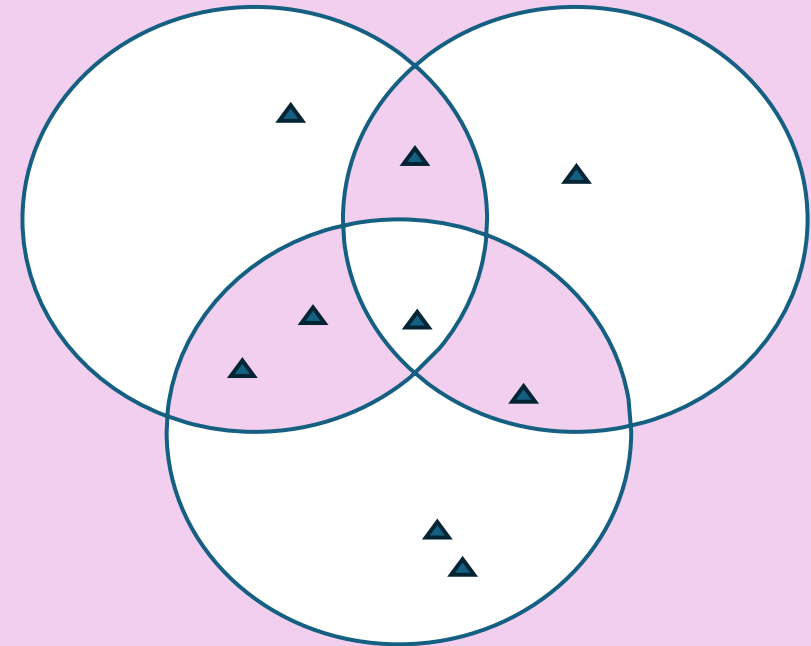
# Granskingin

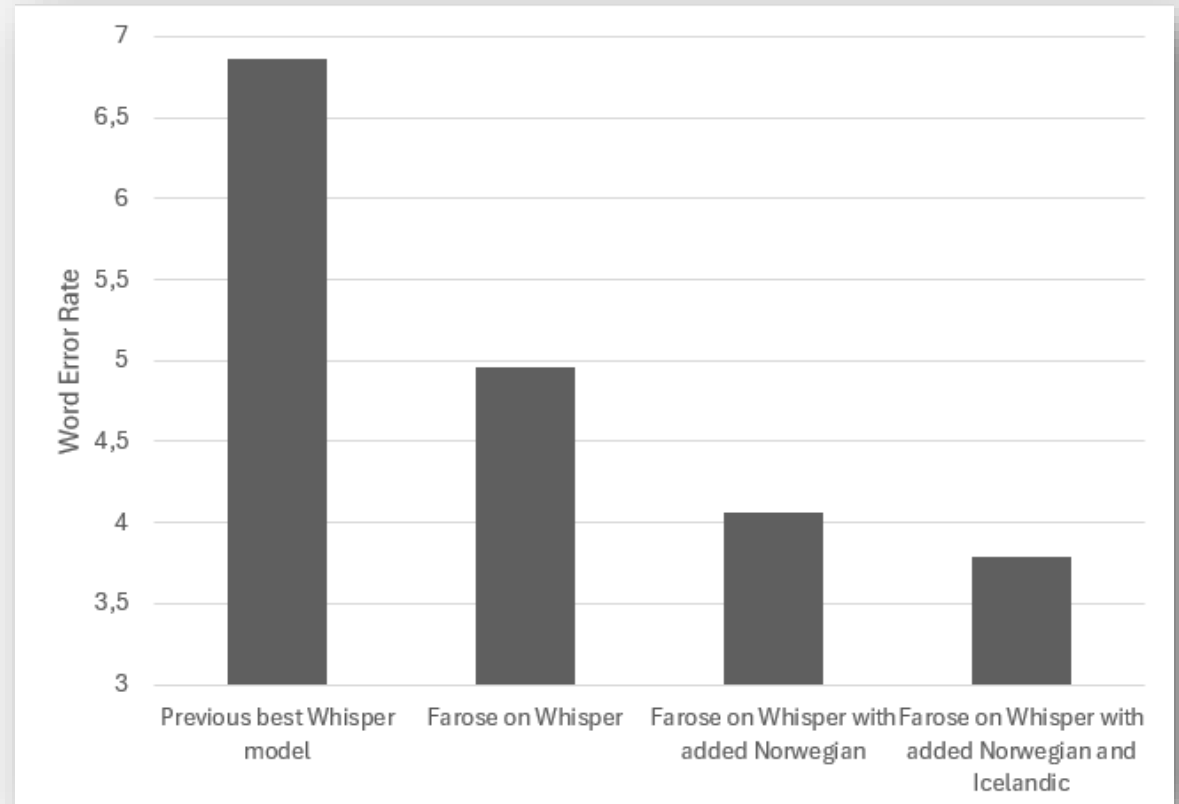
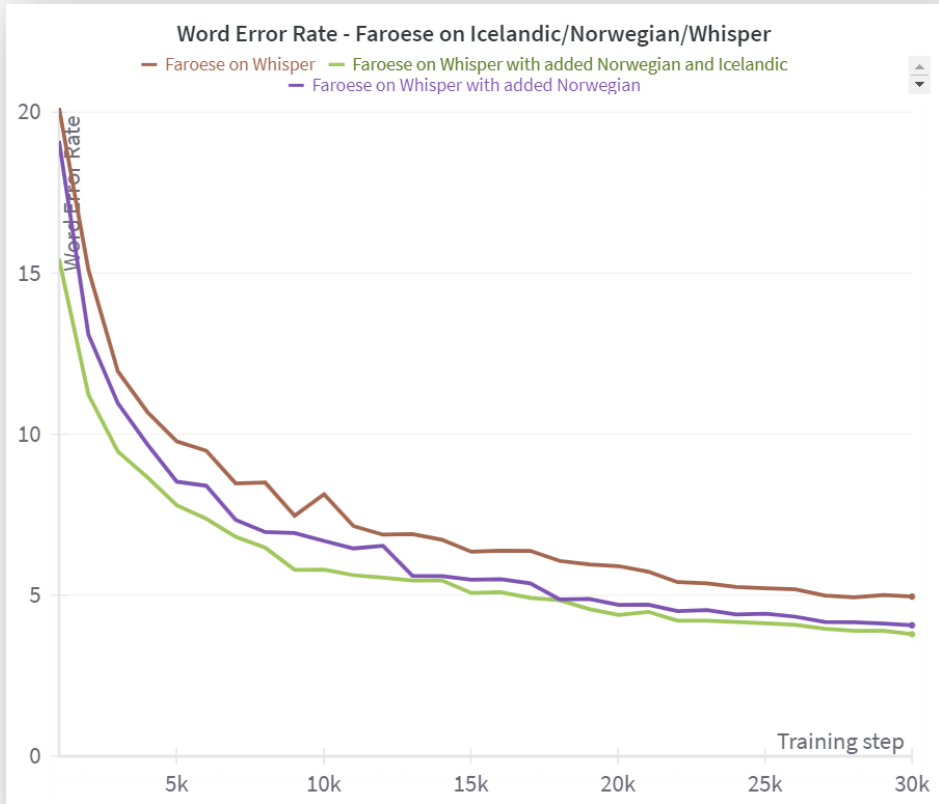
- Old Nordist sum byrjan
- Vitana­rflyting frá størri málum til smærri
- Menna nýggj málmodellir sum innihalda mál sum eru tætt við føroyskt
- Kanna hvørja ávirkan tað hevur
- Kanna fjarstöðuna millum føroyskt og hini máluni



# Granskingin

- Fleirmálmodellir læra talaðar representatióinir millum mál
  - *wav2vec 2.0* (Meta)
    - 56.000 tímar, 56 mál
    - Lærur um mál uttan tekst við at fjala talu undir trening
  - *Whisper* (OpenAI)
    - 660.000 tímar, 99 mál, 1.5B
    - Lærur mál við myndum og tilhoyrandi tekst
- Distillatióin av málmyndlum
  - Minka um stødd og tíð
- Dátu augmentering
  - Samla dátur sum longu finnast
  - Gera nýggj dátur út frá tí sum er til
  - Gera nýggj dátur við Tekst-til-Talu tøkni







# Ávirkan av granskingini í heiminum

- Vísindaliga ávirkan: Betri avrik við færri dátur fyri lítil mál
- Tekniska ávirkan: Gera nýggj Talu-til-Tekst modell tøk fyri øllum at brúka
- Sosiala ávirkan: Betri atgongd til góða Talu-til-Tekst tøkni fyri øll
  - Fáa tekst úr føroyskari talu
  - Hjálp til fólk við niðursettari ella ongari hoyrn, orðblind, v.fl.
  - Tekst til alt hjá KVF, sjónvarp og útvarp
  - Løgtingið, beinleiðis og savnið
  - Brúka tekstin til greining/gransking
  - Til tey sum vilja læra føroykst

# Taka saman um og spurningar

- Brúka nútímans máltøkni at fáa føroyskt inn í tann talgilda heimin
- Áhugavert øki sum er í stórari menning
- Granskingin kemur við onkrum nýggjum
  
- Spurningar?

# Takk fyri meg

[www.setur.fo](http://www.setur.fo)

[davidl@setur.fo](mailto:davidl@setur.fo)

