

# NorGPT: Large Generative Language Models for Norwegian

Jon Atle Gulla
jag@ntnu.no

NorwAI

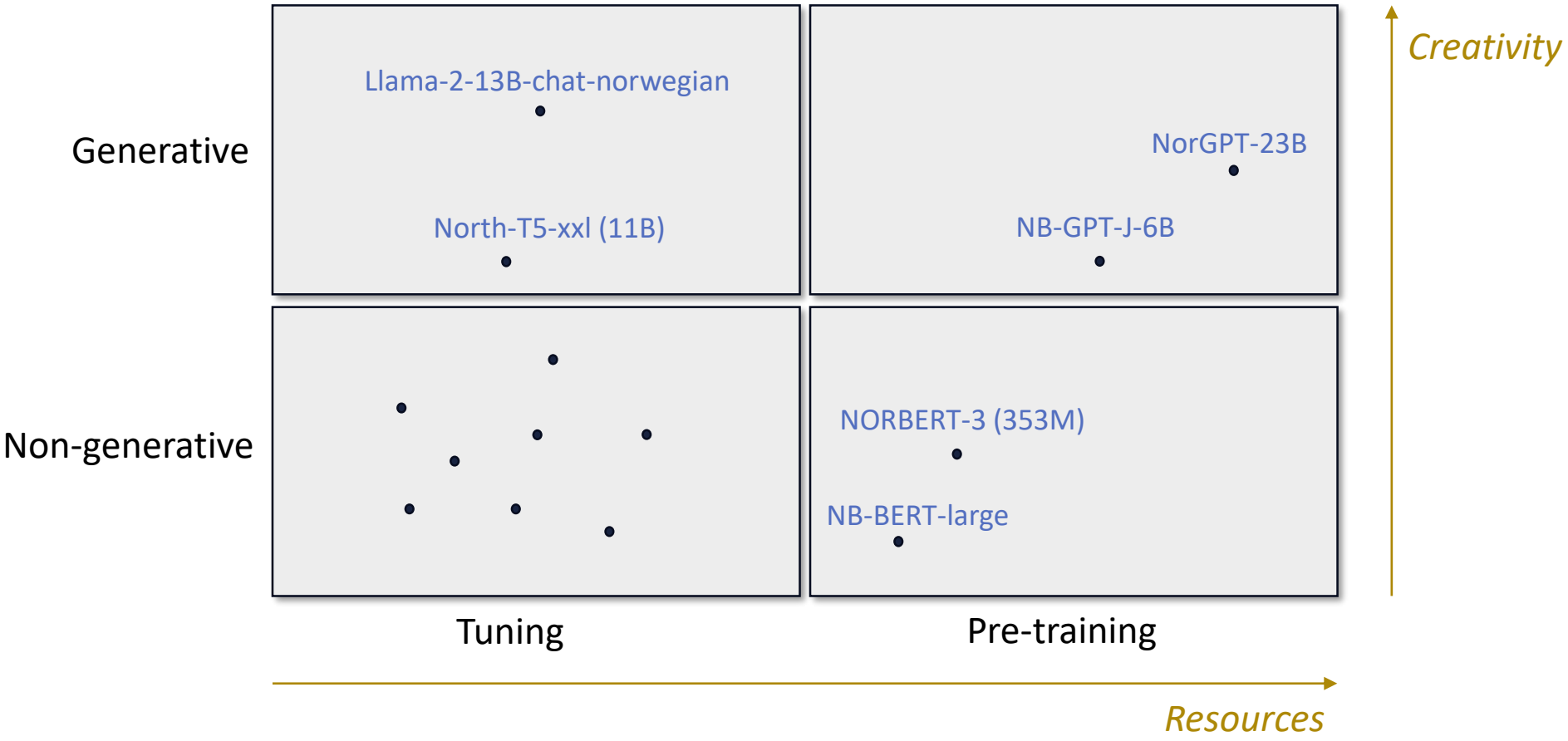
Norwegian Research Center for AI Innovation

NTNU

sfi Centre for Research-based Innovation

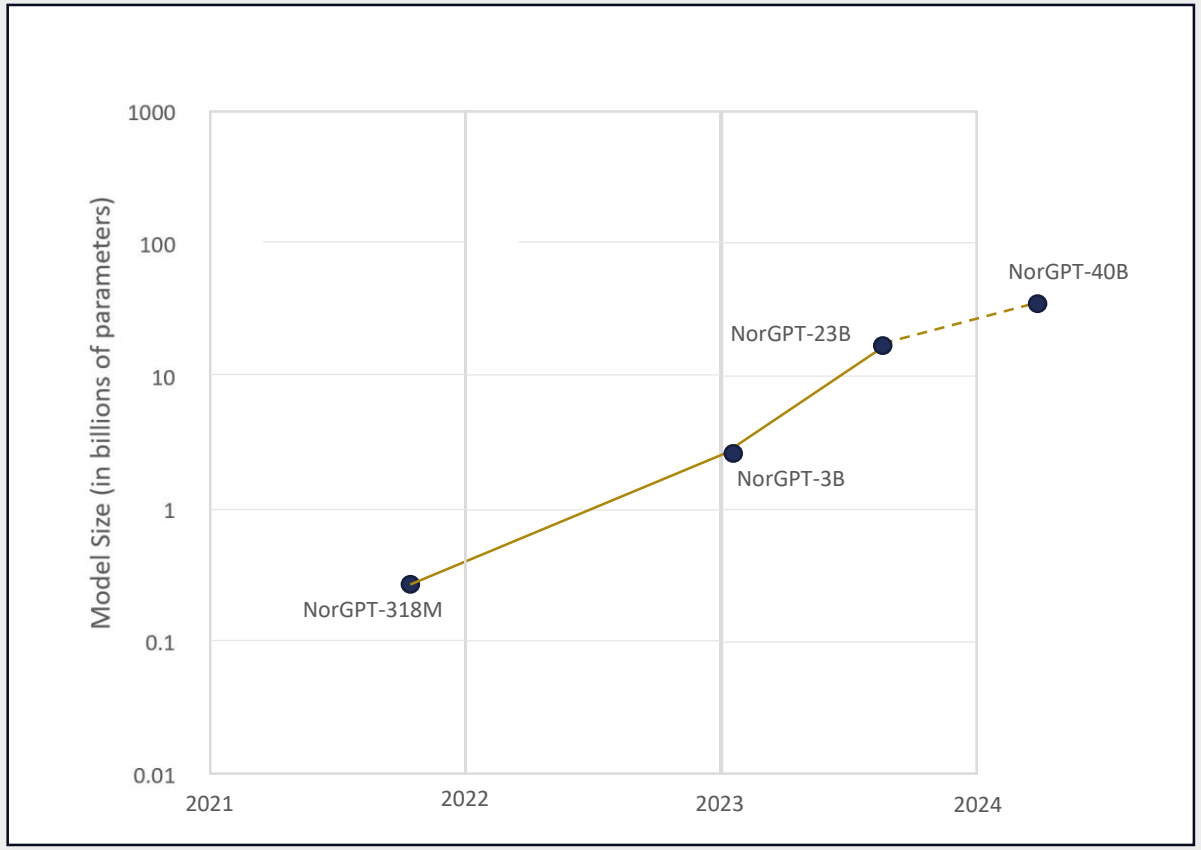
The Research Council of Norway

# Norwegian Language Models

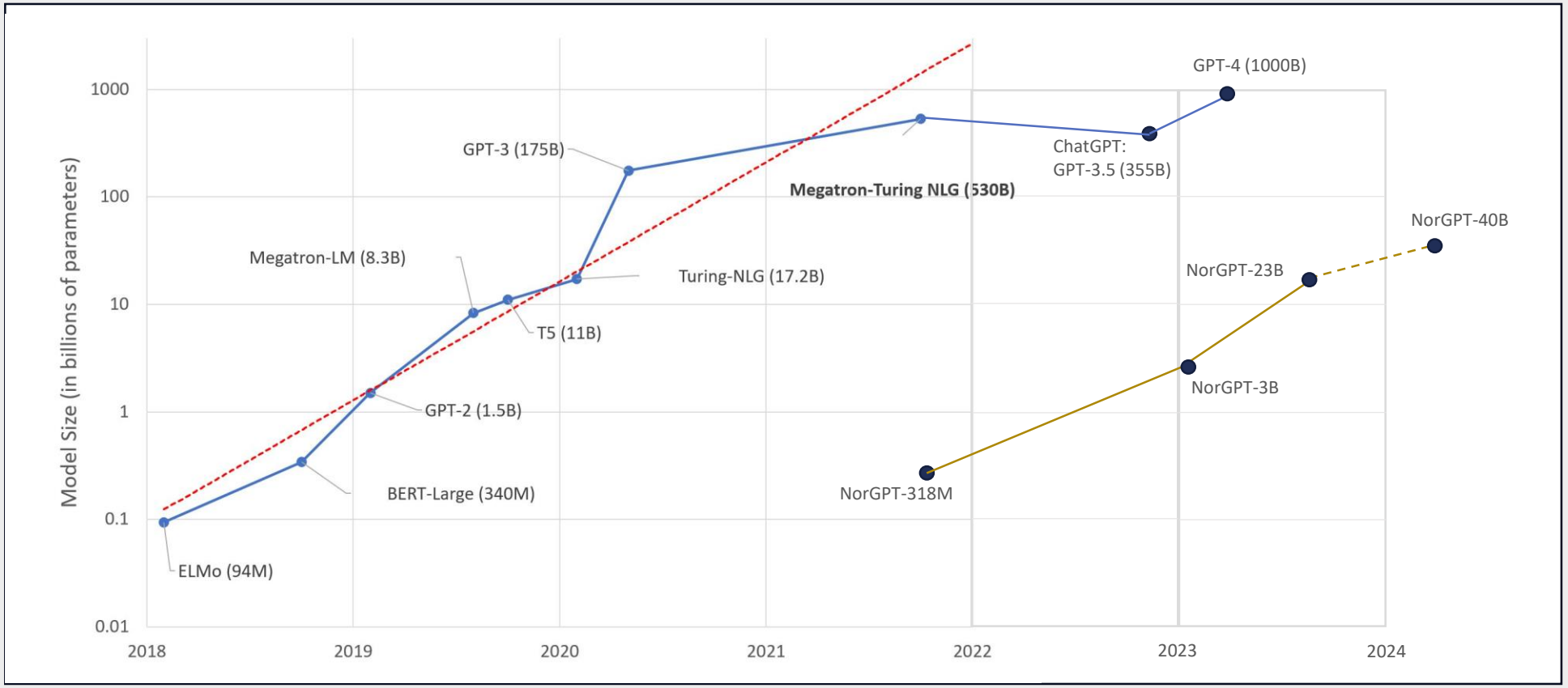


# NorGPT from NorwAI

- National generative language models for bokmål/nynorsk
- Foundation models
- From 318M to 40B parameters

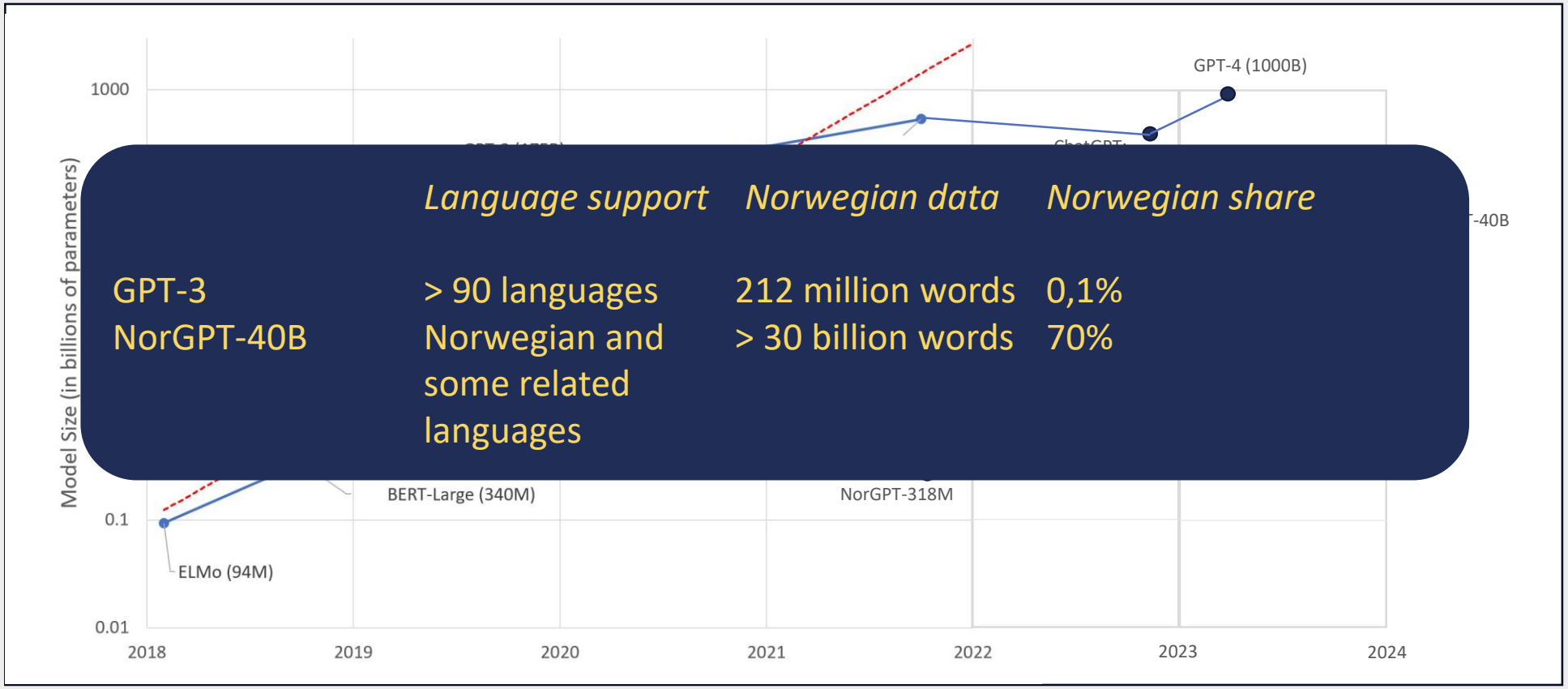


# International Models




https://huggingface.co/blog/large-language-models

# International Models




| | *Language support* | *Norwegian data* | *Norwegian share* |
|---|---|---|---|
| GPT-3 | > 90 languages | 212 million words | 0,1% |
| NorGPT-40B | Norwegian and some related languages | > 30 billion words | 70% |

https://huggingface.co/blog/large-language-models

# Norwegian Research Center for AI Innovation

Universities & institutes

NTNU

University of Stavanger

NR

SINTEF

Media & finance

DNB

Schibsted

NRK

SpareBank 1 SMN

*LAP SOC DATA*



Accelerate the innovation of sustainable and trustworthy AI solutions across Norwegian industries

| | |
|---|---|
| 261 MNOK | 8 years |
| 20-40 PhD candidates | 500+ MSc students |
| Hosted by NTNU | Multi-disciplinary |
| AI industry leaders | Open source innovation |

Industry 4.0

KONGSBERG

DNV

DigitalNorway

DigitalNorway

COGNITE

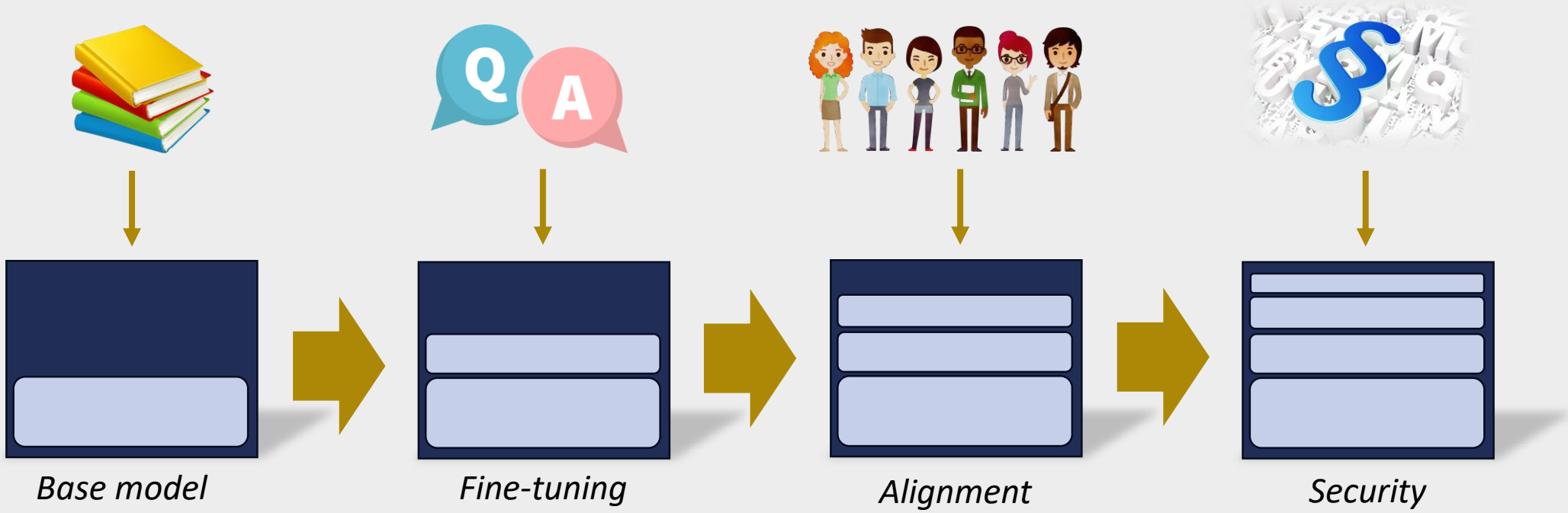
ANEO

telenor

*TRUST STREAM HYB DATA*

# Building Useful Language Models

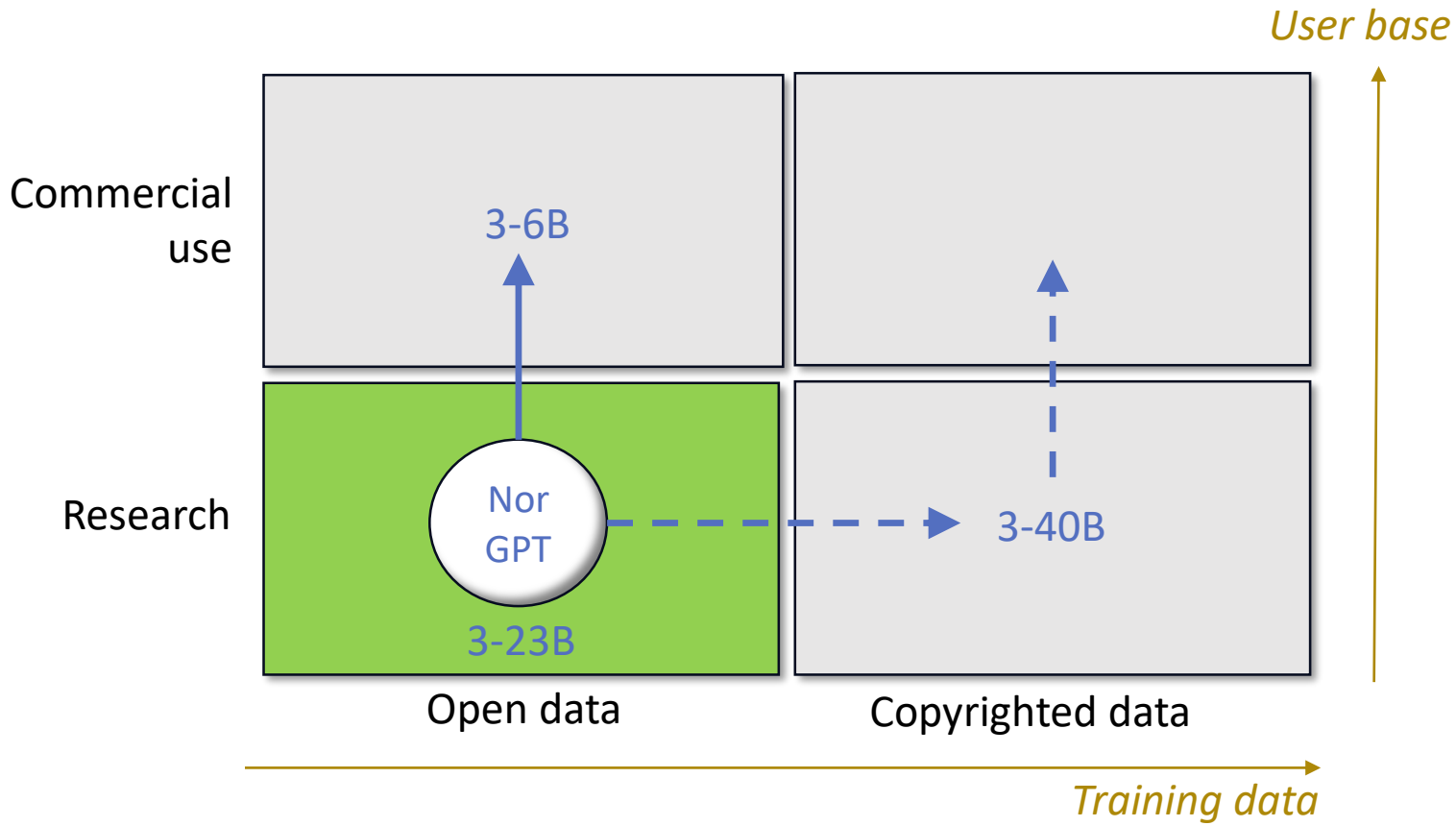
Competence — AI/NLP




Architecture — GPT transformers

Infrastructure — Idun supercomputer cluster

# Moving forward... Norway




*Academic partners*

*Data partners*

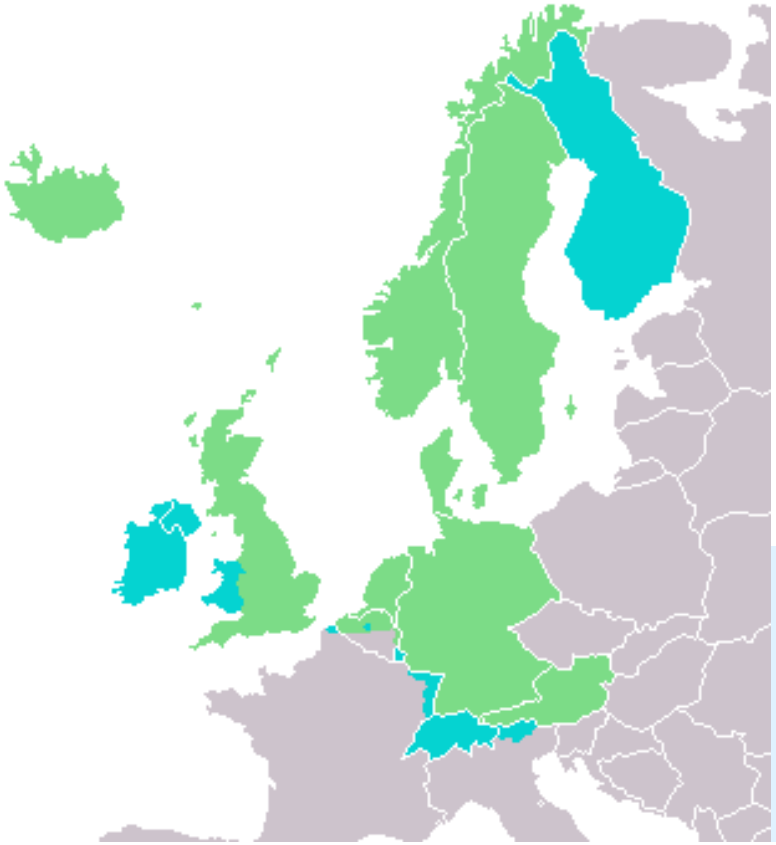
*User partners*

# Moving forward... Europe

**TrustLLM**

*Transfer effect between related languages*

- EU project starting up 2023
- Collaboration on generative language models among Germanic countries
- Partners from Norway, Sweden, Denmark, Iceland, Netherlands and Germany
- Coordinated by Linkoping University



Horizon Europe
Framework Programme for Research and Innovation (2021-2027)

# Challenges Ahead



- Infrastructure and costs
- Copyrighted content
- Low-resource languages like Sámi

*Jon Atle Gulla (jag@ntnu.no)*