# A Novel Approach for Leveraging Agent-Based Experts on Large Language Models to Enable Data Sharing Among Heterogeneous IoT Devices in Agriculture

Nur Arifin Akbar(✉), Biagio Lenzitti(✉), and Domenico Tegolo(✉)

Dipartimento Matematica e Informatica, Universitá degli Studi di Palermo, 90123 Palermo, Italy
{nurarifin.akbar,biagio.Lenzitti,domenico.tegolo}@unipa.it

**Abstract.** The rapid adoption of Internet of Things (IoT) devices in agriculture has led to the generation of diverse data types, creating challenges in data sharing and integration across heterogeneous platforms. This paper presents a novel approach to facilitate data sharing among heterogeneous IoT devices in agriculture using agent-based experts built on large language models (LLMs).

Background: Traditional methods of data sharing in agriculture face limitations due to the lack of standardization and interoperability among IoT devices. Previous approaches, such as model fine-tuning and prompt engineering, have shown promise but struggle with open-ended agricultural queries and context comprehension.

The proposed Agent-based Data Sharing (ADS) framework combines semantic web technologies with agent-based design and LLMs to enable seamless information exchange, decentralized data sharing, and knowledge transfer through intelligent expert agents. This approach leverages the strengths of LLMs in understanding text and their extensive training data while addressing the challenges of data interoperability and context-aware decision-making in agriculture.

Using synthetic agricultural data, we evaluated the framework's performance in disease diagnosis and precision farming recommendations. The results demonstrate significant improvements in data integration, interoperability, and decision-making efficiency. With extensive data sharing, mean performance scores increased by 16% for disease diagnosis and 25% for precision farming compared to baseline scenarios.

The framework's ability to manage diverse devices and handle natural language queries through agent-based experts highlights its potential for real-world agricultural applications. This approach could support the advancement of smart farming through IoT applications and pave the way for improved efficiency in sustainable agriculture. However, challenges such as data privacy, standardization, and incentive structures need to be addressed in future research.

**Keywords:** large language models · agriculture · expert agents · data sharing · heterogeneous devices · agent-based systems


A. Artale et al. (Eds.): AIxIA 2024, LNAI 15450, pp. 12–22, 2025.
https://doi.org/10.1007/978-3-031-80607-0_2

## 1 Introduction

The advancement of technology, specifically the incorporation of IoT devices within agriculture, has made it possible to collect a considerable volume of heterogeneous data. However, the standardization and interoperability issues of these devices prevent us from meeting the main goal of data sharing and exchange. Large language models can help fill the gaps because they have the ability to understand and process natural language.

In this capacity, LLMs can be very effective for LLM-based agricultural applications, tapping into their ability to read beyond the text and answer questions accurately. However, for some specific applications, such as crop yield prediction or disease recognition in crops, their value is proportional to the number of examples that can be collected and how deeply the matrix understands the specifics of the problem at hand. Implementing LLMs to answer open-ended agricultural questions depends on how rich the examples provided are and how well the models understand the situation at hand.

Model fine-tuning is one of the most common ways to adapt LLM for agricultural problem solving [7]. However, existing methodology, such as prompt engineering and in-context learning, have been recently developed and applied in some cases [1, 14]. These techniques strive to increase the performance of LLM while addressing concerns such as data security and bias reduction issues [22]. Prompt engineering generates requests or directions to shape the output of LLM, thereby increasing the likelihood of accurate output in various tasks [6].

Nevertheless, these approaches tend to have many areas for improvement in responding to unstructured agricultural questions or issues and evaluating context comprehension. Giving more problem descriptions may also help block the agent's comprehension, and due to the limited variety of external agricultural knowledge bases available, there may also be limitations in the range of consultable resources.

This paper will present a new approach to agriculture data sharing by proposing an agent-based data sharing (A.D.S.) framework, which uses agents for data sharing between heterogeneous IoT devices based on generative AI on a large language model. Our primary contributions to this paper are as follows:

a. *Introducing the ADS Framework:* The combination of agent-based systems and LLM enables decentralized information and knowledge transfer among agricultural IoT devices.
b. *Improving Data Interoperability:* Using the strengths of LLM and agent-based systems to overcome data interoperability issues and contextually driven agricultural decision-making processes.
c. *Experimental Validation:* We illustrate the A.I. through experiments on agronomically sworn synthetic data, and show how the problem diagnosis or precision farming recommendations improve when agents share the data in the ADS. Paradigm.

# 2 Related Works

## 2.1 In-Context Learning for Agricultural Applications: Capabilities and Limitations

In-context learning in agriculture has the potential to improve its energy efficiency, resource allocation, and promote environmentally friendly attitudes through smart agricultural activities. Context learning can interact with various channels such as IoT sensors or intelligent monitoring systems, thereby enhancing data collection, reasoning, and decision-making for farmers. This may result in better production yields, less wastage of resources, and positive stewardship of the environment [23].

There are constraints on the use of in-context learning in the field, particularly in its application to agriculture. One major limitation is the issue of data standards, and interoperability is another. Agricultural databases are heterogeneous in terms of data format, quality, and availability. In order to maintain the quality and credibility of the learning setups, combining and using these dataset for agricultural decision-making requires significant effort in data processing and standardization procedures [23].

## 2.2 Retrieval-Augmented Generation: Enhancing LLMs with External Agricultural Knowledge

The retrieval-augmented method generation is a new and hopeful technique to improve Large Language Models (LLMs). In this method, external knowledge from agriculture can be added to the input through attention blending or output interpolation[15]. This way, LLMs can give accountability for information and create more precise and suitable for farming purposes [24].

Different approaches have been applied to retrieval-augmented generation in the field of agriculture, such as:

a. *BM25 Algorithm:* Used to select the most similar farming details to a query by considering term frequency and document length [13].
b. *Sentence-BERT (SBERT):* A sentence-level embedding model used for effective example retrieval using contrastive learning [11].
c. *Dense Retrieval Models:* Utilizing feedback-driven dense retrievers for farming-related tasks, significantly impacting practical learning in context [11].

# 3 Agent-Based Data Sharing Framework

## 3.1 A Conceptual Overview of the ADS Framework

The ADS framework models the system with autonomous agents that represent the diverse players in agriculture, such as a farmer, researcher, or policymaker, using an agent-based architecture. There's a clear division of task, including goals, knowledge, and faculties, allowing them to cooperate in a distributed manner [2]. Within this framework, the decentralized data pool acts as a distributed database where agents can store and access shared data without a central point of control. Such a structure guarantees the control of the data with the individual agents, yielding independence and privacy in data use. Each agent makes an autonomous decision on what information to disclose to protect their own databases [Fig. 1].

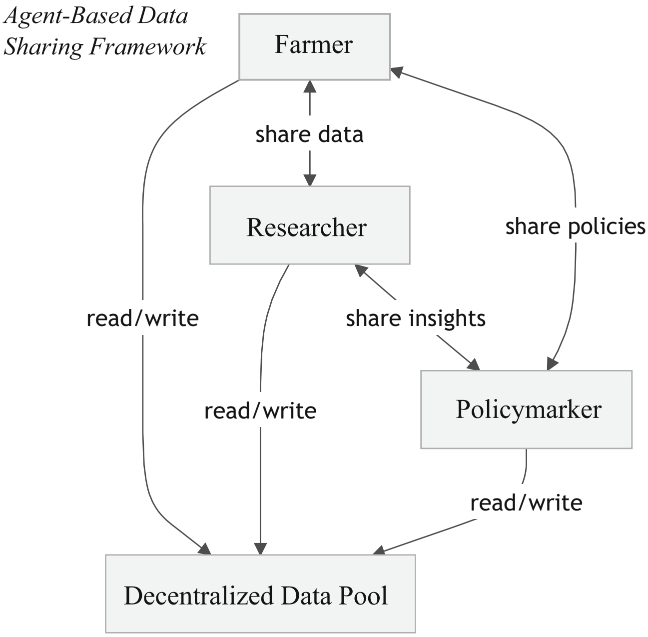


**Fig. 1.** Interconnected Agricultural System with three agents - Farmer, Researcher, and Policymaker - sharing data and insights, all linked to a Decentralized Data Pool for data storage and access.

### 3.2 Data Retrieval and Storage

Each agent in the framework is responsible for creating and maintaining its own dataset containing information such as crop yields, soil content, weather, or market price. This innovative method of producing and storing agriculture information captures and adds more localized agricultural data to the existing sets of data. When an agent tackles a wide-ranging agricultural problem, it initiates a complex process of product retrieval to identify the most pertinent instance, knowledge, or concept from a vast pool of resources (Fig. 2). This includes techniques like semantic space matching and context ranking, which aim to fit the query to the extent of available information. Such algorithms accomplish the task by reasoning about the asker's goals and objectives, the questions' reasons and context, and the agent's area of knowledge, thereby selecting and rendering the most attractive facts from the knowledge base. This, in turn, improves the agent's performance in context retrieval, aiding in additional decision-making aimed at improving agricultural practices.

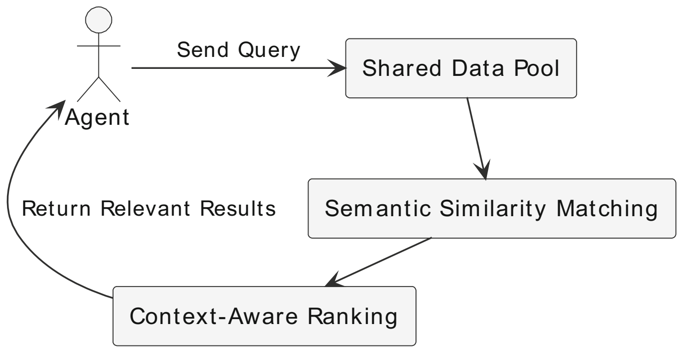


**Fig. 2.** Agent interacting with a Shared Data Pool, The Shared Data Pool processes this question through a two-step retrieval process involving Semantic Similarity Matching and Ranking Algorithms with Context.

# 4 Experiment

## 4.1 Dataset Description

For testing the proposed ADS framework, we utilized DataDreamer, an open-source Python library [3], to generate synthetic datasets simulating real-world agricultural scenarios using the Phi-3 foundational model [17]. The details of these datasets are summarized in Table 1.

## 4.2 Experiment Setup

In our experiment, we approached two main areas of application: disease detection and agronomic precision recommendations. For the disease detection area, we created agents well-skilled in diagnosing plant diseases through the incorporation of LLM-symptom analysis, climate, and plant metadata. In the area of providing precision farming recommendations, agents rendered advisory services by analysing data sourced from soil sensors, weather stations, and management activities in particular regions.

We created an instruction set to operationalize the agents, tailored to each agent's domain of knowledge. They used the OpenAI Compatible API [14] with the Llama 3 8B model. Furthermore, the agents used a code interpreter tool in real-time data analysis and computation, allowing them to analyse a large amount of data.

In order to check the data distribution effect on the agents' work, we implemented three different variants of data sharing:

a. Baseline (No Sharing): The agents worked separately with no data exchange.
b. Limited Sharing (Partial): The agents engaged in modest data sharing with one another in such matters as soil or weather data.
c. Extensive Sharing (All): The agents exchanged all available data and resources internally, without any restrictions on external sources.

**Table 1.** Dataset Description

| Component | Sample Number | Features | Target Variable | Data Split |
|---|---|---|---|---|
| Crop Yield Prediction | 10,000 | Historical yield data, weather conditions (temperature, precipitation, humidity), soil characteristics, management practices (irrigation, fertilization, pest control) | Crop yield (tons per hectare) | Training: 70%, Validation: 15%, Testing: 15% |
| Plant Disease Diagnosis | 5,000 | Visual symptoms (leaf discoloration, lesions, wilting), environmental conditions (temperature, humidity), plant metadata (species, growth stage) | Disease class (10 distinct classes) | Training: 60%, Validation: 20%, Testing: 20% |

In each of the data sharing scenarios, we posed a combination of 50 open-ended agronomy-related questions to the agents, simply focusing on disease diagnosis and precision agriculture, among other aspects. It allowed us to evaluate the impact of changes to levels of data sharing on the agents' response quality, completeness, and context relevance [Fig. 3].

### 4.3 Evaluation Methodology

To evaluate the effectiveness of data sharing among the agents, we developed a scoring formula that assesses their responses based on five key criteria: relevance ($s_r$), accuracy ($s_a$), completeness ($s_c$), clarity ($s_l$), and originality ($s_o$). Each criterion was assigned a score ranging from 0 to 1. The overall response score was calculated using the formula.

$$\text{Score} = \frac{w_r s_r + w_a s_a + w_c s_c + w_l s_l + w_o s_o}{\sum_{i \in \{r,a,c,lo\}} w_i} \tag{1}$$

where $w_i$ are the weights assigned to each criterion.

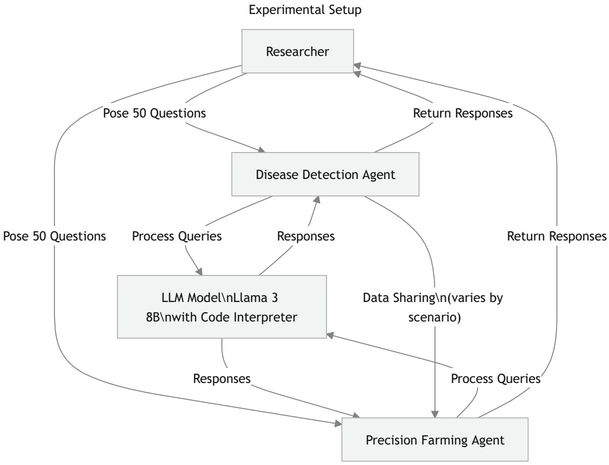


**Fig. 3.** The diagram experimental setup involves two specialized agents: one for disease detection and another for precision farming recommendations. The agents interact with the Llama 3 8B model via the OpenAI Compatible API, utilizing a code interpreter tool for real-time data analysis.

For simplicity and to assign equal importance to all evaluation criteria, we set all weights to 1.

$$(w_r = w_a = w_c = w_l = w_o = 1) \tag{2}$$

The denominator ensures that the resultant score is not greater than 1. We define the criteria as follows: relevance measures the response's ability to address the question; accuracy evaluates the truthfulness of the provided information; completeness gauges the extent of the answer to the question; clarity scrutinizes the answer and its presentation; and originality gauges the creativity and perception inherent in the response. We used this linear sum formula to add up the scores for each criterion and get the overall score.

## 4.4 Results

When comparing all three scenarios, the evaluation indicated very positive changes in agent performance as a function of the level of data sharing. The average performance measures of each domain with respect to the data sharing scenarios have been given in Table 2.

**Table 2.** Mean Performance Scores Across Data Sharing Scenarios.

| Domain | Baseline | Limited Sharing | Extensive Sharing |
|---|---|---|---|
| Disease Diagnosis | 0.7474 | 0.8190 | 0.9137 |
| Precision Farming | 0.5933 | 0.7208 | 0.8447 |

To illustrate the impact of data sharing on the quality of agent responses, we present a comparative analysis based on the agents' answers to the following question:

**"What is the optimal soil pH for growing tomatoes?"**

The agents' responses and corresponding scores under each data sharing scenario are summarized in Table 4, while the details shown in Table 3.

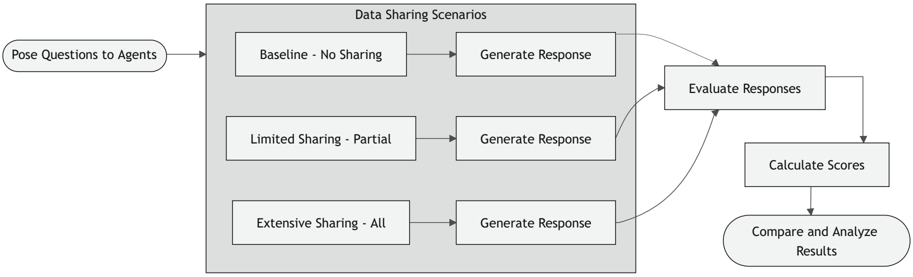


**Fig. 4.** Evaluation Process of Agent Responses

In the baseline scenario, agents operate independently without exchanging data. Partial sharing allows limited data exchange, such as weather or soil information, while keeping proprietary practices private. Full sharing involves complete pooling of all data and resources among agents. These scenarios represent a spectrum from isolated operation to full collaboration, enabling assessment of data sharing's impact on agricultural decision-making and overall system performance.

The results of the performance evaluation showed significant improvements in agent performance when extensive data sharing was employed [Fig. 4]. As a result, the mean performance metrics for each domain under different sharing scenarios were as follows.

**Table 3.** Scoring of Shared Data (Evaluated by Agent)

| Sharing Scenario | Relevance | Accuracy | Completeness | Clarity | Originality |
|---|---|---|---|---|---|
| Baseline (No Sharing) | 0.65 | 0.70 | 0.60 | 0.68 | 0.62 |
| Limited Sharing | 0.78 | 0.80 | 0.75 | 0.77 | 0.73 |
| Extensive Sharing | 0.90 | 0.92 | 0.88 | 0.91 | 0.89 |

In the disease diagnosis domain, mean performance scores increased approximately by 22% from the mean of 0.7474 to that of 0.9137. In precision farming, scores increased approximately 42% to 0.8447 from 0.5933. These huge gains show that the agents have a better capacity to deliver accurate, complete, and contextualized responses when presented with a larger amount of data.

**Table 4.** Agents' Responses and Scores Across Data Sharing Scenarios

| Sharing Scenario | Response | Explanation | Score |
|---|---|---|---|
| Baseline (No Sharing) | The optimal soil pH for growing tomatoes is between 6.0 and 6.8 | The LLM-generated response is mostly accurate but lacks some specificity | 0.7 |
| Limited Sharing (Partial) | The optimal soil pH range for growing tomatoes is between 6.0 and 6.8. Tomatoes prefer slightly acidic soil for optimal growth and nutrient uptake | The LLM-generated response is accurate and provides additional relevant information | 0.85 |
| Extensive Sharing (All) | The optimal soil pH range for growing tomatoes is between 6.0 and 6.8. Tomatoes prefer slightly acidic soil for optimal growth and nutrient uptake. Soil pH affects the availability of essential nutrients, and maintaining the proper pH range ensures the best possible growth, yield, and fruit quality | The LLM-generated response is highly accurate complete, and provides detailed relevant information | 0.95 |

## 5 Conclusion

The experimental results show that there is a greater improvement in response quality when agents share more information in both the disease diagnosis and precision farming areas. These details hold greater significance in agricultural decision-making, as even the smallest details significantly influence the process and outcome of the decisions.

**Acknowledgments.** This project has received funding from the European Union's Horizon 2021 research and innovation program under the Marie Skłodowska-Curie grant agreement No 101073381.

**Disclosure of Interests.** The authors declare that they have no competing interests.

## References


1. Luo, M., Xu, X., Liu, Y., Pasupat, P., Kazemi, M.: In-context Learning with Retrieved Demonstrations for Language Models: A Survey, https://arxiv.org/abs/2401.11624 (2024). https://doi.org/10.48550/ARXIV.2401.11624
2. Ault, A., Palacios, S., Evans, J.: Agriculture data sharing: conceptual tools in the technical toolbox and implementation in the open Ag data alliance framework. Agron. J. **114**(5), 2681–2692 (2022). https://doi.org/10.1002/agj2.21007
3. Patel, A., Raffel, C., Callison-Burch, C.: DataDreamer: A Tool for Synthetic Data Generation and Reproducible LLM Workflows, https://arxiv.org/abs/2402.10379 (2024). https://doi.org/10.48550/ARXIV.2402.10379
4. Condran, S., Bewong, M., Islam, Z., Maphosa, L., Zheng, L.: Machine learning in precision agriculture: a survey on trends, applications, and evaluations over two decades. IEEE Access **10**, 73786–73803 (2022). https://doi.org/10.1109/ACCESS.2022.3188649
5. Cooper, N.: Harnessing large language models for coding, teaching, and inclusion to empower research in ecology and evolution. Methods Ecol. Evol. (2024). https://doi.org/10.1111/2041-210X.14325
6. Dwivedi, S.: Breaking the bias: gender fairness in LLMs using prompt engineering and in-context learning. Rupkatha J. Interdisc. Stud. Humanit. **15**(4) (2023). https://doi.org/10.21659/rupkatha.v15n4.10
7. Zheng, H., et al.: Learn From Model Beyond Fine-Tuning: A Survey, https://arxiv.org/abs/2310.08184 (2023). https://doi.org/10.48550/ARXIV.2310.08184
8. Griend, R.: Artificial intelligence and scholarly publication in foot & ankle international and foot & ankle orthopaedics. Foot Ankle Int. **45**(3), 207 (2024). https://doi.org/10.1177/10711007241232288
9. Haileslassie, A., Mekuria, W., Schmitter, P., Uhlenbrook, S., Ludi, E.: Review of lessons learned in changing agricultural landscapes in Ethiopia: what worked well and what did not work so well? (2020). https://doi.org/10.20944/preprints202010.0124.v1
10. Hou, J.: Assessing large language models in mechanical engineering education: a study on mechanics-focused conceptual understanding (2024). https://doi.org/10.31219/osf.io/d3nc6
11. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. https://arxiv.org/abs/1908.10084 (2019). https://doi.org/10.48550/ARXIV.1908.10084
12. Li, W.: Segment, anything model, can not segment anything: assessing AI foundation model's generalizability in permafrost mapping. Remote Sens. **16**(5), 797 (2024). https://doi.org/10.3390/rs16050797
13. Sawarkar, K., Mangal, A., Solanki, S.R.: Blended RAG: Improving RAG (Retriever-Augmented Generation) Accuracy with Semantic Search and Hybrid Query-Based Retrievers, https://arxiv.org/abs/2404.07220. (2024). https://doi.org/10.48550/ARXIV.2404.07220
14. Allen, M., Pearn, K., Monks, T.: Developing an OpenAI Gym-compatible framework and simulation environment for testing Deep Reinforcement Learning agents solving the Ambulance Location Problem, https://arxiv.org/abs/2101.04434. (2021). https://doi.org/10.48550/ARXIV.2101.04434
15. Radeva, I.: Web application for retrieval-augmented generation: implementation and testing. Electronics **13**(7), 1361 (2024). https://doi.org/10.3390/electronics13071361
16. Raiaan, M.: A review on large language models: architectures, applications, taxonomies, open issues, and challenges (2023). https://doi.org/10.36227/techrxiv.24171183.v1
17. Abdin, M., et al.: Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone, arXiv:2404.14219. (2024)

18. Shutske, J.: Editorial: harnessing the power of large language models in agricultural safety & health. J. Agric. Saf. Health **29**(4), 205–224 (2023). https://doi.org/10.13031/jash.15841
19. Sivarajkumar, S.: An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: algorithm development and validation study. JMIR Med. Inform. **12**, e55318 (2024). https://doi.org/10.2196/55318
20. Stoyanov, S.: Using LLMs in cyber-physical systems for agriculture - ZEMELA (2023). https://doi.org/10.1109/BDKCSE59280.2023.10339738
21. Ubah, C.: Evaluation of AI models to update cybersecurity curriculum. J. Colloquium Inf. Syst. Secur. Educ. **11**(1), 8 (2024). https://doi.org/10.53735/cisse.v11i1.183
22. Woo, B.: Transforming nursing with large language models: from concept to practice. Eur. J. Cardiovasc. Nurs. (2024). https://doi.org/10.1093/eurjcn/zvad120
23. Qamar, T., Bawany, N.Z.: Understanding the black-box: towards interpretable and reliable deep learning models. https://doi.org/10.7717/peerj-cs.1629. (2023)
24. Lewis, P., et al.: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, https://arxiv.org/abs/2005.11401. (2020). https://doi.org/10.48550/ARXIV.2005.11401