データリポジトリとLLM勉強会 趣旨説明

北本 朝展（国立情報学研究所）

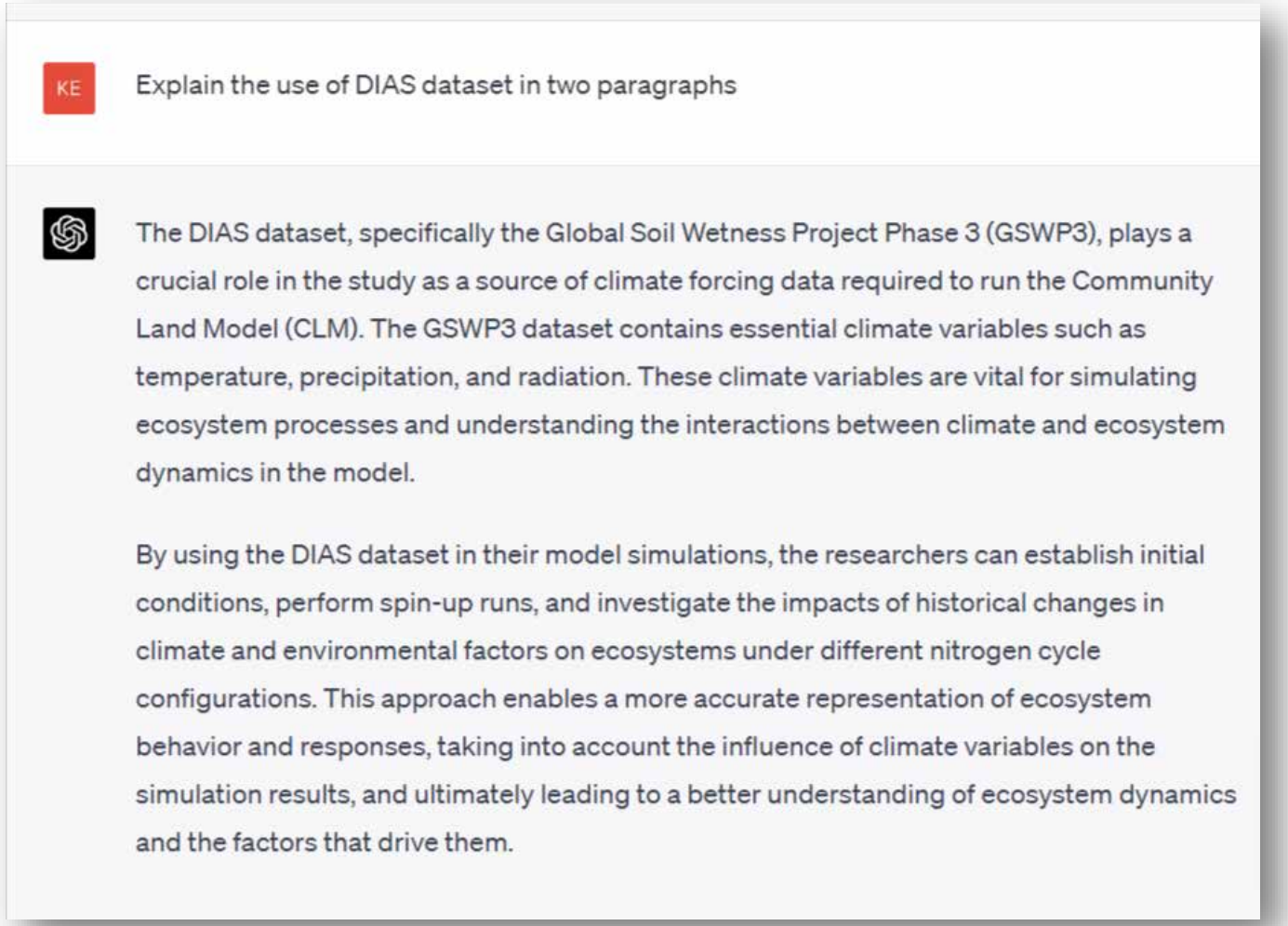
https://dias.ex.nii.ac.jp/

大規模言語モデル（LLM）

https://llm-jp.nii.ac.jp/

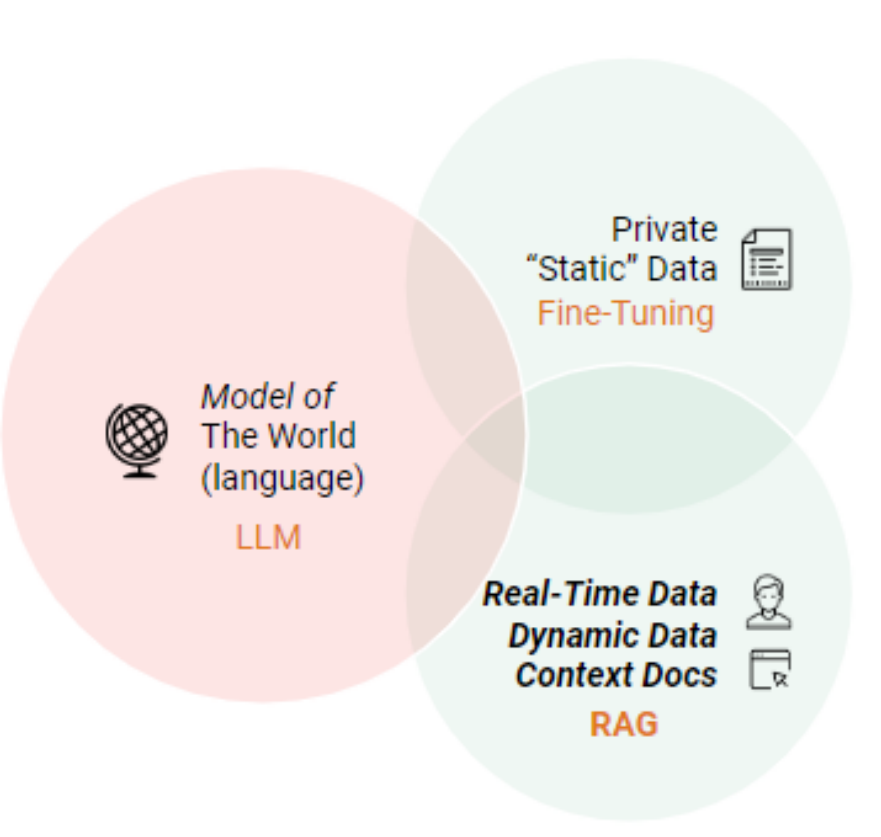
1. ChatGPT等の大規模言語モデルが社会に広く浸透
2. GPT（General-Purpose Technology）は全分野に影響
3. 日本の学術界でもLLMを構築するプロジェクト開始
4. データリポジトリにも大きな影響を与えるのは確実

タスクの効率化



1. メタデータ作成の効率化
2. データセット論文からのメタデータ自動生成
3. メタデータの翻訳
4. 関連論文の要約
5. データセット利用方法の整理

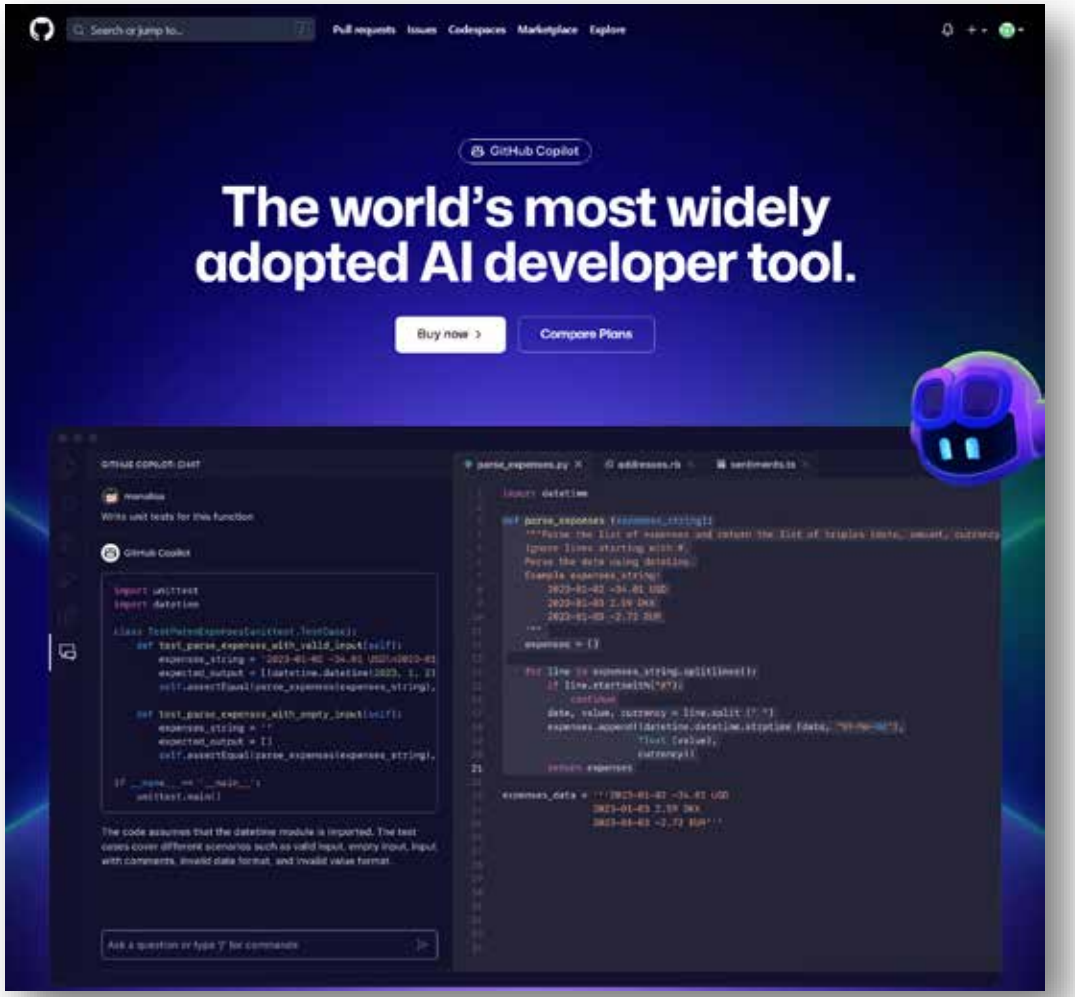
検索の高度化



https://www.hopsworks.ai/dictionary/retrieval-augmented-generation-llm

1. 自然言語文の直接入力による意味的な検索
2. 自然言語文をクエリ言語（例：SPARQL）に変換
3. 検索結果の整理や要約、複数DBを統合した検索結果表示
4. ユーザのレベルに応じた検索結果生成

利活用の円滑化



1. LLM拡張用のメタデータの付与
2. データセットのスキーマを反映したコード生成
3. データセット分析・可視化のためのコード生成

LLMの強みと弱み

1. LLMは「言語モデル」であり、自然言語／人工言語で表現された「言語の世界」を対象としている
2. 「言語の世界」で完結するタスクは強い（定型的なメール返信、一般的な悩み相談など）
3. 特定分野の正確な知識や、知識の更新が求められるタスクでは、外部のサービスと連携する必要がある
4. LLM単体で完結するタスクと、LLMを外部サービスで補強すべきタスクとを区別する必要がある

データリポジトリとLLM勉強会

https://dias.ex.nii.ac.jp/llm/

1. **データリポジトリを対象に、LLM（生成AI）をどのように活用するかを考える勉強会**
2. DIAS（Data Integration and Analysis System）の取り組みに閉じず、みんなでアイデアを共有する
3. 気軽に情報交換できる場とし、実験的な試みやコードについても共有する（参考：LLM勉強会）
4. 皆さんが興味をもつテーマについて、情報共有と気軽な議論が行えるようにしたい