

言語処理学会第30回年次大会 併設ワークショップ JLR2024  
「日本語言語資源の構築と利用性の向上」  
2024年3月15日  
LT

## 農研機構における農業用生成AI構築の取り組み

農研機構 農業情報研究センター 上級研究員 桂樹哲雄  
農研機構 農業情報研究センター 主任研究員 小林暁雄  
北海道大学 大学院情報科学研究院 准教授 坂地泰紀

## BRIDGE AI課題について

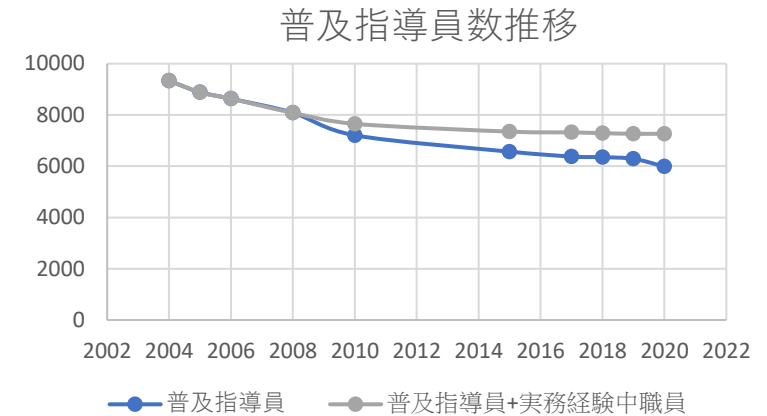
本研究課題は、内閣府のBRIDGEの中で実施している「AI農業社会実装プロジェクト」の一部。

- 内閣府「研究開発とSociety5.0との橋渡しプログラム(BRIDGE)」
  - <https://www8.cao.go.jp/cstp/bridge/index.html>
- 農林水産省が実施する施策「AI農業社会実装プロジェクト」
  - サブ課題1: 自然環境下での葉菜類の生育収量シミュレータの開発
  - サブ課題2: 施設園芸における果菜類の収量・品質制御技術の開発
  - サブ課題3: 食料安全保障に向けた主要穀物の安定多収化技術の開発
  - サブ課題4: 普及指導員・営農指導員向け生成AIの開発
- 2023年11月から実質スタート

# 普及指導員・営農指導員について

## 普及指導員（都道府県）

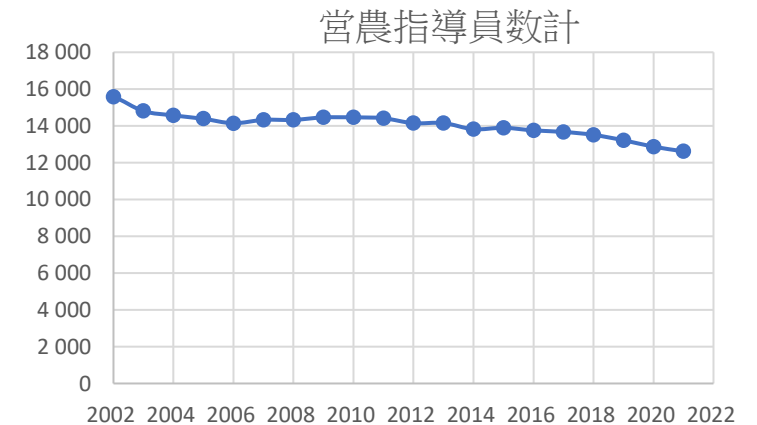
- 国家資格を持つ都道府県の職員
    - 令和元年(2019)4月現在6,102人
    - （実務経験中職員等を合わせて7,293人）
  - 農業者と直接接して以下のことを実施
    - 農業技術指導
    - 経営相談
    - 農業に関する情報提供
- 技術や広域的な情報に強みを持つ



※協同農業普及事業をめぐる情勢（農林水産省 R2.12 生産局 技術普及課）等より作成

## 営農指導員（農協）

- 農協の営農指導業務を担当する職員
    - 2021年度12,621人
  - 組合員の営農支援活動を実施
    - 農業技術指導
    - 出荷計画（定植前に計画、出荷時には計画との差分を調整）
    - 技術の導入指導
    - 農地利用調整等
- 地元密着で営農・販売に強みを持つ



※e-stat(総合農協統計表)より作成

## 背景・目的

- **背景**：現場にAIを広めるためには**普及指導員(都道府県)・営農指導員(農協)の存在が欠かせない(ラストワンマイル)**。普及指導員は全国で7000人 → 毎年200人程度減少
- **目的**：普及指導員・営農指導員数の減少が深刻であり、**AIチャットボットで支援**する。

## 概要

- **モデルの開発**：日本語汎用生成AI（大規模言語モデル）に対し、農業知識データを用いた**ファインチューニング**を実施し、「**普及指導員・営農指導員向けAI**」を開発する。**農業共通のAI**と、**各地域・作目に特化したAI**とを開発する。
- **データ収集**：全国の普及員向け掲示板や問題集、栽培マニュアル、SOP、栽培暦等の**農業知識データを収集**する。
- **アプリケーション開発**：開発したAIモデルを**WAGRIから提供**し、**チャットアプリ等に実装**する。
- **フィードバック**：チャットアプリを**ユーザ約5000名に無償提供**し、**正誤フィードバックを収集**する。また、**公設試の協力**により、**正解データを収集**する。これらのフィードバックは**AIモデル開発に反映**する。

### 生成AIモデルの開発

- データ収集
  - 全国の普及員向け掲示板
  - 普及員向け問題集
  - 栽培マニュアル
  - 農研機構SOP
  - 栽培暦等
- ファインチューニング
  - 日本語汎用モデル → 全国共通農業モデル
  - 全国共通農業モデル → 各地域・作物特化モデル



AIスパコン紫峰



### フィードバックによる改良

- チャットツール開発
  - WAGRIへのモデルの実装
  - チャットツールにモデルを組み込み
- チャットツールを利用した改良
  - 5000名のユーザによる正誤フィードバック
  - 公設試の協力による正解データ
- AIモデルへの反映



チャットツール

開発は「生成AIモデルの開発」と「フィードバックによる改良」の繰り返し

# データを集めています！

## 集めるデータ

- 栽培技術ドキュメント
  - 各都道府県公設試、JAの栽培マニュアル、栽培暦
  - 新聞・図書
    - 日本農業新聞、農文協（農山漁村文化協会）など。交渉中。
- 栽培・防除管理データ
  - 栽培日誌、業務報告
- 出荷・販売データ
  - JAからはなかなか出してもらえない。政府統計の総合窓口(e-stat)や農水省のデータなどから補完？

## 現状

- Webクロールは一旦完了
  - 全国の普及員向け掲示板、普及指導員資格試験問題集、栽培マニュアル、農研機構SOP、農研機構病虫害カタログデータ、病虫害年報、各種メディア等
- 共同研究機関、協力機関からのデータ提供（敬称略）
  - JAおきなわ、三重県農業研究所、石川県農林総合研究センター、佐賀県農業試験研究センター、岐阜県農業技術センター
- データ提供の交渉予定（敬称略）
  - 群馬県、栃木県、北海道、千葉県、愛媛県、宮崎県、宮城県、鹿児島県、十勝農業協同組合連合会、

## 営農支援ツールFARMCHAT/FARMBOXを利用

- いずれもソフトビルが提供する営農支援ツール
  - JAおきなわでは2017年から運用
- FARMBOX
  - 営農指導者向けアプリ
  - 巡回記録を投稿する形式
  - グループ間で共有可能
- FARMCHAT
  - 農業者向けアプリ。チャット形式で生産者とやりとり



## JAおきなわから提供されるデータ例(予定)

- データ利用契約の締結最終段階
- 栽培マニュアル（かぼちゃ、マンゴーから）
  - JAおきなわの営農指導員ではなく、元沖縄県の普及指導員が作成し、JAに提供
    - 栽培に関する知識は県の普及指導員の方が豊富。退職後ならば協力しやすいという背景もある？
- 巡回記録
  - 指導員が圃場に行って記録。ほぼ写真付き。→マルチモーダルなモデルが必要？
  - 約40万件ほど。実際に使えるのはその1/4程度？
- その他データ
  - 販売に関するデータは現状使用不可 → 市況データなどから補足？



## 栽培データはどこに？

- 農研機構にあるでしょ？
  - LLMに使えるデータはほとんどない。
    - 研究者は自分の研究（育種、栽培）にしか興味がない。
      - マニュアル等があるのは農研機構の育成品種のみ
    - 形質データなどはそもそもたくさんとれるものではない。
  - 誰かにデータを渡すことに抵抗がある。
- 都道府県の公設試やJAにあるのでは？
  - あるところにはある。
  - ただ、出してくれない。特に、他の団体（都道府県、JA等）が絡むと出しづらい。

## 日本の農業の構造自体の問題

- 隣の県、隣のJAは敵
  - 他県に資するかもしれないものは共有しない（つらい過去がある模様）
- 事業者の野心が小さい傾向にある
  - 中小規模が多い → ある程度事業が安定し、収入を得られるようになったらそこで満足する
  - データを集めて大規模化するという野心が小さい → データが集まらない。

こんな状況の中データ集めとモデル開発を頑張ってます。ご協力よろしくお願いします。