

J-AgriQA: 農業水利施設に関する日本語データセットの作成と今後の展望
 Creation of a Japanese Dataset on Agricultural Water Management
 Facilities and Future Prospects

○箱石健太* 菅田大輔* 一言正之* 塩野隆弘* 工藤晶子* 小美野聡子* 市原東** 横山恵一**
 HAKOISHI Kenta, SUGETA Daisuke, HITOKOTO Masayuki, SHIONO Takahiro, KUDO Akiko,
 OMINO Satoko, ICHIHARA Azuma, YOKOYAMA Keiichi

1. はじめに

近年、生成 AI 技術の発展により我々人間が生み出す情報と AI が生成する情報の差がなくなりつつある。特に ChatGPT をはじめとする大規模言語モデル（以下 LLM）は人間が作成した文章と見分けがつかないほどの精度を誇っており、日々進化している。我が国の農業農村整備事業においては「農業農村整備に関する技術開発計画」を策定し、その中で AI 技術をはじめとしたデジタル技術の活用による省力化・効率化を目指している。このような背景を踏まえ LLM の農業農村工学分野での利活用が期待されるが、LLM がどの程度の精度があるのかを評価するには、主観的、定性的な評価をするほかに、定量的な評価を実施するためには、ベンチマークが必要である。ベンチマークの例として一般用語や常識の理解を評価するために JCommonsenseQA というデータセットを活用する手法がある。また土木分野といった特定分野に特化したデータセットを用いて評価も行われつつある。一方で農業農村工学の分野ではこのようなデータセットについては発表されていない。本研究では、農業農村工学分野における「農業水利施設の保全管理」を対象を絞り、LLM に対して評価可能なデータセットを作成し（以下 J-AgriQA）、対象データの解説、評価方法などとともに今後の展望について述べる。

2. 対象データ

「農業水利施設の機能保全の手引き」の資料を対象とした（表 1）。

表 1 対象データ一覧 (Table1 Target data list)

No.	資料名	更新日時	ページ数	QA 数
1	総論編	令和 5 年 4 月 26 日	101	256
2	パイプライン編	平成 28 年 8 月	84	85
3	開水路編	平成 28 年 8 月	90	101
4	頭首工編	平成 28 年 8 月	104	108
5	水路トンネル編	平成 28 年 8 月	120	130
6	ポンプ場（ポンプ設備）編	平成 25 年 4 月	77	80
7	除塵設備編	平成 25 年 4 月	66	67
8	電気設備編	平成 25 年 5 月	131	131
9	水管理制御設備編	平成 25 年 5 月	83	83
計			856	1,041

*日本工営株式会社 Nippon Koei Co., Ltd. **株式会社ライフスペース LifeSpice INC.

キーワード：農業水利施設，維持管理，生成 AI，大規模言語モデル，自然言語

3. J-AgriQA の作成

表 1 の資料から農業水利施設の保全管理に関する QA (質疑応答) タスクのデータセットを 1,041 件作成した。基本方針として 1 ページあたり 1QA とし、総論編については保全管理の基本的な考え方や用語等が記されているため、重点的に作成した。また QA の根拠を確認するために資料名、章名、QA の基となった文章、ページ数を含めた形として作成した。作成したデータは今後 GitHub などに誰しもがダウンロードできる形とし、PullRequest によるデータレビューが可能な形で公開する予定である。

4. J-AgriQA を用いた LLM の評価事例

J-AgriQA の一部のデータを抜粋し、一般的な LLM である ChatGPT (GPT-3/GPT-4) と農業水利施設に特化した LLM (以下 AgriAI) の 2 モデルを評価した。なお AgriAI は、Retrieval-Augmented Generation (RAG) の手法を用いて、「農業水利施設の機能保全の手引き」に関する資料を外部知識として ChatGPT に与えた AI である。一般的な性能をもつ ChatGPT と、特化した知識を持つ AgriAI で質疑応答にどのような回答精度の差が生じるかを確認することが目的である。データ例とそれぞれの回答結果を表 2 に示す。

表 2 J-AgriQA の一部のデータに対する正答と AI の回答結果
(Table2 Correct answers and AI answer results for some data of J-AgriQA)

No.	質疑内容例	正答	GPT-3/GPT-4	AgriAI
1	河川流域のあらゆる関係者が流域全体で行う治水対策を何と行うでしょうか。	流域治水	総合治水/流域管理	流域治水
2	農業水利施設の(略)ストックマネジメントの取組について、点検、補修、機能診断、施設監視、関係機関との役割分担などともに策定する計画とは？	機能保全計画	ストックマネジメント計画/維持管理計画	機能保全計画

表 2 の通り ChatGPT より農業水利施設に特化した AgriAI のほうが正しく回答していることがわかる。ただし本ケースはデータの一部を抜粋し実施しているに過ぎないため、本結果のみで AgriAI のほうが優れているとは言えない。

5. 今後の展望と課題

本研究は J-AgriQA の作成を行い、一部のデータを用いて ChatGPT と AgriAI の簡易的な性能確認を実施した。今後は J-AgriQA の全てのデータを利用し、詳細な LLM の精度検証を実施していく。これにより、LLM の定量的な評価が実施可能となり、農業農村工学分野での生成 AI の利活用の効果、課題が明確になることを期待する。

また対象データは「農業水利施設の保全管理」の一部の資料に限定しているため、さらにデータを拡充していく必要があるが、こういった資料・データは扱いにくいファイル形式で、かつ分散管理しているケースがあるため、これらを整理するための労力をいかに省力化するかが課題である。加えて資料が改訂されるたびに J-AgriQA を改訂する必要があるため、この維持管理についても検討していく必要がある。