

# **Keeletehnoloogia areng ja eestikeelne tehisintellekt: Kust tuleme ja kuhu liigume?**

Kairit Sirts (sirts@ut.ee)

TÜ keeletehnoloogia kaasprofessor

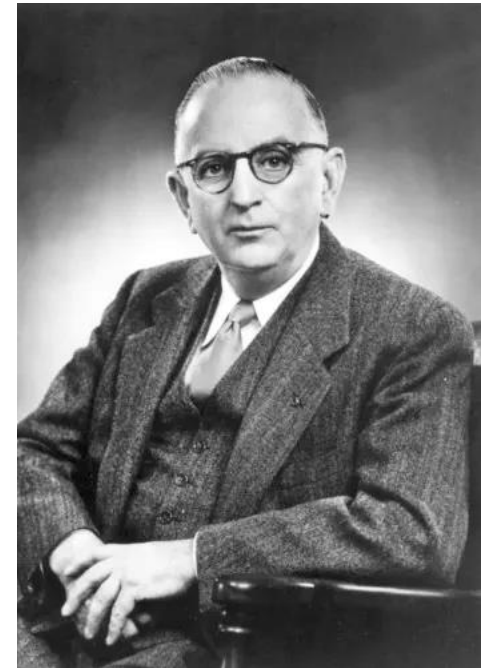
Mäluasutuste talveseminar, 17.02.2025

# Warren Weaver masintõlke

1947 – kirjas Norbert Wiener'ile

*One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'*

*Loomulikult tekib küsimus, kas tõlkimist võib vaadelda krüptograafia probleemina. Kui ma loen venekeelset artiklit, siis mõtlen: „See on tegelikult kirjutatud inglise keeles, kuid kodeeritud veidrates sümbolites. Nüüd hakkab seda lahti kodeerima.“*



# Frederick Jelinek kõnetuvastusest

1980ndate algus

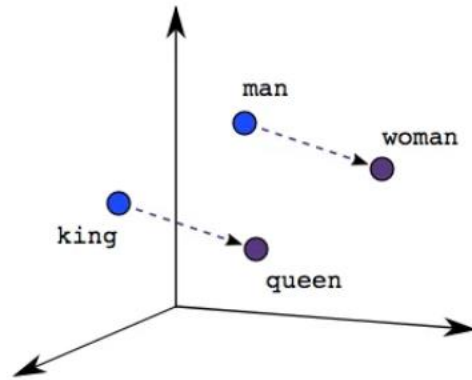
*Every time I fire a linguist, the performance of the speech recognizer goes up.*

*Iga kord, kui vallandan lingvisti, kõnetuvastuse kvaliteet paraneb.*

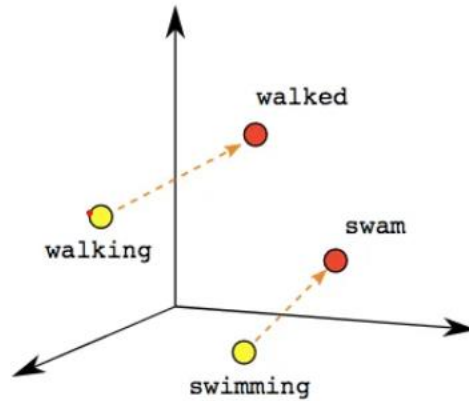
*Iga kord, kui lingvist grupist lahkub, kõnetuvastuse kvaliteet paraneb.*



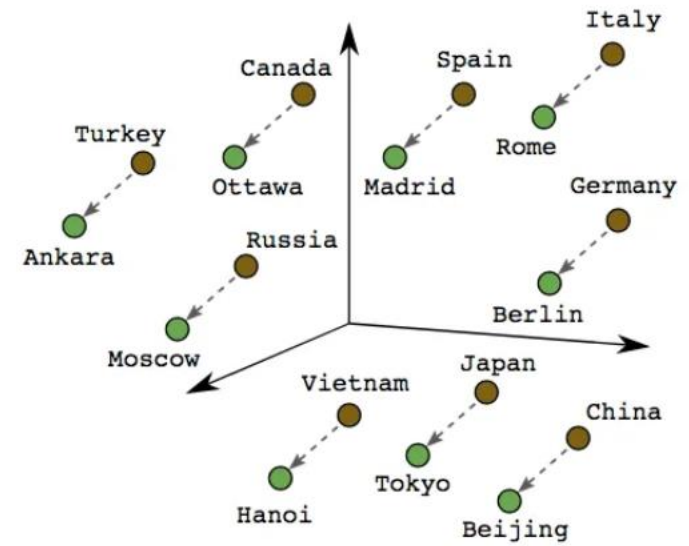
# Sõnavektorid (2013)



Male-Female



Verb Tense



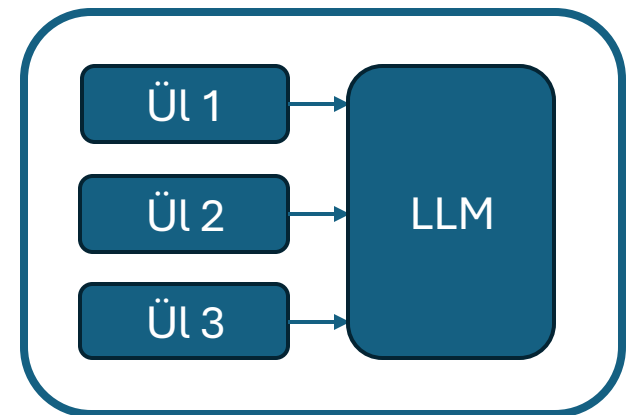
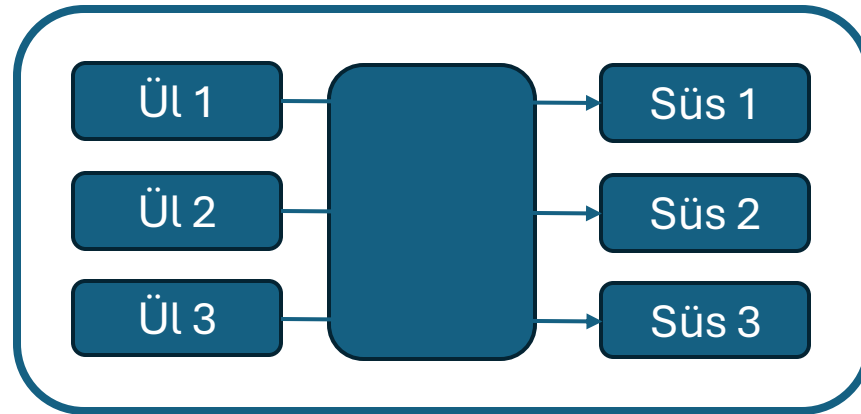
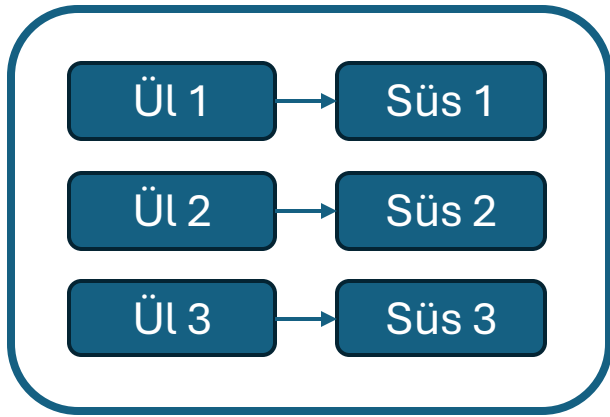
Country-Capital

# Suured keelemudelid

30. november 2022



# ChatGPT



# neurotolge.ee

**TARTU**   
NEUROTÕLGE

[Info](#)

[Koostöö](#)

[API](#)

[EN](#)

[ET](#)

[Kõik demod](#)

Sisendkeel : eesti



Väljundkeel : inglise

Kirjuta siia oma tekst ja me  
tõlgime selle ära!

0 / 7000 



Mis teeb Neurotõlke eriliseks?

# tekstiks.ee



ESILEHT

DEMO

ET

LOGI SISSE

Päevakaja 16.05.mp3

Kestus: 00:03:30

Ülesse laetud: 22. juuni 22

Muutmise režiim:

Tavaline (tekstis vigade parandamine)

Märgendamine (teksti ja kõne põhjalik märgendamine)



LAADI ALLA

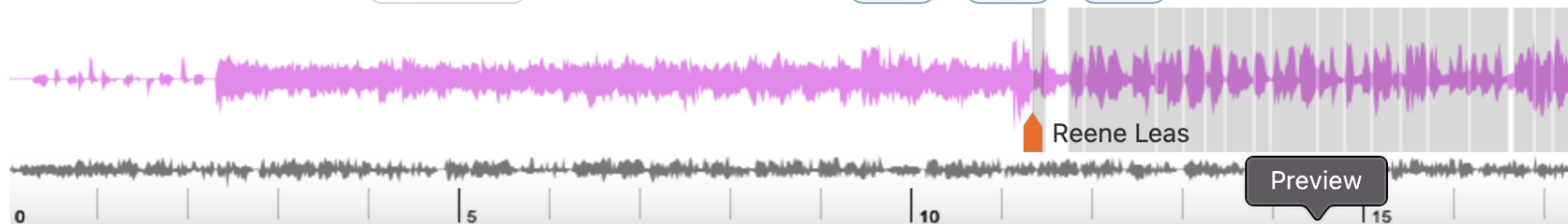
Reene Leas

00:11

Ta on 16. mai kell sai kuus ning Päevakaja võtab kokku tänase päeva tähtsamad uudised. Mina olen Reene Leas. Valitsus sidus lisaelarve vastuvõtmise riigikogus usaldusküsimusega.



1.0x



# EKI keelekorpused

- Eesti keele ühendkorpuste sari [↗](#) on kaasaegseim, mahukaim ja žanriliselt mitmekesiseim eestikeelsete digitekstide kogu (loe lähemalt Koppel ja Kallas 2022 [↗](#)).

– Eesti keele ühendkorpus 2013 (563 miljonit sõna)

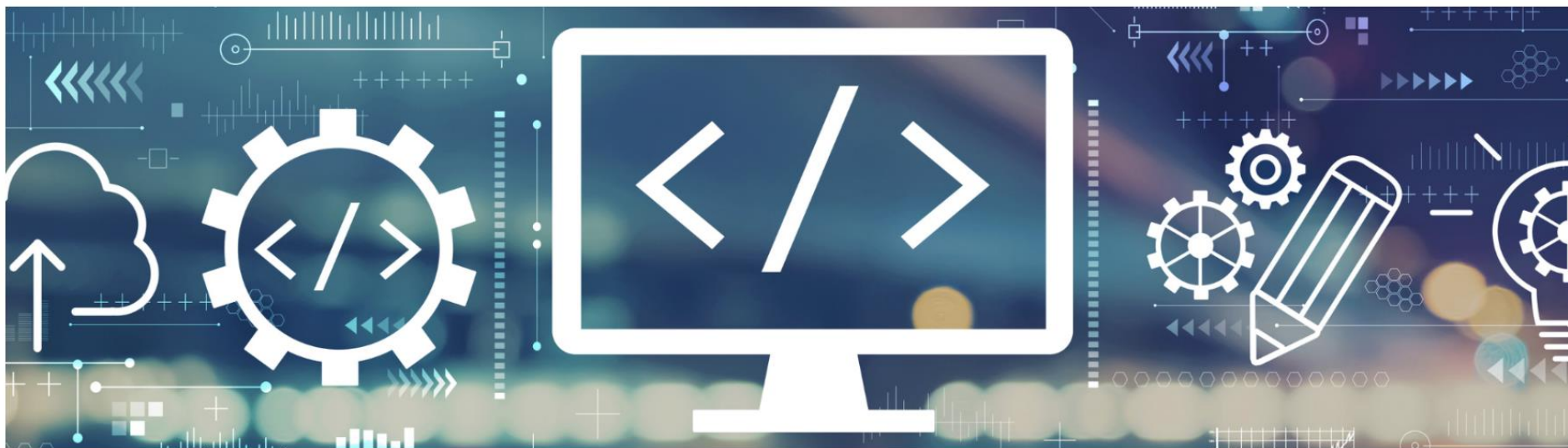
– Eesti keele ühendkorpus 2017 (1,1 miljardit sõna)

– Eesti keele ühendkorpus 2019 (1,5 miljardit sõna)

– Eesti keele ühendkorpus 2021 (2,4 miljardit sõna)

– Eesti keele ühendkorpus 2023 (3,8 miljardit sõna)

# Eesti keeletehnoloogia programm



Riiklik teadus- ja arendustegevuse programm „[Eesti keeletehnoloogia 2018-2027](#)” (EKT) toetab eesti keeletehnoloogia valdkonna arengut. Selleks luuakse programmi toel uusi keeletehnoloogilisi rakendusi, parandatakse olemasolevate rakenduste kvaliteeti ning võetakse neid kasutusele võimalikult paljudes valdkondades.

# Eesti keele toetus suurtes generatiivsetes vabavaralistes keelemudelites



**TAL  
TECH**

**EESTI  
KEELE  
INSTITUUT**



TALLINNA ÜLIKOOI

# vestle.tartunlp.ai


 Vestlemine  Meist


## Keelemudelite edetabel: aita valida parimat eestikeelset keelemudelit!


### Kuidas see töötab?

- **Pimetest:** Esita oma küsimus. Sinu küsimusele vastavad kaks anonüümset keelemudelit.
- **Vali parim:** Vali kahest vastusest parim. Kui sa kohe valikut ei oska langetada, võid vestlust jätkata kuni oled otsuseni jõudnud.
- **Aus mäng:** Sinu valikute põhjal koostame mudelite edetabeli. Palun tee oma otsus vastutustundlikult.

### Vestlema!


 Kliki siia, et näha võrdluses olevaid mudeleid


 Mudel A

 Mudel B

Kas sa tead midagi mäluasutuste talveseminarist?

**Saada**

 Uus vestlus

 Genereeri vastus uuesti

 Jaga

# vestle.tartunlp.ai

## Vestlema!

Kliki siia, et näha võrdluses olevaid mudeleid

Mudel A

Mudel B

Kas tead midagi mäluasutuste talveseminarist?

Yes, I can provide information about the winter seminar for memory championships. Please provide me with specific details about the event you are interested in, such as the year, location, or specific competitors, and I will do my best to help you.

Kas tead midagi mäluasutuste talveseminarist?

Mäluasutuste talveseminar on iga-aastane üritus, kus kogunevad mäluasutuste esindajad ja ekspertid, et arutada mäluasutuste arengut, ühisprojekte ja ühistegevusi. Seminaril on võimalus jagada kogemusi, õppida teineteiselt ja leida uusi lahendusi mäluasutuste ülesannete täitmiseks.

Talveseminaril käsitletakse sageli teemasid nagu:

- o Mäluasutuste roll ühiskonnas ja nende tulevikuväljavaated
- o Digiteerimine ja digitaalne säilitamine
- o Kogude hooldus ja konserveerimine
- o Kasutajateenindus ja külastajate võrgutamine
- o Koostöö teiste asutustega ja organisatsioonidega

Seminaril esinevad tavaliselt ekspertid erinevatest valdkondadest, nagu ajalugu, raamatukogundus, arhiivindus, muuseumitöö ja digiteerimine. Lisaks esitlustele on seminaril sageli ka töötubadeid, aruteluid ja ühisprojektide tutvustusi.

Mudel A: Mistral-Nemo-Instruct-2407

Mudel B: Llama-3.3-70b-instruct

A on parem

B on parem

Viik

Mõlemad on halvad