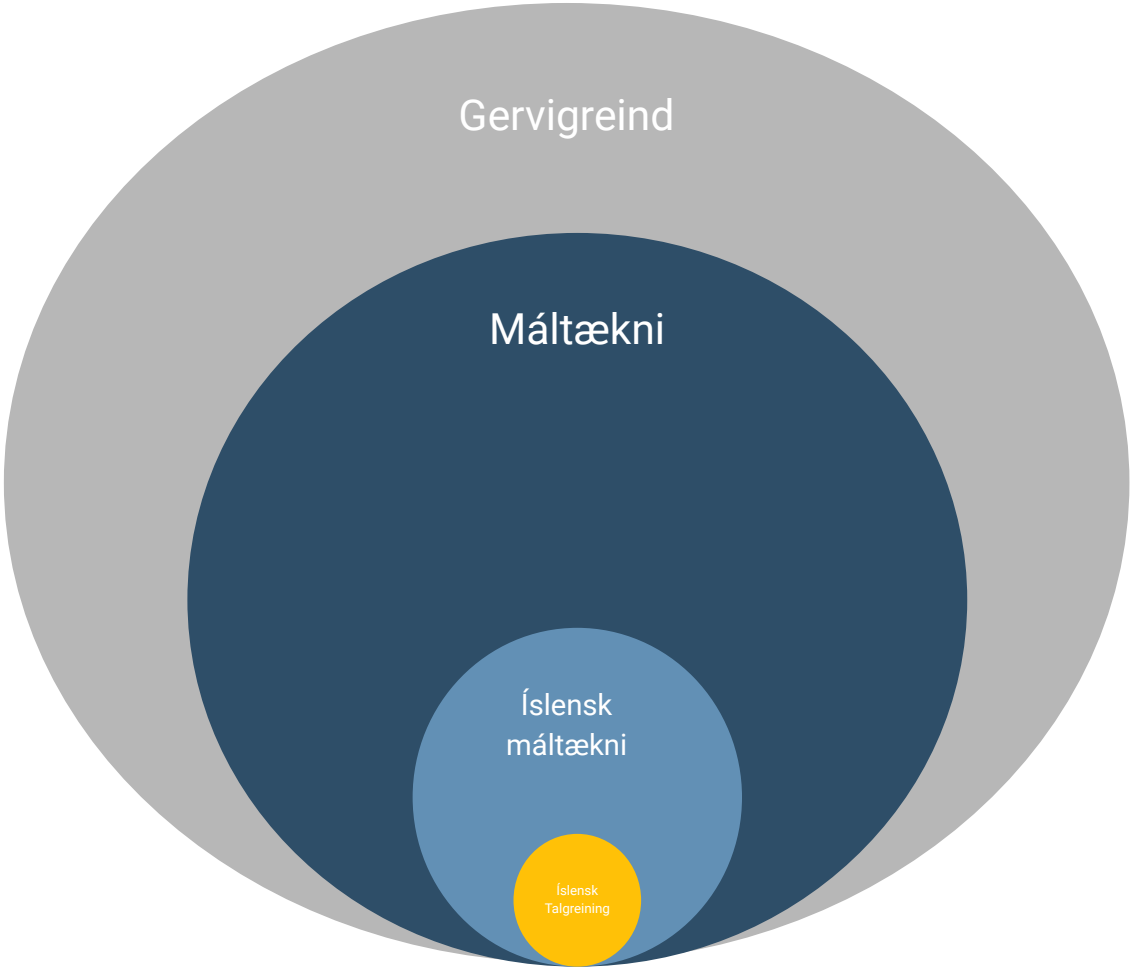
Ský hádegisfyrirlestur
Vélnám

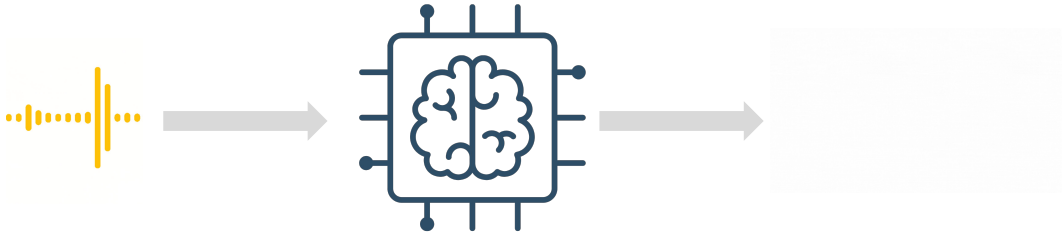
19.apríl 2023

Yfirlit

- Talgreining 101
- Hvað er hægt að gera við talgögn.
- Hvernig er talgreining að nýtast í dag á íslenskum markaði.

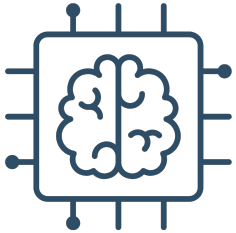

Gervigreind
Máltækni
Íslensk
máltækni
Íslensk
Talgreining

Hvað er talgreining?



Hvað er talgreining?

IBM shoebox



Djúp tauganet

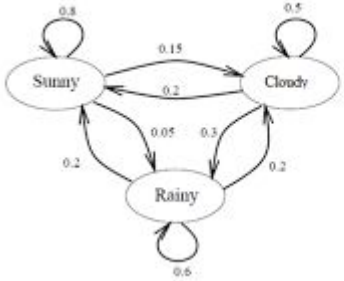
1950

1990

2010

2018 -

HMM - GMM líkön

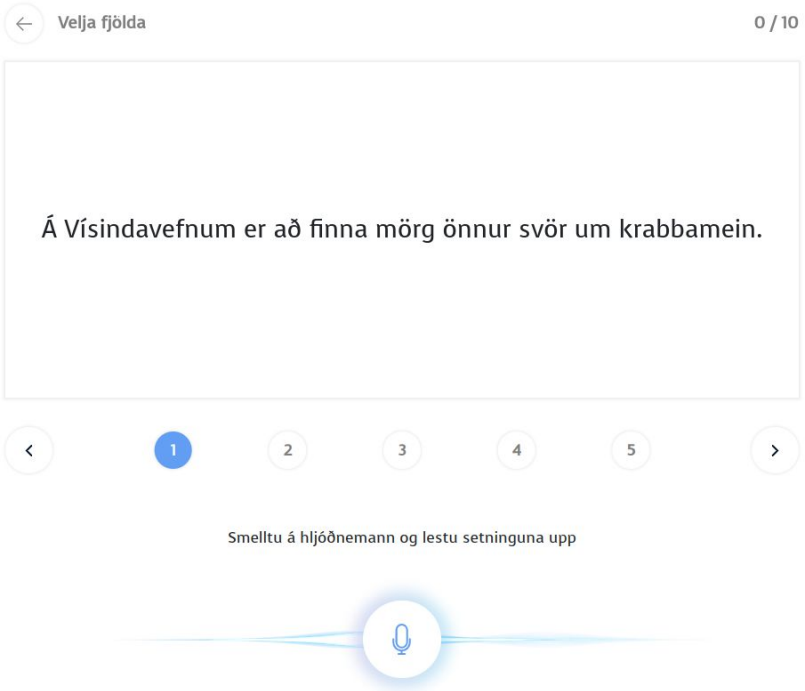


Transformer líkön

Talgreining 101

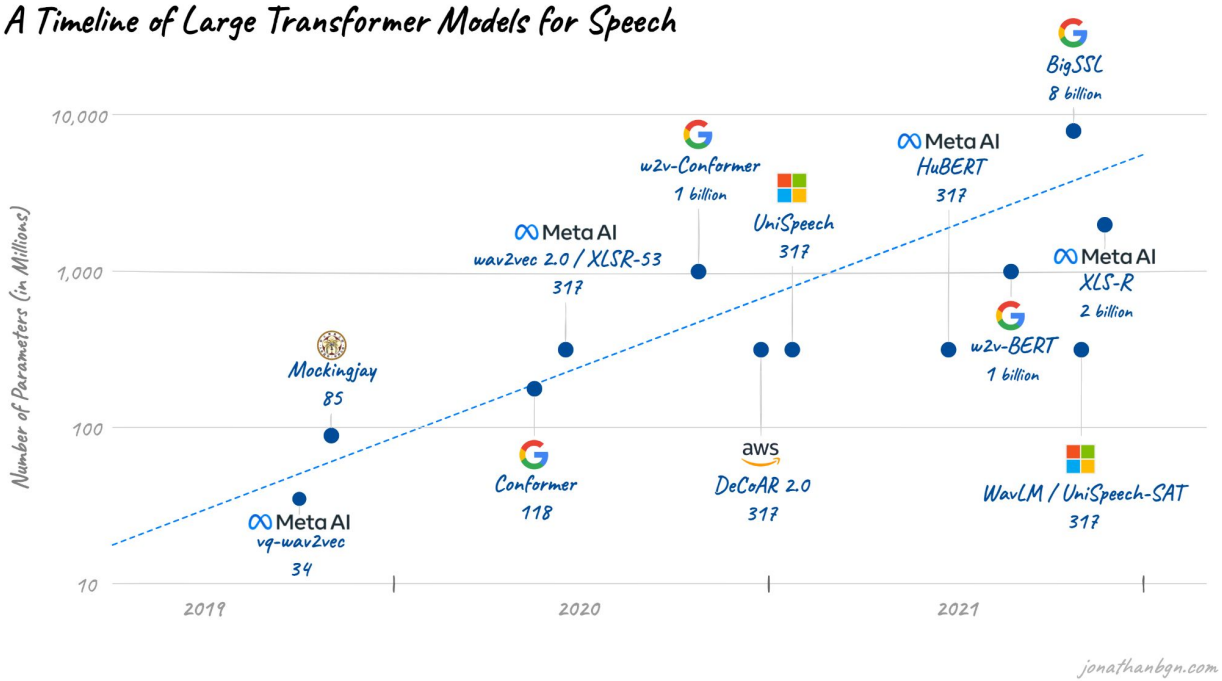
Talgreining 101

- Máltæknáætlun Stjórnvalda.
- Íslensk talgagnasöfn.
  - ~3000 klukkustundir.
  - Samrómur.
  - Alþingi.

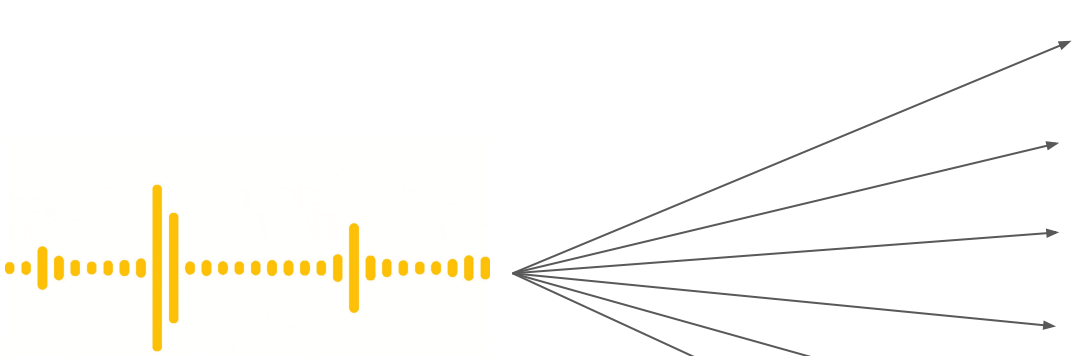


Gagnagnótt

- GMM-líkön 10-100 klst.
- Fyrstu tauganetin 1000-3000 klst.
- Wav2vec2.0 54.000 klst.
- XLS-R 500.000 klst.
- Whisper 680.000 klst.



Meira en talgreining



Talgreining

Mælenda aðgreining

Vélræn þýðing

Sannprófun mælanda

Tungumála greining

Hljóðflokkun

Dæmi um talgreining í noktun á Íslandi

Einfalda störf

- Alþingisræður
  - Lögbundin skylda til að birta ræður þingfunda.
  - Áður fyrr störfuðu ritarar og yfirlesarar. Nú er einungis yfirlesarar.
- Diktering lækna
  - Það þarf að huga að öryggi vinnslunnar.
  - Erfitt að nálgast gögn.
  - Lækníska.

Dæmi:

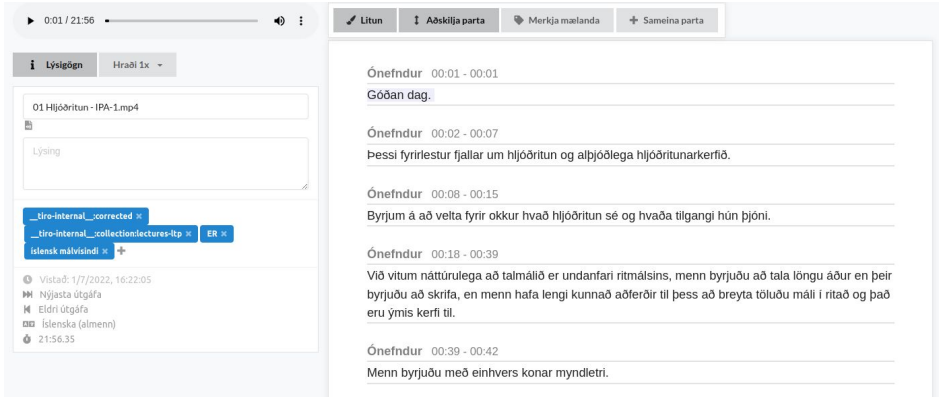
„seint væ positivt neer test, ei þreifi eymsl yfir a c liðnum, þreifi eymsl yfir löngu bicepssin“

Auka aðgengi að efni

- Rauntímatextun á myndefni.
- Mikilvægt fyrir heyrnalausa og heyrnaskerta.
- Áskoranir: SMS íslenska.
- Eykur textun á íslensku efni.

Greining á talgögnum

- Greining og vinnsla á viðtölum.
- Ísmús - Sögulegar upptökur.
- Samræður í þjónustuverum.



www.tiro.is tiro@tiro.is www.talgreinir.is


fá upptöku af því að sjá hvernig það gengur að túlka þetta því ég
tala oft ansi