

国産農業用 LLM のためのインストラクションデータ構築と構築された LLM システムの評価

石原潤一¹ 小林暁雄¹ 桂樹哲雄¹ 大友将宏¹ 橋本祥² 阪本浩太郎³
杉村安都武⁵ 米丸淳一¹ 安藤まや⁴ 後藤美智子⁴ 関根聡⁴ 川村隆浩¹
¹ 国立研究開発法人 農業・食品産業技術総合研究機構 ² 筑波大学大学院
³ 株式会社 BESNA 研究所 ⁴ 株式会社いちから ⁵ 三重県農業研究所
{ishiharaj612, akio.kobayashi, t.katsuragi, masahiro.otomo}@naro.go.jp

概要

農研機構は、内閣府「研究開発と Society 5.0 との橋渡しプログラム (BRIDGE)」における AI 農業社会実装プロジェクトにて、農業分野に特化した日本語大規模言語モデルの開発を進めている。本研究では、三重県農業研究所より提供を受けたイチゴに関するマニュアル類を元にインストラクションデータを構築を行い、これを用いて Elyza-8B モデルに対しインストラクションチューニングを実施した。本稿では、これらのデータおよびモデルの構築の概要について解説したのち、実際にこのモデルが専門知識を含んだ回答ができているかを LLM as a Judge にて評価する手法を提案する。

1 はじめに

昨今の不安定化する世界情勢の中で、国内での食料の安定供給を維持するため、2024 年の通常国会にて農政の憲法といわれる食料・農業・農村基本計画が 25 年ぶりに改正された。併せて、スマート農業を振興させる新たな法的枠組みも制定されたことで、スマート農業に関する新技術を普及させ、生産・流通・販売方式の変革などの取り組みを強力に促進させていくことが定められた。さらに、みどりの食料システム戦略や気候変動への対応など、近年になり次々と新しい技術が生まれている。しかし、国内就農者の平均年齢は 68.7 歳 (2023 年) [1] と高齢であり、今後 20 年間で現在の 3/4 ほどの就農者が離農するという予測が立てられている [2]。同様に、農業従事者に対して営農指導を行う国家資格を持つ普及指導員数も減少が進んでいる [3]。このような状況を打開しなければ、新規就農者の早期育成や、既存の農業者へのスマート農業やみどりの食料シス

テム戦略などに関する適切な情報提供が滞ると懸念されている。

そこで、農研機構は、内閣府「研究開発と Society 5.0 との橋渡しプログラム (BRIDGE)」[4] の一環として、日本の各産地や作目に応じた精緻な農業知識を正しく学習した農業用大規模言語モデル (LLM) の開発を進めている [5]。このプログラムは、官民研究開発の投資拡大が見込まれる領域における研究開発等を推進し、Society 5.0 への橋渡しを実現するための施策である。この施策のプロジェクトの 1 つである農林水産省実施施策「AI 農業社会実装プロジェクト」[6] では、農業従事者の減少による農業労働力の減少を補うための AI 技術の確立を目的としている。その中の実現目標課題の一つとして、地域の特性に応じた正確な知識と、スマート農業やみどりの食料システム戦略に関する新技術情報などを収集・整理し、それらを適切に引き出せる対話式システムとして、普及指導員が現場の状況に応じて効率的かつ適切に対策を判断するための補助的役割を担う能力を持った LLM (農業用 LLM) を構築することを目指す。農業に特化した LLM の構築は各国で進められている [7, 8, 9, 10] が、本研究では、既存の農業用 LLM にはない、普及指導の補助に焦点を当てたシステム構築を進めている。ここで想定する補助は、普及指導員が営農指導の際に農業者から問い合わせがあった際に、指導員自身の知識や観点が不足していたために適切な指導ができないような状況に対し、適切な知識や観点を農業用 LLM が回答することを目指している。この回答を得ることで、普及指導員が実際の圃場の状況なども加味したうえで適切な指導に結びつけられると期待される。また、今後指導に必要な知識を効率よく習得してもらう際にも有用となり得る。

このような農業用 LLM の第 1 弾として、三重県農業研究所と連携し、三重県産イチゴに関するモデルの構築を行った。このモデルを構築するにあたり、まず、三重県農業研究所が独自に作成したイチゴの栽培マニュアル（三重県イチゴマニュアル）を用いて、普及指導員をサポートする振る舞いを表現するためのインストラクションチューニング用データを構築した。

モデル構築では、この三重県イチゴマニュアルと農研機構が構築した農業に関する百科事典である農業技術辞典 NAROPEDIA[11] を用いて Elyza-8B[12] モデルに事前学習を行い、インストラクションデータによるインストラクションチューニングを行った。さらに事前学習に用いたデータを RAG[13] に用いることで、三重県産イチゴに特化したモデルを実現した。

また、同モデルについて、2024 年 10 月 21 日よりイチゴ農家を普及指導員が指導を行う際の補助としての実用性を調査するための試験運用を開始した。

本稿では、まず本農業用 LLM のモデルの構成に言及し、具体的に三重県イチゴマニュアルからのインストラクションデータの構築手順や構築されたデータの詳細について解説する。次に、普及指導員が現場で尋ねられる質疑応答集を独自に作成し、普及指導員の回答を規範として、本農業用 LLM の回答と比較する評価手法を解説する。並行して、既存の大規模自然言語モデルでも同様の実験を行い、モデル間の比較を行うことで、本農業用生成 AI がどのくらい普及指導員を模倣できるか相対的評価を行った。

2 農業データの集積とモデルの学習

農研機構は全国の公設試験研究機関や農業法人などと協力し、農研機構内外から農業に関する膨大なデータを収集し、農研機構が管理する大規模データ共有サービスである、農研機構統合データベース[14] への集約を進めている。

BRIDGE「AI 社会実装プロジェクト」で収集されるデータも同データベースへの保管が進められている。これらのデータには、NAROPEDIA の情報や、地方公共団体の公設試験研究機関や JA などが管理する栽培マニュアルや栽培歴、営農指導記録などの情報が含まれている。三重県イチゴマニュアルも同データベースに格納され、当該プロジェクトコンソーシアムのオープン・クローズ戦略に基づいてコ

ンソーシアムメンバー内で共有されている。

2.1 インストラクションデータ構築

三重県イチゴマニュアルに基づくインストラクションデータの構築について解説する。まず、三重県イチゴマニュアルとしては、三重県農業研究所より、イチゴ全般に関する三重県での高設栽培についてのマニュアルと、三重県が開発した品種である「かおり野」についての栽培マニュアルの二つを用いた。それぞれ、イチゴの栽培方法だけでなく、土壌整備や病害虫防除に関する情報が記載されており、「かおり野」は図表込み 1 カラムの 10 ページ、「高設栽培」も図表込み 1 カラムの 232 ページとなっている。ここからイチゴ栽培に関して実際に普及指導員が農業者から受けるような質問と、それに対し回答を行う際に参考となるような回答例を構築し、三重県イチゴモデルへのチューニングに利用する。

この際に、三重県農業研究所の協力の下、普及指導員に対しアンケートを行い、実際に農業者から受けたイチゴに関する質問のうち、すぐには回答できず、後ほど調査を行い回答したような質問を収集した。また、このような質問だけではイチゴ全般に関する質問を網羅できないため、三重県イチゴマニュアルなどを参照しつつ、アノテーターが追加の質問を作成した。農業知識を持つアノテーターを集めることは難しいため、豊富な農業知識を持つアノテーターが他のアノテーターの作成した質問に対しチェックを行い、本課題にとって適切であるものだけを採用した。

また、質問に対する回答は、構築の際の基準としたデータに基づき、以下の 3 種類に分けられる。

「かおり野」マニュアルに基づく回答 かおり野品種に関する質問のうち、マニュアルから回答に関する情報が取得できたもの

「高設栽培」マニュアルに基づく回答 上記以外で「高設栽培」マニュアルから回答に関する情報が取得できたもの

その他イチゴに関する情報に基づく回答 上記以外で、各種イチゴ栽培に関するウェブサイトなどから回答が取得できたもの

これら 3 種類について、アノテーターが回答の構築を行った。期間は約 3 か月程で、構築されたデータ数は延べ 506 件となった。

回答の構築では、まず、質問に対しどのような回

答が必要なのかについてアノテーターが想定することが難しかったため、事前にアノテーター間で具体例を共有することで、回答文章の品質を担保した。回答精度については、先述した農業知識のあるアノテーターが質問同様にチェックを行うことで担保した。また、三重県イチゴマニュアルが非公開データであるため、マニュアルに基づくデータについては、マニュアルと同一の文言を避けるなど、著作権に配慮して構築を行った。三重県イチゴマニュアル以外の情報については、農業従事者からの質問は、三重県に限らず発生しうる質問が想定されること、また、圃場の状況によっては同種の問題であっても解決アプローチが異なることが想定されるため、三重県のイチゴ栽培にこだわらず情報を収集し、質問・回答データの構築を行った。それぞれ構築できた件数を表 1 に示す。表より、実際の質問に基づくデータの件数が全体の 1 割程度となっている。また、三重県イチゴマニュアル 2 種から回答が作成されている割合は多いものの、提供されたマニュアル以外からも情報を収集する必要があったことが分かる。また、新規作成した質問については、その他イチゴに関する情報に基づく回答が最も多くなっている。これらの結果から、三重県イチゴマニュアルは三重県での栽培に特化した様々な情報が含まれているものの、それ単体でイチゴ栽培に関する様々な質問に対応するようなインストラクションデータを構築するには不足している結果となった。

表 1 インストラクションデータ内訳

質問データ種別	回答データ出典	構築件数
実際の質問	「高設栽培」マニュアル	39
	「かおり野」マニュアル	18
	その他の情報	10
	合計（出典重複有）	55
新規に作成	「高設栽培」マニュアル	88
	「かおり野」マニュアル	100
	その他の情報	267
	合計（出典重複有）	451

2.2 モデルの構築

本農業生成 AI は、国産の事前学習済み日本語大規模言語モデル (Llama-3-ELYZA-JP-8B) をベースとして、NAROPEDIA を用いて継続事前学習を行ったものである。継続事前学習は、5,493 件のデータに対し、バッチサイズ 900 で実施した。Llama-3-ELYZA-JP-8B から継承した性能を破滅的忘却によって失うことを避けるため、損失関数の減少

率が一定以下に収束した 96 イテレーションのモデルを採用した。インストラクションチューニングでは、構築したインストラクションデータから 9 割をランダムサンプリングして使用し、バッチサイズ 480 で 256 イテレーションの学習を行った。継続事前学習およびインストラクションチューニングの両方では、データセットの規模が小さいことを考慮し、学習率を $1e-5$ 、ウォームアップステップを 10 イテレーションに設定した。学習は NVIDIA V100 GPU を 48 台搭載した計算環境で実施した。

また、このモデルをベースとして、別途 NAROPEDIA と三重県イチゴマニュアルを用いて構築した RAG システムも構築し、より正確かつ具体的な回答を行えるチャットシステムを構築した。このシステムは三重県のイチゴ栽培に特化した回答を行えるシステムとなっているため、このシステムの性能を適切に評価するために、既存の評価手法を当てはめることは難しい。このため、新たに本システムのような特定ドメインの知識の要点を抑えているかどうかを評価するための評価手法を考案した。

3 システムの評価

3.1 評価の方法

特定ドメインに特化したモデルでは、抑えておくべき要点となる専門知識が含まれている回答が良い回答である可能性が高いと考えられる。特に、本システムでは、普及指導員が農業者の質問に対して回答する際に、抑えておくべき要点が網羅されていることが好ましい。

そこで、既存の LLM を用いてテストデータ（以降、正解）およびシステムの回答（以降、システム出力）を、小さな意味単位（ナゲット）[15, 16] に分割した。これらのナゲット群を抑えておくべき要点の一覧と捉え、正解のナゲット中の何割がシステム出力に含意されているか、システム出力中の何割が正解に含意されているかを既存の LLM により求め、それぞれの割合を Recall、Precision として捉え、F1 値を求めた。この手法により、特定ドメインに特化したシステム出力に対して、そのドメイン知識を含んだ回答ができていないか否かを評価することが可能となる。

ナゲット分割については、述語項構造をナゲットと捉え、述語項構造解析を行うようなプロンプトを作成し（付録 A.1.1）、LLM に出力させた。ナゲット

分割と含意関係認識の例を表 2 に示す。本システムの回答精度を既存の LLM の精度と比較するため、ベースモデルとして使用した Llama-3-ELYZA-JP-8B および、OpenAI GPT-4o mini[17] にも回答を出力させた。そのうえで、上記で得られた回答ごとの評価方法を、これらの既存モデルの回答にも適用し、macro-averaging と micro-averaging をそれぞれ算出することで、本システムと各モデルの回答精度を比較した結果を表 3 に示す。ナゲット分割および含意関係認識についても、OpenAI GPT-4o mini を用いた。

表 2 正解とシステム出力のナゲットの比較

【質問】
イチゴの育成に好適土壌かどうかは pH 値を管理すればよいか？
【正解】
イチゴの育成ステージに合わせて好適土壌の水分は変化します。多く水分が必要なのは育苗期で、収穫期では最も少なくなります。その土壌水分の吸収は pH 値で管理できます。...(中略)... また、育苗期や定植初期の段階の水分の多少が、根の発生に大きく影響することも忘れてはなりません。
【正解のナゲット】
<ul style="list-style-type: none"> ・好適土壌の水分は育成ステージに合わせて変化する ・育苗期では多くの水分が必要で、収穫期では最も少なくなる ・水分吸収は根域の通気性や土壌溶液濃度にも関係する ・育苗期や定植初期の水分の多少が根の発生に影響する ・土壌水分の吸収は pH 値で管理できる
【システム出力】
イチゴの生育に適した土壌の pH は 5.5~6.5 です。pH だけではなく、土壌の施肥状態や土壌構造、土壌菌叢など多くの要素が関与します。直感的にイチゴの育成に好適土壌か否かは pH 値のみで判断せず、土壌診断を行うとよいでしょう。...(中略)... 生育ステージによって好適水分範囲は異なり、育苗期が最も多く、収穫期が少ないことを念頭に管理する必要があります。
【システム出力のナゲット】
(含意されていたシステム出力ナゲット例)
<ul style="list-style-type: none"> ・イチゴの育成に好適土壌か否かは土壌診断を行うとよい ・培地の土壌溶液濃度は作物に大きな影響を与える ・水分管理は重要で、根域の通気性や土壌溶液濃度を考慮する ・適切な土壌水分は生育段階によって異なる ・水分吸収は pH 値以外に根域の通気性や土壌溶液濃度と関連する
(それ以外のシステム出力ナゲット例)
<ul style="list-style-type: none"> ・イチゴの生育に適した土壌の pH は 5.5~6.5 である ・土壌の施肥状態や土壌構造、土壌菌叢などの要素が関与する ・イチゴ高設用培土に全国農業協同組合の提案条件が求められる

3.2 評価の結果と他モデルとの比較

表 3 に示すように、micro f1 と macro f1 の値を見ると、現段階では GPT-4o mini のスコアが最も高く、次いで本システムの回答精度が高くなった。ここで、本システムの回答精度がベースモデルとして使用した Llama-3-ELYZA-JP-8B の F1 値より優れていたことは、本システムの構築が目的通り行われていたことの証左といえる。また、Recall はともに本システムが最も高くなっており、本システムが多くの要点を出力していたことが分かる。表 2 の例のよう

表 3 モデルの評価

	本システム出力	GPT-4o	Elyza8B
macro f1	0.181	0.223	0.142
macro precision	0.190	0.426	0.294
macro recall	0.263	0.203	0.140
micro f1	0.216	0.276	0.196
micro precision	0.202	0.389	0.293
micro recall	0.233	0.214	0.148

に、本システムは具体的な値や詳細な要点など、正解にも含まれていないような情報も出力しており、内容の正誤は確認する必要があるものの、本研究目的に沿った回答をしていると考えられる。

なお、本システムの回答精度が GPT-4o mini の回答精度を下回った理由の 1 つに、前者の具体性の高さが挙げられる。例えば、「イチゴが通常より小さい上に硬くなっている。栽培方法に問題があったのか?」という質問に対し、本システムの出力のナゲット「病気が原因で果実が小さくなる」「萎黄病が原因で果実が小さくなる」「炭疽病が原因で果実が小さくなる」が、正解のナゲット「果実が小さくなる」に含意できなかった。本研究の目的である、普及指導員の補助という観点で考えると、これらのシステム出力におけるナゲットを一律に不正解としてしまうのは目的から外れてしまっていると思われる。このため、今後の課題としては、例えば正解のナゲット「果実が小さくなる」に含意されると判断されるようなプロンプトチューニングを含意判定用の LLM に実施するなどの対応を検討したい。

4 おわりに

本稿では、BRIDGE プロジェクトの一環である農林水産施策「AI 農業社会実装プロジェクト」を実装させるべく、普及指導員の営農指導を補助する農業用生成 AI の構築およびその精度評価を報告した。生成 AI の社会実装では、データの量と質が非常に重要となる。農研機構は、今後も全国の各産地と協力し、農業データのさらなる収集を進め、他作目や他県にも横展開していく予定である。そして、真に利用されるモデルを構築するため、本稿で紹介したモデル評価方法は、その一助となる。現段階では、本農業生成 AI の回答精度は、GPT-4o mini の精度より低く、また安全性の評価などに課題が残る。本農業生成 AI の精度をさらに高めていくため、農業現場において、就農者および普及指導員の使用を通じて得た情報の正誤や使い勝手などのフィードバックを反映させる予定である。

謝辞

本研究は、内閣府「研究開発と Society 5.0 との橋渡しプログラム (BRIDGE)」における農林水産省実施策「AI 農業社会実装プロジェクト」の助成を受けて実施された。本研究の一部は農研機構の AI 研究用スーパーコンピュータ「紫峰」を利用して実施した。

参考文献

- [1] 農林水産省大臣官房統計部経営・構造統計課センサ統計室. 農業労働力に関する統計. <https://www.maff.go.jp/j/tokei/sihyo/data/08.html>.
- [2] 農林水産省農林水産技術会議事務局. 農林水産研究イノベーション戦略 2024. <https://www.affrc.maff.go.jp/docs/press/attach/pdf/240604-2.pdf>.
- [3] 農林水産省農産局技術普及課. 協同農業普及事業をめぐる情勢. <https://www.maff.go.jp/j/seisan/gizyutu/hukyu/h.about/attach/pdf/index-1.pdf>.
- [4] 内閣府科学技術・イノベーション推進事務局. 研究開発と society 5.0 との橋渡しプログラム bridge. <https://www8.cao.go.jp/cstp/bridge/index.html>.
- [5] 桂樹哲雄, 小林暁雄, 坂地泰紀. 農研機構における農業用生成 ai 構築の取り組み. 日本語言語資源の構築と利用性の向上 - JLR2024 ワークショップ, 2024.
- [6] 農林水産省大臣官房政策課技術政策室情報化推進班. AI 農業社会実装プロジェクト. https://www8.cao.go.jp/cstp/bridge/keikaku/r5-20_bridge_r6.pdf.
- [7] Ruoling Peng, Kang Liu, Po Yang, Zhipeng Yuan, and Shunbao Li. Embedding-based retrieval with llm for effective agriculture information extracting from unstructured data, 2023.
- [8] Ranjan Sapkota, Rizwan Qureshi, Syed Zohaib Hassan, John Shutske, Maged Shoman, Muhammad Sajjad, Fayaz Ali Dharejo, Achyut Paudel, Jiajia Li, Zhichao Meng, Ferhat Sadak, Muhammad Usman Hadi, and Manoj Karkee. Multi-modal llms in agriculture: A comprehensive review. September 2024.
- [9] Matheus Thomas Kuska, Mirwaes Wahabzada, and Stefan Paulus. Ai for crop production – where can large language models (llms) provide substantial value? **Computers and Electronics in Agriculture**, Vol. 221, p. 108924, 2024.
- [10] Saed Rezayi, Zhengliang Liu, Zihao Wu, Chandra Dhakal, Bao Ge, Chen Zhen, Tianming Liu, and Sheng Li. Agribert: Knowledge-infused agricultural language models for matching food and nutrition. In Lud De Raedt, editor, **Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22**, pp. 5150–5156. International Joint Conferences on Artificial Intelligence Organization, 7 2022. AI for Good.
- [11] 独立行政法人農業・生物系特定産業技術研究機構. 最新農業技術事典 NAROPEDIA. 農山漁村文化協会 (農文協), 2006.
- [12] Masato Hirakawa, Shintaro Horie, Tomoaki Nakamura, Daisuke Oba, Sam Passaglia, and Akira Sasaki. <https://arxiv.org/abs/2401.12345>, 2024.
- [13] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In **Advances in Neural Information Processing Systems**, Vol. 33, pp. 9459–9474, 2020.
- [14] Takahiro Kawamura, Tetsuo Katsuragi, Akio Kobayashi, Motoko Inatomi, Masataka Oshiro, and Hisashi Eguchi. Development of an information research platform for data-driven agriculture. **International Journal of Agricultural and Environmental Information Systems (IJAEIS)**, Vol. 13, No. 1, pp. 1–19, 2022.
- [15] Ani Nenkova and Rebecca Passonneau. Evaluating content selection in summarization: The pyramid method. In **Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004**, pp. 145–152, Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics.
- [16] Teruko Mitamura, Hideki Shima, Tetsuya Sakai, Noriko Kando, Tatsunori Mori, Koichi Takeda, Chin-Yew Lin, Ruihua Song, Chuan-Jie Lin, and Cheng-Wei Lee. Overview of the NTCIR-8 ACLIA tasks: Advanced cross-lingual information access. In Noriko Kando, Kazuaki Kishida, and Miho Sugimoto, editors, **Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access, NTCIR-8, National Center of Sciences, Tokyo, Japan, June 15-18, 2010**, pp. 15–24. National Institute of Informatics (NII), 2010.
- [17] Josh Achiam et al. GPT-4 Technical Report. **arXiv e-prints**, p. arXiv:2303.08774, March 2023.

A 付録 (Appendix)

A.1 ナゲット抽出プロンプト

A.1.1 システムプロンプト

あなたは日本語の文章に対して述語項構造解析を行い、述語項構造から単文を作るアノテータです。あなたは、例えば、次のようなタスクが実行できます。

Task: ""述語項構造解析""

Input: ""アメリカの戦争に興味があり、本屋を回った。""

Output: ""[

(述語: 戦争 (事態性名詞), ガ格: アメリカ),

(述語: 興味があり (動詞), ガ格: 私, 二格: アメリカの戦争),

(述語: 回った (動詞), ガ格: 私, ヲ格: 本屋),

]""

Task: ""述語項構造からの単文生成""

Input: ""(述語: 戦争 (事態性名詞), ガ格: アメリカ)""

Output: ""アメリカが戦争する""

Task: ""述語項構造からの単文生成""

Input: ""(述語: 興味があり (動詞), ガ格: 私, 二格: アメリカの戦争)""

Output: ""私がアメリカの戦争に興味ある""

Task: ""述語項構造からの単文生成""

Input: ""(述語: 回った (動詞), ガ格: 私, ヲ格: 本屋)""

Output: ""私が本屋を回った""

Task: ""文章からの単文抽出""

Input: ""アメリカの戦争に興味があり、本屋を回った。""

Output: ""[

"アメリカが戦争する",

"私がアメリカの戦争に興味ある",

"私が本屋を回った",

]""

A.1.2 ユーザプロンプト

以下の文章から、述語項構造を抽出し、その述語項構造から単文を生成して、生成した単文のみを返してください。

文章: ""{text}""

A.2 含意関係認識

文章は Nugget に近い内容を含んでいますか？

文章: ""{text}""

Nugget: ""{nugget}""