NorLLM—Åpen norsk språkmodell fra NorwAI

Benjamin Kille

NorwAI Norwegian Research Center for AI Innovation

NTNU

Norwegian Centre for Research-based Innovation

AI is here!

NorLLM

To be available in Q2 2024

MISTRAL AI_

	GPT-2	LLAMA 2	Mistral
Size	7B parameter	7B parameter	7B parameter
License	Commercial use allowed	Commercial use allowed	Commercial user allowed
Setup	Pre-training	Continued Pre-training	Continued Pre-training

Use **gigantic** collections of “text” to **generalize** structures

1 **Gunnar og Per finansierer forskning og forskningsbasert innovasjon.**

2 **Gunnar og Per** ?

finansierer 0.05

snakker 0.02

ekorn 0.08

... ...

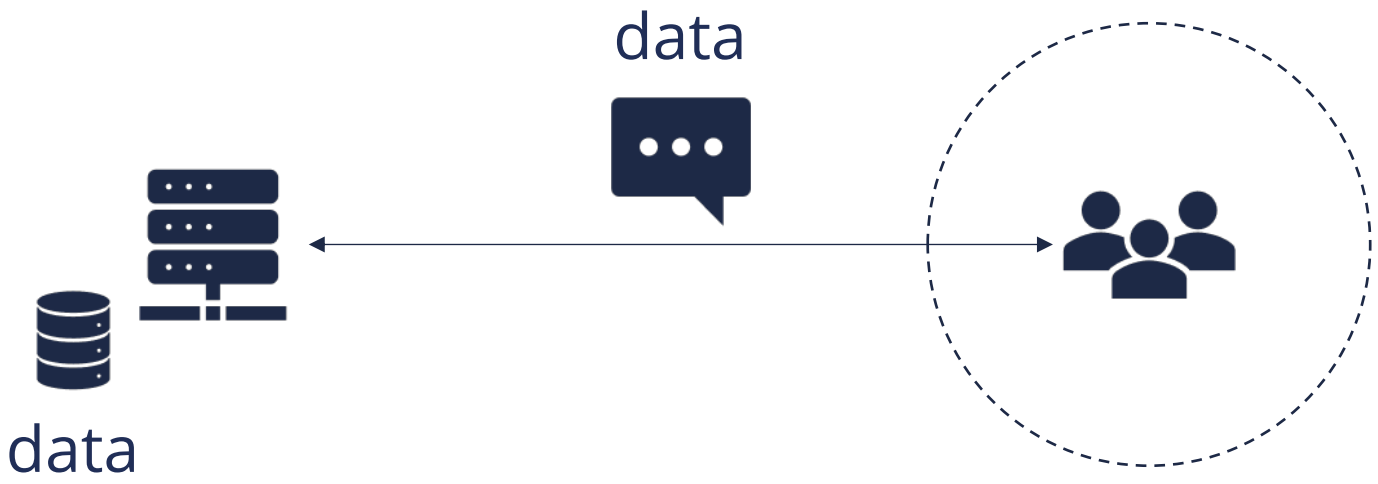
\+ Correct Token predicted

− Incorrect Token predicted

adjust the model!

Data Protection
Privacy
Legal Settings
Data Ownership

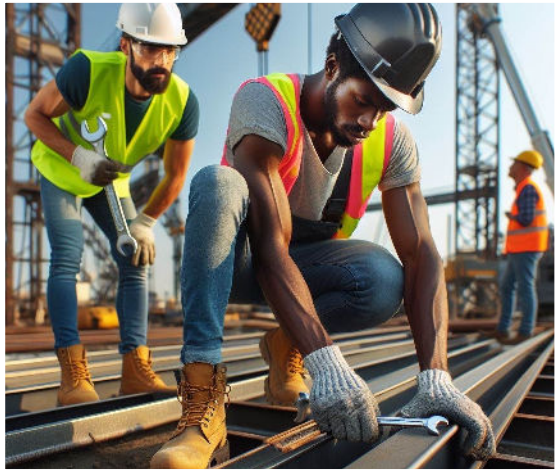
Better Norwegian Data
=
Better Norwegian



How to get the Model to do what you want?

Fine-tuning

Instructing



Alignment

MIMIR

National Library of Norway

UNIVERSITY OF OSLO

Data	Training	Evaluating
• Data set with open license • Data set with copyright protected material	• Building different models • IDUN	• Benchmarking • Measuring Benefit of copyrighted material

TrustLLM

Germanic Languages

Shared Culture and Values

Trustworthiness

Establishing Trust for Technology to Succeed

LLM Eco-system

Bringing Different Actors Together

The Role of NTNU

Build Large Models

Educate

Dialog with Partners

Get in Touch!

Benjamin Kille
benjamin.u.kille@ntnu.no