NorjordAI - AI chat about Norwegian agriculture	**Summary** Adapt LLM to the domain of Norwegian Agriculture to answer questions from farmers in Norwegian **Authors/Contributors:**

What is the question we want to answer? What is the best way to adapt LLM to the domain of Norwegian agriculture for the question-answering task?

Why is this question important? Because it can be useful for farmers, it can help them get relevant information

What work does this question build out? What papers should one read to catch up on the state of the literature relevant to this question?

Research papers to read:

- A dataset dedicated to the training of large- language models for agronomic management practices and production in Norwegian agriculture
- YourBench: Easy Custom Evaluation Sets for Everyone
- NLEBench+NorGLM: A Comprehensive Empirical Analysis and Benchmark Dataset for Generative Language Models in Norwegian
- Small Languages, Big Models: A Study of Continual Training on Languages of Norway
- Fluent Alignment with Disfluent Judges: Post-training for Lower-resource Languages
- RAG vs Fine-tuning: Pipelines, Tradeoffs, and a Case Study on Agriculture
- Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena
- Prometheus: Inducing Fine-grained Evaluation Capability in Language Models
- RAG-Driven Memory Architectures in Conversational LLMs—A Literature Review With Insights Into Emerging Agriculture Data Sharing
- The 3-billion fossil question: How to automate classification of microfossils
- Why Language Models Hallucinate

- ShizishanGPT: An Agricultural Large Language Model Integrating Tools and Resources
- The Art of Asking: Multilingual Prompt Optimization for Synthetic Data
- A Smart Agricultural Knowledge Management Framework To Support Emergent Farmers In Developmental Settings
- On the Theoretical Limitations of Embedding-Based Retrieval
- CogniLoad: A Synthetic Natural Language Reasoning Benchmark With Tunable Length, Intrinsic Difficulty, and Distractor Density
- Language Specific Knowledge: Do Models Know Better in X than in English?
- Tokenization Is More Than Compression

Is this a publishable question (has not been published before/can the findings be shared externally)? Yes, LLM for Norwegian agriculture has't been explored enough, and the results of experiments can be useful in building an AI chat for agriculture: either for farmers or for advisors. Moreover, a synthetic dataset of questions with answers can be useful to test other models. Additionally, these methods can be useful for building an updated dataset in the future.

What is the simplest experimental setting in which this hypothesis can be tested?

1. To create a synthetic QA dataset from the text data about Norwegian Agriculture
2. To test different Norwegian LLMs
3. To test the RAG method with different similarity functions
4. Report results

What are the key baselines and benchmarks we would need to compare against in your experiments?

- Benchmarks:
    1. Questions on a single topic
    2. Questions from 3 websites
- Baselines:
    - Test Norwegian LLMs before RAG

What datasets, models, eval, annotation, resources do we plan to use to

validate this hypothesis? If not relevant, you can answer N/A.

1) Datasets

- Covering a single topic
- Covering all the topics from 3 websites:
    - Generated with NotebookLM
    - Generated with Yourbench using a Norwegian model

2) Model

- Cohere models
- NorMistral-11b-thinking
- NorskGPT
- Viking
- Llama 2 13b Chat Norwegian from Ruter
- NorwAI

3) Evaluation Metrics

- <mark>Ragas metrics</mark>
    - Faithfulness (answer accuracy)
    - Answer Relevancy (answer-to-question fit)
    - Context Recall (retrieval of all needed info)
    - Context Precision (signal-to-noise in context
- Language Model Evaluation Harness
    - RAG Evaluation Harness
- LLM-as-a-judge
    - GPT score
- Word-based evaluation
    - Meteor
    - BLEU
    - ROUGE
- Embedding-based evaluation
    - <mark>Bert score</mark>
    - Mover score

4) Annotation Resources

- NotebookLM

- Yourbench

5) Evaluation Resources

- DeepEval
- EvalAssist
- AlpacaEval

Further work:
Knowledge assimilation: Implementing a knowledge-guided agricultural large language model