Multi-Modal LLMs in Agriculture: A Comprehensive Review

Ranjan Sapkota, *Member, IEEE*, Rizwan Qureshi, *Senior Member, IEEE*, Muhammad Usman Hadi, *Member, IEEE*, Syed Zohaib Hassan, Ferhat Sadak, Maged Shoman, Muhammad Sajjad, *Member, IEEE*, Fayaz Ali Dharejo, *Senior Member, IEEE*, Achyut Paudel, Jiajia Li, *Graduate Student Member, IEEE*, Zhichao Meng, John Shutske, and Manoj Karkee

***Abstract*—Given the rapid emergence and applications of Multi-Modal Large Language Models (MM-LLMs) across various scientific fields, insights regarding their applicability in agriculture are still only partially explored. This paper conducts an in-depth review of MM-LLMs in agriculture, focusing on understanding how MM-LLMs can be developed and implemented to optimize agricultural processes, increase efficiency, and reduce costs. Recent studies have explored the capabilities of MM-LLMs in agricultural information processing and decision-making. Despite these advancements, significant gaps persist, particularly in addressing domain-specific challenges such as variable data quality and availability, integration with existing agricultural systems, and the creation of robust training datasets that accurately represent complex agricultural environments. Moreover, a comprehensive understanding of the capabilities, challenges, and limitations of MM-LLMs in agricultural information processing and application is still missing. Exploring these areas is crucial to providing the community with a broader perspective and a clearer understanding of MM-LLMs' applications, establishing a benchmark for the current state and emerging trends in this field. To bridge this gap, this survey reviews the progress of MM-LLMs and their utilization in agriculture, with an additional focus on 11 key research questions (RQs), where 4 RQs are general and 7 RQs are agriculture focused. By addressing these RQs, this review outlines the current opportunities and challenges, limitations, and future roadmap for MM-LLMs in agriculture. The findings indicate that multi-modal MM-LLMs not only simplify complex agricultural challenges but also significantly enhance decision-making and improve the efficiency of agricultural image processing. These advancements position MM-LLMs as an essential tool for the future of farming. For continued research and understanding, an organized and regularly updated list of papers on MM-LLMs is available at https://github.com/JiajiaLi04/Multi-Modal-LLMs-in-Agriculture**

***Note to Practitioners*—Motivated by the need to optimize agricultural practices, this paper investigates the use of Large Language Models (MM-LLMs) to improve efficiency and decision-making in agriculture. We delve into critical RQs to reveal the capabilities and challenges of MM-LLMs, and their potential applications in the agricultural sector. Looking ahead, our findings suggest a promising future for the integration of MM-LLMs in agriculture, potentially revolutionizing how we manage and operate farms.**

***Index Terms*—Large language models (LLMs), generative artificial intelligence, Agriculture, ChatGPT, deep learning, machine learning, computer vision, precision agriculture, vision-language models.**

Received 10 September 2024; revised 23 February 2025 and 26 June 2025; accepted 7 September 2025. Date of publication 19 September 2025; date of current version 28 October 2025. This article was recommended for publication by Associate Editor W. Beksi and Editor K. Liu upon evaluation of the reviewers' comments. This work was supported in part by the National Science Foundation (NSF); in part by United States Department of Agriculture (USDA); in part by the National Institute of Food and Agriculture (NIFA), through the "Artificial Intelligence (AI) Institute for Agriculture" Program, Accession Number 1029004 for the Project Titled "Robotic Blossom Thinning with Soft Manipulators" under Award AWD003473, Award AWD004595, and Award USDA-NIFA; and in part by United States Department of Agriculture National Science Foundation (USDANSF), Accession Number 1031712, under the Project "ExPanding University of Central Florida (UCF) AI Research To Novel Agricultural EngineeRing Applications (PARTNER)" under Grant 2024-67022-41788. *(Corresponding authors: Ranjan Sapkota; Manoj Karkee.)*

Ranjan Sapkota and Manoj Karkee are with the Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY 14850 USA (e-mail: rs2672@cornell.edu; mk2684@cornell.edu).

Rizwan Qureshi is with the Center for Research in Computer Vision, University of Central Florida, Orlando, FL 32816 USA.

Muhammad Usman Hadi is with the School of Engineering, Ulster University, BT15 1AP Belfast, U.K.

Syed Zohaib Hassan is with the Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL 32816 USA.

Ferhat Sadak is with the Department of Mechanical Engineering, Bartin University, 74100 Bartin, Türkiye.

Maged Shoman is with Oak Ridge Innovation Institute, University of Tennessee, Knoxville, TN 37996 USA.

Muhammad Sajjad is with the Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong, China.

Fayaz Ali Dharejo is with the Computer Vision Laboratory, CAIDAS, IFI, University of Würzburg, 97070 Würzburg, Germany.

Achyut Paudel is with Orchard Robotics, Seattle, WA 98164 USA.

Jiajia Li is with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 USA.

Zhichao Meng is with the College of Optical and Mechanical Engineering, Zhejiang A&F University, Hangzhou 311300, China.

John Shutske is with the Biological Systems Engineering Department, University of Wisconsin–Madison, Madison, WI 53706 USA.

Digital Object Identifier 10.1109/TASE.2025.3612154

I. Introduction

LANGUAGE processing refers to computational techniques that analyze, understand, and generate human language, employing a variety of algorithms and methods within the broad domain of machine learning (ML) [1], [2]. Machine learning itself is an extensive field, involving the development of algorithms that allow computers to learn from and make decisions based on data or examples [3]. Traditional ML techniques include a variety of statistical methods such as decision trees, which work utilizing partitioning of the data into subsets based on features, leading to conclusions about target values [4]. Likewise, Markov Decision Processes (MDPs) provide a framework for decision-making where outcomes are partly random and partly under the control of a decision-maker [5]. Support Vector Machines (SVMs) are another critical ML technique, used for classification and regression by finding the best boundary that separates data into classes [6]. Similarly, Bayesian analysis involves statistical inference that updates the probability of a hypothesis as more evidence becomes available [7]. These methodologies form the foundation of traditional ML, focusing on extracting patterns from data and applying them to predictive models. ML evolved quickly over the last decade enabled by the advancement of deep learning (DL) neural networks. The advancement also led to the rise of specialized subsets of ML techniques such as Large Language Models (MM-LLMs), which are specifically designed to process vast amounts of textual data, enabling enhanced interaction and comprehension of human language across various applications.

The language processing with ML encompasses tasks such as text generation [8], sentiment analysis [9], and machine translation. The automated conversion of text from one language to another [10], enables machines to interact with and comprehend human language effectively [11], [12]. Traditional Language Models (LMs) have been foundational in language processing as they provided a basis for text generation and understanding [13].

With the advent of increased computational power, advanced ML techniques, and access to large-scale data, the emergence of MM-LLMs has marked a significant transition. These models, equipped with expansive and diverse training data, have demonstrated an impressive ability to simulate human linguistic capabilities, leading to transformative changes across multiple domains, such as national security [14], healthcare [15], robotics [16], autonomous vehicle technologies [17], road and traffic safety [18], security and privacy [19] and so on [20].

In agriculture, the application of ML, including DL, techniques has revolutionized various aspects such as crop monitoring, pest management, and yield prediction, establishing these fields as some of the most significant beneficiaries of artificial intelligence (AI) technologies [21], [22]. Among these technologies, MM-LLMs such as the Generative Pre-trained Transformer (GPT) series have emerged as groundbreaking tools [23], [24]. These models leverage transformers to process and analyze vast amounts of agricultural data, offering insights previously unattainable through traditional techniques such as manual sampling, visual inspection, and basic statistical analysis [23]. MM-LLMs can respond to free-text queries without training to specific tasks, demonstrating remarkable adaptability and versatility. For instance, ChatGPT, a prominent example of a generative AI chatbot, showcases the potential of MM-LLMs through its sophisticated fine-tuning processes such as adjusting the model's parameters post pretraining on a task-specific dataset, to improve its understanding and generation of language in specific contexts [24], allowing its adoption to agricultural applications among other industries.

MM-LLMs have advanced significantly beyond traditional data processing applications such as statistical regression, spreadsheet-based data manipulation, and rule-based expert systems [25], bringing deep contextual understanding and predictive analytics into the agricultural sector. These models, including the widely recognized GPT series from OpenAI, analyze large datasets to predict outcomes, understand complex patterns, and provide actionable insights [26]. In agriculture, this capability translates to improved decision-making in areas such as crop rotation, soil management, and predictive maintenance of farming equipment [27]. The ability of MM-LLMs to interpret and generate human-like text further enables them to assist in real-time problem-solving and customer service, enhancing the responsiveness of businesses within the agricultural supply chain. The profound impact of MM-LLMs lies in their ability to transform vast and diverse data into coherent, insightful, and immediately actionable information, thus driving efficiency and productivity in agricultural practices [28], [29].

Despite their potential, MM-LLMs in agriculture face several shortcomings, including data privacy concerns, high computational costs, and the need for extensive domain-specific data for accurate predictions. Additionally, the models can sometimes generate inaccurate or biased outputs due to insufficient or unrepresentative training data, which may lead to undesirable farming outcomes. Moreover, integrating MM-LLMs into existing agricultural systems can be complicated and requires significant adaptation. Further exploring MM-LLMs is essential to develop more robust, efficient, and context-aware systems for enhancing the "farms of the future", and reducing resource waste while increasing yield and quality, thus increasing overall environmental, economic, and social sustainability in agricultural practices [30], [31].

Additionally, despite the technical advancements in MM-LLMs, their application in agriculture should be guided by the practical challenges in farming operations and the needs of farmers and other agricultural practitioners. Farmers, as primary end users, require solutions that are not only technologically advanced but also accessible and relevant to their daily challenges [32]. For instance, in PA (Precision Agriculture), MM-LLMs can be tailored to provide predictive analytics for optimal planting times and fertilizer application rates, directly influencing yield and resource management [33], [34]. Furthermore, MM-LLMs can be utilized to offer real-time language translation services, helping farmers in multilingual regions communicate more effectively with suppliers and clients, thereby enhancing operational efficiency [34]. It is also noted that integrating farmers' long experience, expertise and feedback during the development phase of MM-LLMs could lead to innovations that can address real problems faced by farmers rather than bringing technological solutions to problems that do not exist in farms. Some example solutions where this collaborative approach has been or will be beneficial include voice-activated systems for hands-free operation in the field or customized alert systems for pest outbreaks and weather changes, which are critical for timely interventions [34], [35]. LLM-driven crop input recommendation and market analysis tools, informed by farmer feedback, can predict prices and demand fluctuations, helping farmers make better production and sales decisions. Involving farmers in development ensures models address real challenges and accelerates technology adoption.

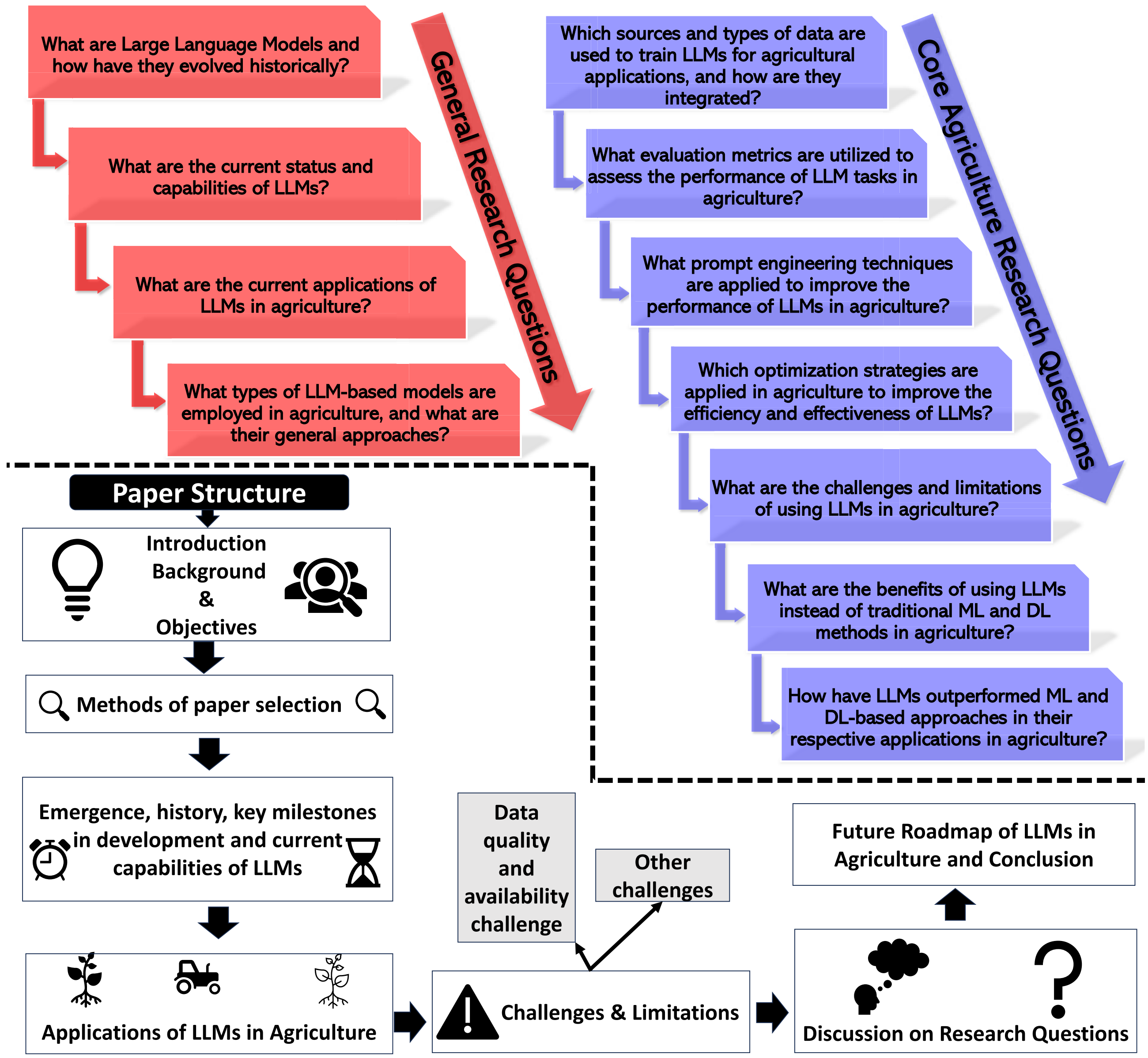

Fig. 1. Research questions (RQs) flowchart highlighting the core agricultural and general RQs addressed in the study, illustrating the comprehensive inquiry into the applications of MM-LLMs in agriculture.

To have a broader understanding and to fill the gaps identified, this study surveys the state-of-the-art use of MM-LLMs in agriculture. The primary focus is on answering the RQs (Researech Questions), as illustrated in Figure 1. By addressing these questions, this review bridges knowledge gaps by systematically exploring LLM applications, benefits, challenges, and future directions in agriculture.

A. Method of Paper Selection

A systematic bibliographic analysis was conducted to comprehensively review LLM integration in agriculture. The research involved two main steps: collecting relevant works and performing detailed review and analysis. Primary searches used Google Scholar, supplemented by secondary searches in ScienceDirect (215 papers), IEEE Xplore (103 papers), PubMed (126 papers), Web of Science (30 papers), and Scopus (37 papers), as shown in Figure 2. The search utilized the query: ['Large Language Models' OR 'LLM'] in/and ['agriculture'], aiming to exclude papers that referenced MM-LLMs without directly applying them to the agricultural domain. This initial search yielded a total of 460 papers across the five databases. Papers were then screened for relevance, with repeated and irrelevant papers being minimized, resulting in 207 References that were used to build this comprehensive survey.

The methodological framework began by defining the research objective: to determine if gaps exist in understanding MM-LLMs in agriculture. A stepwise approach was used, starting with keyword refinement. The broad query "Large Language Models in Agriculture" yielded numerous results (e.g., 2,315 in ScienceDirect, 1,910 in PubMed). Narrowing

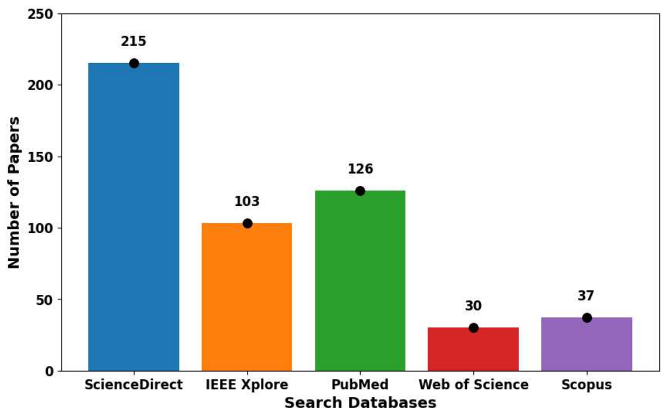

Fig. 2. Search database results for the number of papers.

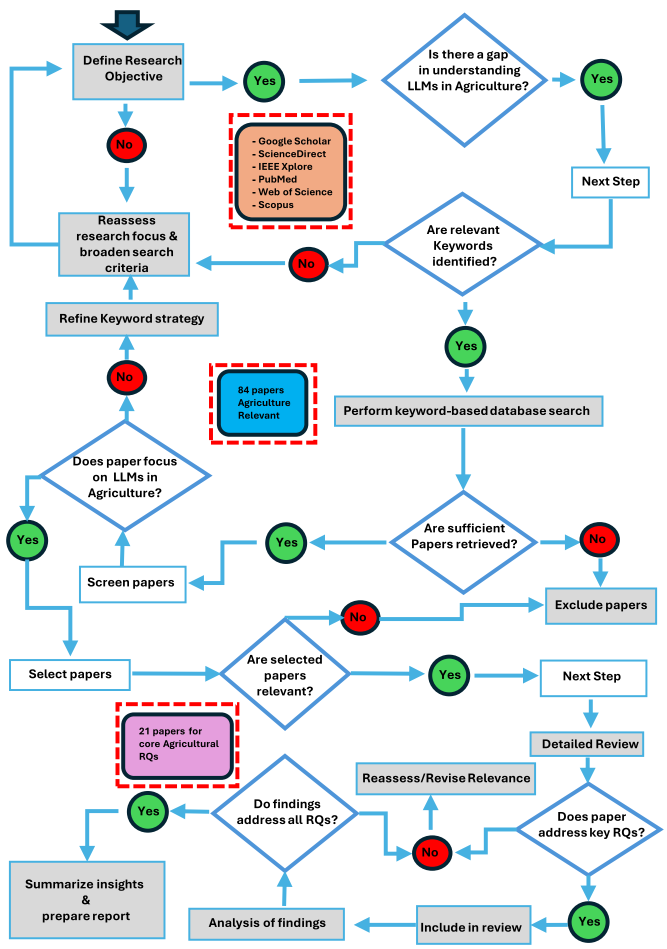

Fig. 3. Flowchart illustrating the systematic process for bibliographic analysis and RQs evaluation in this review study.

to "MM-LLMs in Agriculture" reduced relevant papers to 204 in ScienceDirect and 114 in PubMed. From 207 selected papers, 105 were strongly related to agricultural applications, with 84 directly relevant to agricultural use. As shown in Figures 1 and 3, 21 papers aligned closely with seven core agricultural RQs and were included in the final analysis. Four of these RQs aimed to establish a general understanding of MM-LLMs defining core concepts and evaluating state-of-the-art developments. The review then systematically addressed these RQs, presenting structured findings.

1) Limitation of the Study: This study primarily relied on peer-reviewed academic papers due to their rigorous review process, ensuring reliable findings. However, future research could expand to include non-peer-reviewed sources like technical reports, farmer testimonials, and anecdotal evidence. Incorporating these would capture practical implementations and farmer experiences through qualitative methods or case studies, validating academic insights and enhancing the study's practical relevance. This broader scope would better align with agricultural end-users' needs and ensure the technology's applicability in real-world farming environments.

B. Paper Structure and Content Overview

The paper begins by examining the intricacies of **MM-LLMs**, divided into two subsections: *Emergence to State-of-the-art Capabilities of MM-LLMs* and *Vision Language Models (VLMs)*. Building on this foundation, the section **Application of MM-LLMs in Agriculture** details specific uses such as *MM-LLMs for crop monitoring and management*, *disease detection and diagnosis*, *precision irrigation and fertilization*, *agricultural information and decision support tools*, and *agricultural extension and technical advisors*. The paper then explores the **Commercialization of MM-LLMs by Private Companies in Agriculture**, highlighting private sector roles in technology deployment. It transitions to a **Discussion on Research Questions** that frames systemic challenges in the field. The **Major Challenges and Limitations** section addresses three critical areas: *Data Quality Challenge*, *Data Availability Challenge*, and *Other Challenges*, outlining key obstacles to LLM integration. The review concludes with **Conclusion and Future Roadmap of MM-LLMs in Agriculture**, discussing anticipated developments and future directions in the agricultural domain.

II. Large Language Models

Language is fundamental to human communication and self-expression, which serves as a crucial method for human-machine interfacing as well [23]. The increasing need for machines to perform complex language tasks has driven the development of more advanced language models that can understand and generate language [2]. Recent advances in language models were primarily enabled by **transformers**, advanced architectures that manage context, dependencies, and relevance between words irrespective of their location or distance in the text. Enhanced computational power and access to large-scale training data (e.g., content on the internet such as Reddit discussions) have also played significant roles [36]. These advancements enable **MM-LLMs** to approximate human-level performance on various tasks such as "Expert Suggestions" or standardized "Test Taking," establishing them as state-of-the-art in AI capable of processing text and generating coherent communication. The evolution of MM-LLMs has transitioned from statistical methods to neural language modeling, advancing from pre-trained language models (PLMs) to the latest, highly sophisticated MM-LLMs such as **Reinforcement Learning** [37].

Traditional language modeling techniques have laid the foundation for modern approaches to natural language processing. These include **n-grams**, which predict the next item in a sequence as a function of the preceding one or several items [38]; **Hidden Markov Models (HMMs)**, which are statistical models that output a sequence of symbols or quantities based on hidden states [39]; and **Maximum Entropy Models (MEMs)**, which predict outcomes based on the principle of making as few assumptions as possible, except for those justified by known data [40]. Similarly, **Conditional Random Fields (CRFs)** have been used in the past for language modeling. CRFs are a class of statistical modeling methods often used in pattern recognition and machine learning where they specify known relationships between observed data and outcomes to be predicted [41]. **Latent Dirichlet Allocation (LDA)** is another generative statistical model that allows sets of observations to be explained by unobserved groups, helping to understand why parts of the data are similar [42]. While effective for shorter texts, these traditional models often struggle with understanding the overall context or meaning of longer texts, as they generally consider only a limited number of words or phrases at a time.

It is also noted that conventional language modeling trains task-specific models in supervised settings, whereas **PLMs (Pre-trained Language Models)** are trained in a self-supervised manner on large corpora of text to learn a generic representation that can be shared across various natural language processing (NLP) tasks [43]. With fine-tuning for downstream tasks, PLMs often achieve performance gains superior to those obtained through traditional language modeling [44]. As the PLMs grow larger, their performance improves, leading to the transition from PLMs to **MM-LLMs** through a significant increase in model parameters and training datasets [45].

The core of MM-LLMs is the **Transformer** architecture, which has revolutionized the field of NLP through its self-attention mechanisms and positional encodings, enabling the model to capture complex relationships within data [46]. Although these models build on foundational elements shared with earlier architectures, the introduction of self-attention and positional encodings marked a significant leap in handling large datasets and scaling computational power. This scaling has allowed MM-LLMs to process and generate language with remarkable accuracy, particularly in understanding long-range dependencies in text [47]. The progression of NLP models reflects these advances, with earlier models like **Recurrent Neural Networks (RNNs)** and **Convolutional Neural Networks (CNNs)** laying the groundwork, but struggling with capturing long-range dependencies essential for tasks such as translation and summarization [47], [48].

The breakthrough in addressing these limitations came in 2017 with Vaswani et al.'s introduction of the Transformer model [46]. Its innovative self-attention mechanism enables it to effectively process every token in the input sequence, capturing long-range dependencies that earlier models could not. This advancement significantly boosted performance across a variety of NLP tasks [49], marking a major evolution in the field. Figure 4 shows the history and evolution of MM-LLMs.

The historical evolution of MM-LLMs can be divided into two principal phases, each highlighting major advancements and technologies:

1) **Foundational Techniques and Statistical Learning:**
   - **Early Foundations (Pre-1990s):**The field began with rule-based systems, notably ELIZA (1966), one of the earliest examples of a natural language processing program [50], [51]. ELIZA used a simple pattern-matching technique to mimic conversation. It operated by recognizing keywords or phrases in the user's input and then applying hard-coded rules to form responses [52]. Despite its ability to simulate dialogue, ELIZA lacked any real understanding of language or context, which limited its conversational depth.
   - **Statistical Methods (1990s–2010s):** This era introduced probabilistic models like n-grams, which predict the likelihood of a sequence of words based on the frequencies of previous word sequences in a text [53]. Hidden Markov Models were also prominent, used primarily for tasks like speech recognition, where they modeled language as a series of states and transitions [53], [54]. Techniques such as Naive Bayes and Support Vector Machines were later used to improve the accuracy of tasks like spam detection [55], [56].
   - **Neural Networks and Word Embeddings (2010s):** Technologies like Word2Vec (2013) and GloVe (Global Vectors for Word Representation, 2014) substantially advanced word embeddings. These models are capable of converting words into vectors of real numbers, which capture semantic meanings and relationships between words. Word2Vec uses contextual insights from texts to produce word embeddings, whereas GloVe constructs an explicit word-context or word-cooccurrence matrix using statistics across the whole text corpus for its embeddings.

2) **The Transformer Era and Advances in Model Scale and Specialization:**
   - **The Transformer Revolution (2017–Present):** As discussed earlier, the introduction of the Transformer model in the paper "Attention Is All You Need" (2017) [46] marked a major shift in NLP. Transformers utilize self-attention mechanisms that assess the importance of all other words for each word in the data sequence, thus allowing the model to process all words simultaneously to generate the next token and capture complex inter-word relationships, regardless of their position in the input sequence. Based on this architecture, models like BERT (Bidirectional Encoder Representations from Transformers, 2018) was developed, which uses a bidirectional training of the transformer to generate a rich contextual understanding of language.
   - **Era of Scale and Specialization (2020s):** Current developments focus on rapidly scaling up the size

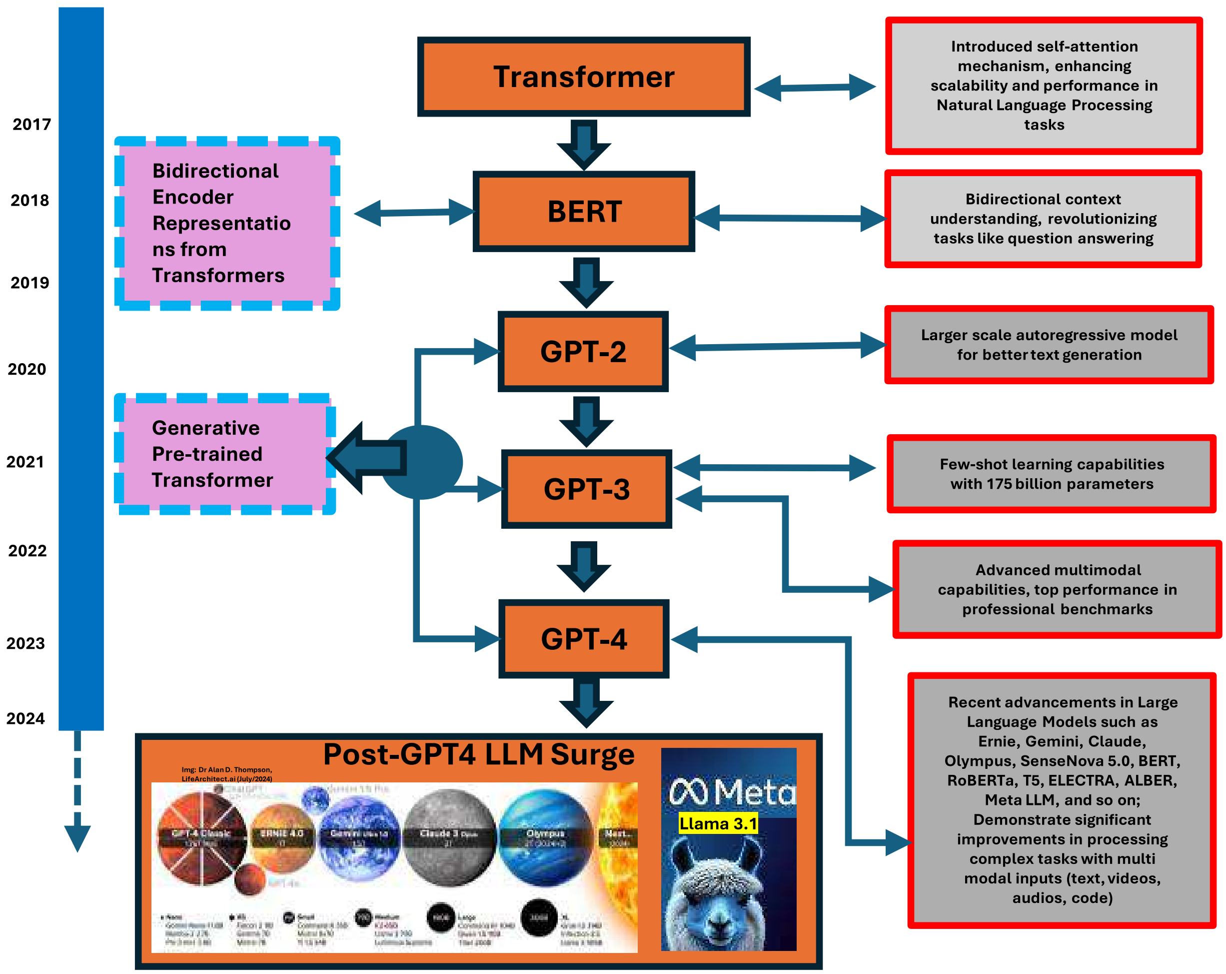

Fig. 4. Flow diagram showing the History of MM-LLMs, starting from transformer to current MM-LLMssurge. The post-GPT4 surge in MM-LLMshas risen to a level where hundreds of MM-LLMs have been released.

of these models, and the volume and diversity of the input data. For instance, models like GPT-3 and PaLM illustrate this trend by leveraging extensive parameters and diverse datasets to improve performance. Additionally, there's a growing emphasis on extending capabilities beyond text, as seen with multimodal models like DALLE and Flamingo, which integrate images, sounds, and other data types to broaden their application scope. These large models leverage their extensive set of trainable parameters to perform a wide array of NLP tasks with minimal task-specific fine tuning, demonstrating remarkable adaptability and proficiency across diverse domains.

As also discussed in the previous paragraph, MM-LLMs are trained using a vast amounts of data from various sources including text, images, audio, and more recently, multimodal inputs, which help MM-LLMs to learn and perform a wide range of tasks such as natural language understanding and generation, translation, summarization, and even complex reasoning [57], [58]. Major examples include generating human-like prose [59], composing music [60], [61], or creating art [62], [63], demonstrating their versatility and depth of learning. Likewise, MM-LLMs excel at question-answering systems, where they use their vast knowledge base to provide answers that are contextually relevant, often outperforming traditional search methods [64]. They also assist in language translation services, achieving near-human accuracy between a wide range of language pairs [65]. Another significant application is in sentiment analysis, where MM-LLMs evaluate and classify opinions from large datasets, providing valuable insights into consumer behavior, as well as in automated code generation, where they assist developers by offering context-aware code suggestions and debugging support [66]. Moreover, these models are increasingly being used in professional fields such as law [67] and medicine [68], where they help to search through large volumes of information to offer preliminary legal arguments or medical diagnostics. In essence, MM-LLMs are reshaping how data is processed, interpreted, and utilized across various sectors, offering enhanced decision-making tools and insights that were previously unattainable.

A. *Capabilities of Large Language Models and Its Relevance to Agriculture*

As discussed before, the recent advancement in MM-LLMs became possible with the introduction of the Transformer architecture [46] which set a new benchmark for handling sequential data through its attention mechanisms. This innovation was further enhanced in 2018 when Devlin et al. introduced BERT (Bidirectional Encoder Representations from Transformers), utilizing deep bidirectional training that allowed the model to assimilate information from both the left and right contexts of a text simultaneously [49]. This approach significantly advanced the model's understanding of context, enabling it to be fine-tuned with just one additional output layer for a broad range of NLP tasks. Continuing this trend, in 2019, Radford et al. introduced GPT-2, which expanded the transformer-based architecture to 1.5 billion parameters, significantly enhancing its text generation capabilities through refined self-attention mechanisms [12], [47]. The same year witnessed the development of Megatron-LM by Shoeybi et al., which scaled up to 8.3 billion parameters. This model not only improved complex pattern recognition but also training efficiency through an innovative parallelization scheme [69]. The advent of GPT-3 by Brown et al. in 2020 brought MM-LLMs to the next level of performance, supported by a model with 175 billion parameters capable of generating high-quality text with minimal fine-tuning, pushing the boundaries of what was achievable in natural language processing [23].

Most recently, GPT-4 has demonstrated the capability of MM-LLMs to process both text and image inputs, showcasing near-human performance across various benchmarks, introducing a shift towards truly multimodal models (OpenAI, 2023). Concurrently, Meta's release of the open-weights MM-LLMs LLaMa, though smaller in size, provides an accessible resource that continues to expand access to cutting-edge AI technologies (Meta, 2023). The evolution of these models highlights the critical role of tokenization in enhancing the flexibility and applicability of MM-LLMs, especially in complex, data-intensive fields like agriculture. Tokenization is the process by which raw data - whether text, images, or temporal sequences - are converted into a structured format of the smallest units of data that MM-LLMs can interpret [70], [71]. This process is crucial for managing the diversity and sparsity of data in broad range of applications including agriculture. This mechanism enables the effective handling of multi-modal time series data highly relevant to agricultural applications. For example, tokenization allows for the uniform representation of disparate data types [72], from satellite imagery to soil moisture levels, facilitating their integration and analysis within a single model. The padding and end-of-data features of tokenization also address issues with data sparsity and variable length data by ensuring seamless processing of variable length data [73]. With these features, tokenization enables MM-LLMs to learn from and generate meaningful predictions across different contexts including agricultural predictions and monitoring [20], [74]. Specifically, by converting complex datasets into tokens, models like Meta-Transformer [70] can efficiently process and learn from unpaired multimodal data, extracting high-level semantic features essential for accurate forecasting and decision-making in agriculture. This capability is transformative for agricultural applications, where an accurate mapping of multi-modal time-series input to desired output can significantly impact farming operations, from predicting crop yields to optimizing resource allocation.

Recent MM-LLMs like OpenAI's GPT-4, DeepSeek-R1, Grok, Claude, Copilot, and Google's PaLM which are at the forefront of NLP are showing remarkable advancements in handling complex tasks with vast parameter scales. However, their increasing integration into daily life raises significant ethical concerns, including potential biases in model training data and the opacity of their decision-making processes. Efforts to mitigate these biases and improve transparency are ongoing, but challenges remain, particularly in regulation and ensuring fairness. Additionally, the environmental and economic costs of training and deploying these resource-intensive models are significant. Despite these challenges, the future of MM-LLMs is likely to see deeper integration with other AI technologies, promising even more sophisticated autonomous systems. This ongoing evolution underscores the importance of continued ethical scrutiny and the development of regulations to prevent misuse and ensure that the benefits of MM-LLMs are realized responsibly and equitably.

In the following section, we review the leading MM-LLMs and their current capabilities, first highlighting the state-of-the-art MM-LLMs developed by leading AI companies: Amazon, OpenAI, Meta, Google, Microsoft, Apple, NVIDIA, and Alibaba in Table II. Subsequently, we provide a discussion on each, outlining their key features and contributions.

- **OpenAI:** The evolution of GPT models by OpenAI marks a new era in the development of language understanding and generation. This post-GPT era demonstrates rapid advancement of model design and capabilities. The capability of MM-LLMs was first demonstrated with GPT-1 in June 2018, which showcased the effectiveness of unsupervised pre-training on large datasets to improve performance on diverse language tasks. GPT-1, with 0.12 billion parameters, laid the groundwork for scalable, task-agnostic language models [75]. Following this, GPT-2 was introduced in February 2019, boasting 1.5 billion parameters trained on 40GB of Internet text. This model showed unprecedented capabilities in generating coherent and contextually relevant text without task-specific training [76]. The advancements continued with GPT-3, released in June 2020, featuring a massive 175 billion parameters. GPT-3 excelled in few-shot learning, achieving state-of-the-art performance on a wide range of NLP tasks, and setting a new benchmark for language models [23]. Following the success of GPT-3, OpenAI introduced GPT-4 in 2023 [77]. GPT-4 represents a significant leap forward as a multimodal model, capable of processing both text and visual inputs, thereby expanding its utility across a broader range of applications. GPT-4 demonstrated human-level performance on various professional and academic benchmarks, such as the bar exam, where it scored in the top 10% of test takers. The model's ability to handle complex and nuanced scenarios in multiple languages further solidified its position as a cutting-edge AI system. Building on this, OpenAI announced GPT-4o, a groundbreaking advancement that integrates text, audio, image, and video processing capabilities [78].

GPT-4o, which stands for "omni", achieves remarkable performance improvements in multilingual, audio, and visual tasks, making it a versatile tool for real-time, multimodal human-computer interaction.

- **DeepSeek:** DeepSeek has recently captured the attention of the global tech community with its innovative model, DeepSeek-R1 [80]. This model is designed to efficiently manage a diverse array of text-based tasks, excelling particularly in languages such as English and Chinese. This model achieves comparatively high performance in complex reasoning areas like coding, mathematical computations, and scientific explanations, setting a new benchmark in AI capabilities. DeepSeek-R1 model includes 671 billion parameters, but it operates with only 37 billion parameters activated during each forward pass, enhancing its efficiency substantially compared to other recent MM-LLMs.

  The model utilizes a sophisticated Mixture of Experts (MoE) architecture, which utilizes a set of specialized models, termed "experts," where each expert is trained to handle specific types of data or tasks, and a gating network that decides which expert to consult for a given input, optimizing model performance and computational efficiency by dynamically allocating resources based on the complexity of the task at hand [123]. This method not only improves processing speeds but also enhances the model's ability to address complex problem-solving scenarios [124], [125]. Coupled with reinforcement learning techniques, DeepSeek-R1's MoE framework significantly improves its reasoning capabilities, enabling it to adapt and learn from new challenges dynamically. DeepSeek-R1's key performance measures are equal to or better than those of OpenAI's flagship models, while this model was developed on modest computational infrastructure compared to several magnitudes of scale higher computational infrastructure used by thecounterparts. Now available under an MIT license on platforms like Hugging Face, it provides unrestricted commercial use, making this powerful AI tool accessible to a wider range of users from startups to academic labs. The operational cost of DeepSeek-R1 is estimated to be only 15%-50% of that of similar models from companies like OpenAI, which contributes substantially in broadening access to MM-LLMs.
- **Amazon's LLM**: Amazon has significantly contributed to the advancement of MM-LLMs through its innovative research and development [81], [82]. The AlexaTM 20B model, a multilingual sequence-to-sequence (seq2seq) model, excels in few-shot learning tasks, outperforming larger models like GPT-3 in zero-shot settings, and showcases superior performance in machine translation and summarization tasks [83]. This model demonstrates Amazon's capability to achieve high efficiency and effectiveness even with a relatively smaller parameter scale. The HLAT model, designed for efficient and scalable learning, highlights Amazon's commitment to optimizing model training and enhancing performance across diverse applications [81]. Additionally, the multimodal-CoT (Chain-of-Thoughts) model integrates language and vision modalities to improve reasoning accuracy and mitigate hallucinations through a two-stage framework that separates rationale generation from answer inference [82]. The CoT is a prompting technique that enables MM-LLMs to break down complex problems into intermediate reasoning steps, mimicking human-like logical thinking. This structured approach allows models to generate more interpretable and accurate outputs, particularly in tasks requiring multi-step problem-solving [82], [126].
- **Microsoft's LLM**: The Orca 2 model, with its innovative approach to teaching smaller LMs diverse reasoning techniques, achieves performance comparable to much larger models [117]. Florence-2 advances vision tasks with a unified, prompt-based representation, leveraging a vast annotated dataset for superior zero-shot and fine-tuning capabilities [118]. The Multimodal-CoT model integrates language and vision modalities to enhance reasoning accuracy, addressing the limitations of single-modality CoT models [82]. In Orca 2, improved training signals significantly boost smaller LMs' reasoning abilities, surpassing models 5-10x larger on complex benchmarks [117]. Florence-2's extensive dataset of 5.4 billion visual annotations ensures comprehensive performance in vision tasks, reflecting its robust multi-task learning setup [118]. Multimodal-CoT's two-stage framework mitigates hallucination and enhances convergence speed, demonstrating the efficacy of multimodal information fusion [82].
- **Meta AI**: Meta has developed several cutting-edge MM-LLMs that push the boundaries of AI capabilities. The Chameleon family of models, as described in [84], represents a significant advancement in mixed-modal models. Chameleon uses an early-fusion, token-based approach to integrate images and text, enabling it to perform a variety of tasks such as visual question answering, image captioning, text generation, and image generation [127]. The models achieve state-of-the-art performance in image captioning and outperform other text-only models in text generation tasks. Chameleon also demonstrates competitive performance against much larger models in long-form mixed-modal generation evaluations, showcasing its versatility and robustness. Meta Llama 3 and 3.1 [85], [128], is hailed as the most capable openly available MM-LLMs to date. This model excels in instruction fine-tuning through a combination of supervised fine-tuning, rejection sampling, proximal policy optimization, and direct preference optimization. The study by Meta AI presented in [86] introduces a series of long-context MM-LLMs capable of handling up to 32,768 tokens. These models are built from the continual pre-training of Llama 2, with a focus on longer training sequences and an upsampled dataset of long texts. The models demonstrate consistent improvements on regular tasks and significant enhancements on long-context tasks compared to Llama 2. A notable achievement is that the 70B variant surpasses GPT-3.5-turbo-16k's performance on long-context tasks using a cost-effective instruction tuning method.

[89] introduced BlenderBot 3x, an enhanced conversational model trained using organic conversation and feedback data from real users. This approach aims to improve both the skills and safety of the model. Likewise, [129] introduces LLaMA, a series of foundation models from 7B to 65B parameters, trained exclusively on publicly available datasets. LLaMA models, particularly the 13B and 65B variants, outperform GPT-3 and compete with models like Chinchilla-70B and PaLM-540B. Reference [87] discusses LIMA, a 65B parameter LLaMA model fine-tuned with only 1,000 curated prompts and responses. LIMA achieves remarkable performance in following response formats and generalizing to unseen tasks, often preferred over GPT-4, Bard, and DaVinci003 in controlled studies, highlighting the significance of pre-training in MM-LLMs.

Additionally, [90] introduces Atlas, a retrieval-augmented language model designed for few-shot learning in knowledge-intensive tasks. By leveraging a retrieval mechanism, Atlas achieves high accuracy with minimal training examples, demonstrating over 42% accuracy on Natural Questions with just 64 examples, outperforming much larger models like the 540B parameters model. This approach significantly reduces the need for massive parameter counts while maintaining robust performance across various benchmarks such as MMLU, KILT, and Natural Questions. Reference [91] presents the Open Pre-trained Transformers (OPT), which matches the performance of GPT-3 with significantly reduced computational resources, showcasing its efficiency in large-scale model operations.The emphasis on lowering environmental impact while maintaining high performance highlights the significance of sustainable AI practices. The full model weights and training logs are made available, promoting transparency and reproducibility in large-scale model training. Reference [92] introduces InCoder, a generative model capable of both code synthesis and infilling, trained on a diverse corpus of code with bidirectional context, InCoder excels in zero-shot code infilling tasks, including type inference, comment generation, and variable renaming.

- **Google Deepmind:** Google DeepMind has made significant strides in the development of MM-LLMs with their recent publications. The Gemini 1.5 family, introduced by Reid et al. [105], represents a new generation of compute-efficient multimodal models that excel in recalling and reasoning over extensive contexts, including text, audio, and video. These models demonstrate superior performance in long-document QA and long-video QA tasks, achieving near-perfect retrieval accuracy for contexts up to 10 million tokens. In the medical domain, Saab et al. [106] introduced Med-Gemini, which outperforms GPT-4 on several medical benchmarks, showcasing its potential in medical applications such as text summarization and multimodal dialogue. Another notable development is Griffin, a hybrid model combining gated linear recurrences with local attention, as described by de et al. [107]. Griffin exhibits efficient training and inference capabilities, matching the performance of Llama-2 while using significantly fewer tokens. Additionally, the Gemma family, detailed by the DeepMind team [108], offers lightweight open models with strong performance across language understanding, reasoning, and safety benchmarks. These models, ranging from 2 billion to 7 billion parameters, are released with a focus on responsible AI development, aiming to advance the field of MM-LLMs through open and safe innovations.

  PaLI-3, as detailed in [109], introduces a VLM with only 5B parameters, achieving superior performance across various multimodal benchmarks. It emphasizes the effectiveness of contrastively pre-trained Vision Transformers, particularly for tasks like localization and visually-situated text understanding. RT-X, described in [110], employs a large, diverse dataset collected from 22 different robots to train generalizable robotic policies. This model demonstrates positive transfer, significantly improving capabilities across multiple robots and tasks. Med-PaLM M, highlighted in [111], is a multimodal biomedical AI system that integrates clinical language, imaging, and genomics.
- **Apple:** Apple's recent advancements in MM-LLMs are marked by significant contributions across various domains, as demonstrated by five notable papers. The first paper [93] presents a robust vision language model (VLM) trained on diverse multimodal datasets, achieving state-of-the-art performance across a broad range of tasks. It emphasizes the potential of a unified model handling substantially more modalities without performance loss. [94] introduces OpenELM, an open and reproducible language model framework that enhances accuracy and efficiency through a layer-wise scaling strategy, offering comprehensive resources for the research community. Reference [95] explores the construction of multimodal MM-LLMs, highlighting the critical role of data diversity and architectural components in achieving superior few-shot learning results across different benchmarks. Reference [96] innovates in reference resolution, demonstrating significant improvements by converting the problem into a language modeling task, achieving results comparable to GPT-4. Lastly, [97] addresses the challenges in UI comprehension, proposing Ferret-UI, a model that excels in understanding and interacting with mobile user interfaces, outperforming existing models like GPT-4V on numerous tasks.

  Another MM-LLMsfrom Apple, MGIE [98], enhances instruction-based image editing by using MM-MM-LLMs to generate expressive and detailed commands. This approach addresses the challenge of brief human instructions, allowing for more precise image manipulation. The model integrates visual imagination with editing capabilities through end-to-end training, demonstrating significant improvements in automatic metrics and human evaluations across various editing tasks, including global photo optimization and local modifications, while maintaining efficient inference. Ferret, another innovative model from

Apple [99], excels in spatial referencing and grounding within images.

- **NVIDIA:** In their exploration of robot manipulation, NVIDIA presents VIMA, a transformer-based agent capable of handling multimodal prompts encompassing textual and visual tokens. VIMA demonstrates a high level of scalability and data efficiency, outperforming existing designs in zero-shot generalization settings with a significant success rate, and maintaining strong performance with reduced training data [101]. RAVEN explores in-context learning using retrieval-augmented language models, enhancing few-shot learning efficiency [103]. NVIDIA's InstructRetro significantly boosts performance by pre-training a 43B GPT model on 1.2 trillion tokens, achieving notable improvements in QA and summarization tasks [102]. The Nemotron-4 340B model family, under the NVIDIA Open Model License, and the collaborative Megatron-Turing NLG 530B model by Microsoft and NVIDIA, demonstrate high performance on various NLP benchmarks [100], [104].
- **Alibaba:** Alibaba's advancements in MM-LLMs are highlighted through their development of Qwen and SeaMM-LLMs. The Qwen series, encompassing base pretrained language models and chat models fine-tuned with human alignment techniques, showcases exceptional performance across diverse downstream tasks. Qwen-Chat models, particularly those trained using Reinforcement Learning from Human Feedback (RLHF), demonstrate advanced tool-use and planning capabilities, competing effectively even with larger models on complex tasks such as utilizing a code interpreter. Specialized models like Code-Qwen, Code-Qwen-Chat, and Math-Qwen-Chat exhibit significantly enhanced performance, surpassing open-weights counterparts and approaching the capabilities of proprietary models [119]. SeaMM-LLMs addresses the linguistic biases inherent in many MM-LLMs by focusing on Southeast Asian languages. Built on the Llama-2 model, SeaMM-LLMs incorporate an extended vocabulary and specialized instruction tuning to capture the nuances of regional languages. This series respects local cultural norms and legal considerations, resulting in superior performance across a variety of linguistic tasks. SeaLLM-13b models notably outperform ChatGPT-3.5 in non-Latin languages like Thai, Khmer, Lao, and Burmese, demonstrating their effectiveness while maintaining cost efficiency by low computation requirement [120].

  Beyond widely discussed MM-LLMs such as GPT-4 and DeepSeek-R1, a new generation of large-scale general-purpose and multimodal language models has emerged with strong potential to impact agriculture. These include **Claude 3** (Anthropic) [Source Link], **Grok** (xAI) [Source Link], **Microsoft Copilot** [Source Link], **Google Gemini** [105], **Meta's LLaMA 3** [128], **Mistral 7B** [130], **Mixtral** [131], **PaLM 2** [132], **Command R (Cohere)** [Source Link], and **Qwen** [119]. While these models were not originally developed for agriculture, their advanced reasoning, multilingual capabilities, and multimodal input handling make them highly adaptable for agri-tech domains.

  Claude 3 is recognized for its advanced alignment and safety measures, making it ideal for regulated applications such as agricultural risk management, policy compliance, and advisory systems. **Grok**, designed to operate on real-time data from the web and social media platforms like X (formerly Twitter), could support dynamic market forecasting, pest outbreak alerts, and climate-sensitive recommendations [Source Link]. **Copilot**, integrated with Microsoft Office tools, streamlines agronomic documentation, automates spreadsheet analyses, and supports multilingual communication for global agricultural operations.

  Gemini 1.5, with its extended context window and multimodal fusion abilities, shows promise in processing complex, time-series data from drones, sensors, and satellite imagery, which are critical for precision agriculture. Similarly, **LLaMA 3** and **Mixtral**, both instruction-tuned for accuracy and efficiency, are well-suited for conversational interfaces, AI-powered extension services, and crop advisory chatbots. **Command R**, known for Retrieval-Augmented Generation (RAG), can synthesize agronomic reports and scientific documents with real-time evidence integration. **PaLM 2** and **Qwen** provide strong multilingual support and code-generation capabilities that are valuable in modeling crop simulations, deploying edge-AI models, and supporting rural advisory platforms. While limited peer-reviewed agricultural deployments exist for these models, their capabilities suggest significant applicability. Their integration into tools like **ChatAgri**, **CropDoctor-GPT**, **AgriBERT**, and **PLLaMa** would accelerate context-aware decision-making. Future work must focus on model grounding, agricultural fine-tuning, participatory design, and in-field validations to unlock their full impact across precision farming, climate resilience, and digital extension ecosystems.

B. *Vision-Language Models (VLMs)*

Multi-Modal Vision-Language Models (VLMs) extend the capabilities of traditional language models by incorporating multiple types of data, such as text, images, audio, and video [133]. These models represent a significant advancement in agricultural technology by integrating diverse data types to provide comprehensive insights and solutions [134]. By combining textual data with visual and sensory inputs, MM-MM-LLMs can offer more accurate and context-aware recommendations, thereby enhancing decision-making processes in agriculture [135].

Vision-Language Models (VLMs) [136] are a subset of multi-modal models that specifically focus on the integration of visual and textual information to understand and generate multi-modal content. By integrating computer vision and NLP techniques, VLMs can perform tasks such as image captioning, visual question answering [137], and image-based text generation [138]. One notable model is "GLaMM: Pixel Grounding Large Multimodal Model," which introduces a novel approach to vision-language models by focusing on

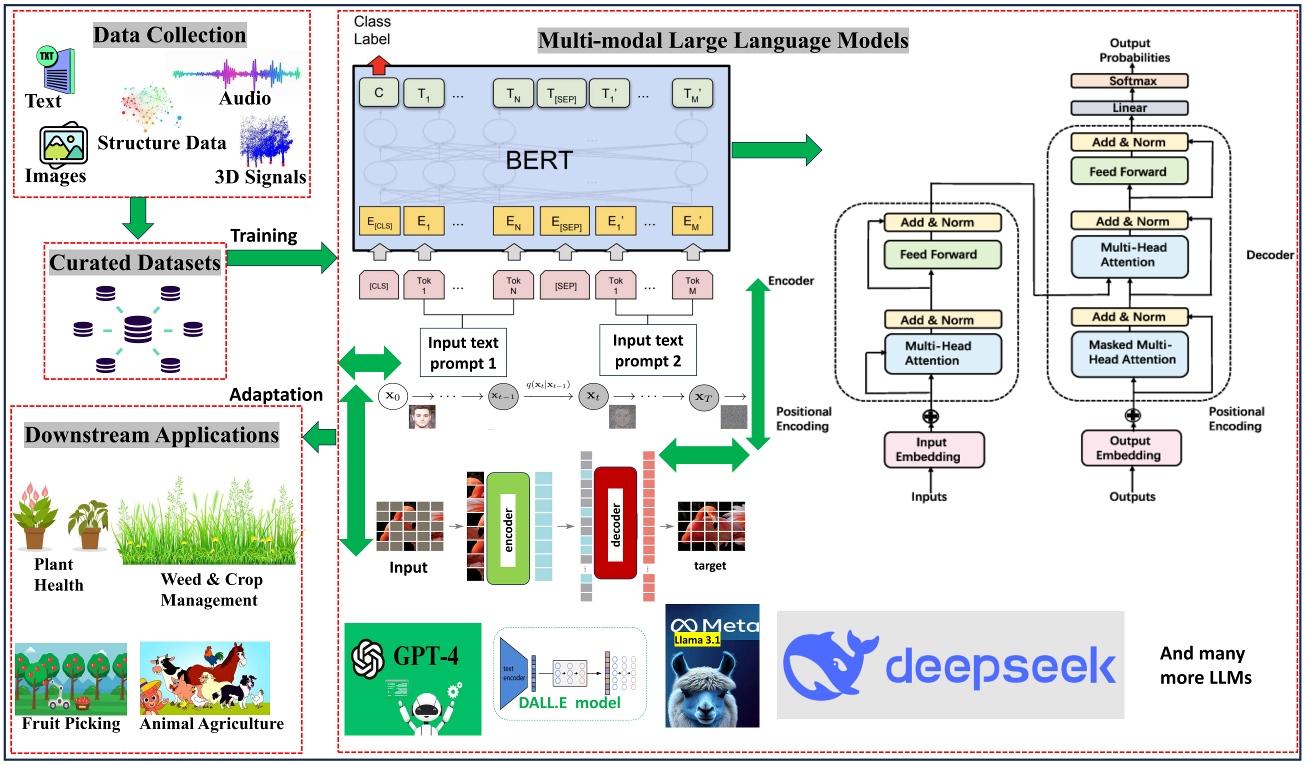

Fig. 5. Pipeline for developing multi-modal MM-LLMs in agriculture, illustrating (multi-modal) data collection, cleaning, training, fine-tuning, and downstream adaptation, modified from [134].

pixel-level grounding to enhance interaction and understanding between visual and textual modalities [139]. The GLaMM model extends the capabilities of MM-LLMs into the vision domain by generating natural language responses that are closely intertwined with corresponding object segmentation masks. This capability is particularly important in handling video data, where temporal dynamics add complexity to the analysis [140].

III. Applications of MM-LLMs in Agriculture

In agriculture, MM-LLMs have been applied across various domains including weed control, pest and disease detection and livestock management. The capability of these models to process extensive and diverse datasets renders them exceptionally useful for predictive modeling and performance analysis. Agricultural MM-LLMs are tailored to address specific challenges such as the limited availability and scattered nature of agricultural data. For example, ChatGPT has shown significant potential in agricultural text classification, effectively overcoming challenges related to data scarcity and cross-linguistic transferability [141]. Similarly, AgriBERT has been fine-tuned to enhance semantic matching between food descriptions and nutrition data, demonstrating its utility in improving agricultural NLP applications [142]. The integration of MM-LLMs with CNNs for diagnostics, as seen with GPT-4 and YOLOPC, exemplifies how these models can enhance pest detection accuracy while reducing model parameters [143].

Furthermore, MM-LLMs like GPT-4 have demonstrated their potential in educational and crop management contexts by successfully passing agriculture exams [144]. Techniques such as RAG and fine-tuning have been utilized to enhance MM-LLMs' efficiency in providing location-specific agricultural insights [145]. Additionally, the development of PLLaMa, an open-weights language model specifically designed for plant science, has significantly advanced the understanding of plant-related topics [146]. These advancements underscore the extensive adaptability and versatility of MM-LLMs in addressing complex agricultural challenges. Figure 5 presents a comprehensive overview of the pipeline for developing (multimodal) foundation models (FMs) in agriculture. This process encompasses stages such as data collection, dataset curation, model training, and downstream fine-tuning.

Table III provides a detailed summary of recent MM-LLMs published in peer-reviewed journals, summarizing the state-of-the-art surveys related to MM-LLMs in agriculture, highlighting their main findings, contributions, and impacts. Table IV presents a summary of LLM-related papers available on arXiv.

LLM integration enables precise monitoring and management of agricultural practices, optimizing resource use and improving crop yields. Furthermore, combining real-time sensor data and MM-LLMs predictive analytics enables dynamic adaptability to changing agriculture conditions, thereby improving forecasting accuracy [134], [170]. In this section, we review the integration of MM-LLMs with proximal sensing technologies in agriculture, categorized into the following subsections: MM-LLMs for Crop Monitoring and Management, MM-LLMs for Disease Detection and

TABLE I
ABBREVIATIONS AND ACRONYMS

Abbreviation	Full Term
AI	Artificial Intelligence
AOM	Action-Object-Member
AUPRC	Area Under the Precision-Recall Curve
AUROC	Area Under the Receiver Operating Characteristic
BERT	Bidirectional Encoder Representations from Transformers
CAP	Common Agricultural Policy
CNN	Convolutional Neural Network
CoT	Chain-of-Thought
CRF	Conditional Random Field
DL	Deep Learning
DSSAT	Decision Support System for Agrotechnology Transfer
FMs	Foundation Models
GANs	Generative Adversarial Networks
G2F	Genomes to Fields
GPT	Generative Pre-trained Transformer
HMM	Hidden Markov Model
IoT	Internet of Things
KCC	Kisan Call Center
LDA	Latent Dirichlet Allocation
MM-LLMs	Large Language Models
LSTM	Long Short-Term Memory
MAP	Mean Average Precision
MAPE	Mean Absolute Percentage Error
MEMs	Maximum Entropy Models
ML	Machine Learning
MM-MM-LLMs	Multimodal Large Language Models
MoE	Mixture of Experts
MSE	Mean Squared Error
NER	Named Entity Recognition
NLP	Natural Language Processing
PA	Precision Agriculture
PLMs	Pre-trained Language Models
QA	Question Answering
RAG	Retrieval Augmented Generation
RL	Reinforcement Learning
RLHF	Reinforcement Learning from Human Feedback
RNN	Recurrent Neural Network
RQ	Research Question
SVM	Support Vector Machine
TTS	Text-to-Speech
UAV	Unmanned Aerial Vehicle
VLMs	Vision-Language Models
YOLO	You Only Look Once

TABLE II
RECENT MM-LLMS FROM MAJOR AI LEADING COMPANIES: AMAZON, OPENAI, META, GOOGLE, MICROSOFT, APPLE, NVIDIA, AND ALIBABA

Lab Name	Model Name	Parameters (B)	Tokens Trained (B)	Paper Reference
OpenAI	GPT-1	0.12	0.005	[75]
OpenAI	GPT-2	1.5	Not specified	[76]
OpenAI	GPT-3	175	570	[23]
OpenAI	GPT-4	Not specified	1,800	[77]
OpenAI	GPT-4o	Not specified	Not Specified	[78]
OpenAI	CriticGPT	Not specified	Not specified	[79]
DeepSeek	DeepSeek-R1	671	14800	[80]
Amazon	Olympus	2000	40000	-
Amazon	HLAT	7	Not specified	[81]
Amazon	Multimodal-CoT	Not specified	Not specified	[82]
Amazon	AlexaTM 20B	20	Not specified	[83]
Meta AI	Chameleon	34	9200	[84]
Meta AI	Llama 3 70B	70	2000	[85]
Meta AI	Llama 2	70	2400	[86]
Meta AI	LIMA	65	Not specified	[87]
Meta AI	LLaMA-65B	175	300	[88]
Meta AI	BlenderBot 3x	150	300	[89]
Meta AI	Atlas	11	40	[90]
Meta AI	OPT-175B	175	300	[91]
Meta AI	InCoder	6.7	Not specified	[92]
Apple	4M-21	3	Not specified	[93]
Apple	Apple On-Device model	3.04	1500	[94]
Apple	MM1	30	2080	[95]
Apple	ReALM-3B	3	134	[96]
Apple	Ferret-UI	13	2000	[97]
Apple	MGIE	7	2000	[98]
Apple	Ferret	13	2000	[99]
NVIDIA	Nemotron-4 340B	340	9000	[100]
NVIDIA	VIMA	0.2	Not specified	[101]
NVIDIA	Retro 48B	48	1200	[102]
NVIDIA	Raven	11	40	[103]
NVIDIA	MT-NLG	530	270	[104]
Google DeepMind	Gemini 1.5	Not specified	Not specified	[105]
Google DeepMind	Med-Gemini-L 1.0	1500	30000	[106]
Google DeepMind	Hawk	7	300	[107]
Google DeepMind	Griffin	14	300	[107]
Google DeepMind	Gemma	7	6000	[108]
Google DeepMind	Gemini 1.5 Pro	1500	30000	[105]
Google DeepMind	PaLi-3	6	Not specified	[109]
Google DeepMind	RT-X	55	Not specified	[110]
Google DeepMind	Med-PaLM M	540	780	[111]
Microsoft	MAI-1	500	10000	[112]
Microsoft	YOCO	3	1600	[113]
Microsoft	phi-3-medium	14	4800	[114]
Microsoft	phi-3-mini	3.8	3300	[114]
Microsoft	WizardLM-2-8x22B	141	Not specified	[115]
Microsoft	WaveCoder-Pro-6.7B	6.7	20	[116]
Microsoft	WaveCoder-Ultra-6.7B	6.7	20	[116]
Microsoft	WaveCoder-SC-15B	15	20	[116]
Microsoft	OCRA 2	7, 13	Not specified	[117]
Microsoft	Florence-2	Not specified	5.4 (visual annotations)	[118]
Alibaba	Qwen	72	3000	[119]
Alibaba	SeaLLM-13b	13	2000	[120]
Anthropic	Claude 3 Opus	137	200,000	[121]
xAI	Grok-1.5	Not specified	128,000	[122]

Diagnosis, MM-LLMs for Precision Irrigation and Fertilization, and MM-LLMs for Agricultural Information and Decision Support Tools.

1) *MM-LLMs for Crop Monitoring and Management:*

Recent advancements in language models such as GPT-4 have shown significant potential in transforming crop production by providing on-demand agronomic expertise [155], [171], Kuska et al. [172] explored the potential of MM-LLMs in revolutionizing agricultural practices. The authors highlight four primary use cases for MM-LLMs in agriculture: consulting and assistance, automated documentation, explanation and education, and interpretation of machine learning (ML) results and forecasts. These use cases include providing tailored agricultural recommendations, translating machine-generated data into accessible text, generating user-friendly guides and tutorials, and converting complex ML outputs into comprehensible information for decision support in areas such as plant disease management and precision agriculture.

Wu et al. [157] explored the integration of reinforcement learning (RL) with LMs for advanced crop management. The study utilizes the Decision Support System for Agrotechnology Transfer (DSSAT) to simulate crop management scenarios, employing a deep Q-network for training management policies. The innovative aspect lies in converting

TABLE III

APPLICATIONS OF MM-LLMS IN AGRICULTURAL INFORMATION DIAGNOSTICS AND HUMAN-MACHINE INTERACTION (PEER-REVIEWED)

Reference	MM-LLMsModel Used	Main Findings, Contributions, and Impact
[143]	GPT-4, YOLOPC	Achieved 94.5% accuracy in pest detection, integrating MM-LLMs with CNNs for advanced agricultural diagnostics.
[147]	BERT-based Seq2Seq	Enabled efficient synthesis of commands, significantly enhancing human-machine interaction in agricultural applications.
[148]	GPT-2, BERT, DistilBERT	BERT outperformed others in sentiment analysis, proving viable for classifying technical agricultural text accurately.
[149]	AgriBERT	Demonstrated superior semantic matching results, enhancing food-nutrition mapping in agricultural contexts.
[150]	Extended BERT with LDA	Achieved 80.43% F-measure in NER, significantly improving agricultural entity recognition accuracy.
[141]	ChatGPT (GPT-3.5, GPT-4)	Effective cross-linguistic text classification, addressing data scarcity and transferability issues in agriculture.
[22]	Generative Pre-trained Transformer (GPT)	Simplified scientific knowledge for farmers, offering an idealized MM-LLMs design process for safe agricultural use.
[151]	ChatGPT	Provided real-time information access, enhancing farm management and simplifying regulatory processes in agriculture.
[152]	GPT-3.5	Enabled real-time pest identification, promoting sustainable farming practices through advanced AI integration.
[153]	ChatGPT	Improved surveillance of agricultural injuries, advancing public health through AI-driven insights.
[154]	GPT-3, LSTM	Enhanced soybean news analysis, leading to improved sentiment analysis accuracy in agricultural contexts.
[31]	YOLOv8, RAG	Achieved precision in coffee disease remediation by combining object detection with MM-LLMs for advanced diagnostics.
[155]	ChatGLM, Llama 2-13B	Improved knowledge service modes, enhancing agricultural technology extension services for farmers.
[27]	Generative AI, MM-LLMs	Provided advisory services for risk assessment, integrated into the smart agriculture platform ZEMELA.
[156]	YOLO11, YOLOv10, DALL-E	Generated annotated dataset with DALL-E; YOLO11 outperformed YOLOv10 in precision and validation.
[157]	RL, LM (DSSAT)	Advanced crop management system with RL, LM; enhanced economic profit and sustainability.
[158]	AgRoBERTa	Introduced AgXQA and AgRoBERTa, significantly outperforming GPT 3.5 in domain-specific QA tasks.
[159]	Meta-LLaMA 3.1, Agricultural-BERT, BERT-based-uncased	Used ensemble method for better query resolution; achieved 93% accuracy and high BLEU, ROUGE scores.
[160]	ChatGPT (GPT-3.5 Turbo, GPT-4)	ChatGPT evaluated for plant disease risk forecasting; GPT-4 superior for strategic advice.
[161]	N/A	Introduced a dataset from Norwegian agricultural websites for LLM training in agronomic practices.
[162]	RL, LLM	Explores integrating RL and MM-LLMs for crop production optimization, offering adaptive solutions and informed decision-making.

state variables from the simulator into language that LMs can understand, facilitating better decision-making. Empirical results from simulations with maize crops in Florida and Spain demonstrate that this approach not only improves economic profitability by over 49% but also reduces environmental impact compared to baseline methods. This study highlights the potential of combining RL and LMs to optimize crop management practices effectively.

TABLE IV

APPLICATIONS OF MM-LLMS IN AGRICULTURAL INFORMATION DIAGNOSTICS AND HUMAN-MACHINE INTERACTION (ARXIV) AS OF 07/04/2024

Reference	MM-LLMs Model Used	Main Findings, Contributions, and Impact
[142]	AgriBERT, GPT, ChatGPT	Improved semantic matching in food domains, enhancing agricultural NLP applications and precision in food-nutrition mapping.
[33]	ChatGPT-style MM-LLMs	Highlights risks and opportunities of MM-LLMs, proposing frameworks for responsible use in agricultural settings.
[144]	GPT-4, Llama 2, RAG, ER	GPT-4 passes agriculture exams, significantly improving educational methods and crop management practices.
[146]	PLLaMa (LLaMa-2 based)	Enhanced understanding in plant science, providing an open-weights model tailored for agricultural research and applications.
[145]	GPT-4, Llama 2-13B	Demonstrated benefits of RAG and fine-tuning, offering improved geographic-specific insights for agriculture.
[163]	Various MM-LLMs	Enhanced explainability in Decision-Driven Technologies (DDTs), aiding in smart agricultural decision-making processes.
[164]	BERT-based models	Classified Plant Health Bulletins accurately, improving plant health management and response strategies.
[165]	YOLO11, Segment Anything Model (SAM)	Introduced a zero-shot, LLM-generated method for automatic annotation and instance segmentation of apples, eliminating the need for manual annotation and field imaging. Demonstrated high accuracy with a Dice Coefficient of 0.9513 and IoU of 0.9303 on real orchard images, showcasing significant advancements in agricultural AI.
[28]	General MM-LLMs	Enabled accurate pest identification, efficiently extracting structured data from agricultural documents.
[166]	MoCo-V2, CNN, Swin Transformer	Improved dataset for agricultural analysis, enhancing aerial image analysis in precision farming.
[167]	TimeGPT	Achieved accurate soil moisture forecasting, advancing time-series models for smart agriculture applications.
[168]	GPT-4o	Enhanced visual perception in piglet activity, improving animal behavior recognition and welfare monitoring.
[169]	LSTM-RASA	Developed an effective farm assistant chatbot, providing timely and expert advice for farmers.

Veena et al. [150] propose a novel approach for Named Entity Recognition (NER) tailored for the agricultural domain. Their study addresses the challenges of developing accurate NER models in agriculture, such as the lack of annotated data, domain-specific vocabulary, and contextual variation. The proposed model leverages an extended BERT model with Latent Dirichlet Allocation (LDA) topic modeling to perform unsupervised NER, focusing on six key entities: disease, soil, pathogen, pesticide, crops, and place. To overcome the absence of a benchmark dataset, the authors created a corpus of 30,000 sentences from recognized agricultural sources and evaluated their model using a test corpus of 700 sentences. The AGRONER model achieved a macro average F-measure of 80.43%, demonstrating its effectiveness. The approach shows significant improvement over BERT without LDA, highlighting its potential for unsupervised domain-specific entity recognition in agriculture.

Zhang et al. [147] address the user-friendliness issues of existing agricultural measurement and control platforms. They propose a NLP pipeline to enhance human-machine interaction. The pipeline uses a dynamic tuple language framework to synthesize natural language commands into structured AOM statements. An end-to-end framework with a special mask mechanism improves the BERT-based Seq2Seq model's ability to capture global sequence relations. Experimental results demonstrate the pipeline's efficacy, showing good performance and reasonable response times.

Likewise, the "My Climate Advisor" system, presented by Nguyen et al. [173], employs NLP to aid climate adaptation in agriculture by providing relevant, trustworthy information to farmers and advisors. This question-answering prototype synthesizes information from peer-reviewed scientific literature and high-quality, industry-relevant grey literature, such as technical reports, white papers, and policy documents, to generate comprehensive, referenced answers. Utilizing open-weights generative models for privacy and intellectual property protection, it combines retrieval augmented generation for precise, verifiable answers. A novel evaluation framework with seven metrics was developed to assess the system's effectiveness, with initial user studies indicating promising results in usability and applicability [173].

2) MM-LLMs for Disease Detection and Diagnosis:

Lu et al. [174] presented a sophisticated approach using multimodal transformer models to enhance the detection of agricultural diseases and provide accurate responses to related queries. By integrating various data modalities, such as images and text, the proposed model leverages advanced ML techniques to identify plant diseases with high precision.

Hue et al. [175] introduce a groundbreaking approach to plant disease diagnosis leveraging the capabilities of GPT-4's multimodal model. The system integrates image recognition with NLP to provide accurate diagnostic information for plant diseases. Utilizing an extensive knowledge base of 1,420 host plants, 2,462 pathogens, and 37,467 pesticide instances, the GPT-4-based system offers rapid and precise identification of plant diseases. This method addresses the shortcomings of traditional visual inspections and complex molecular-based diagnostics by combining the speed and accessibility of ML with the detailed accuracy of molecular data, facilitating effective integrated pest management.

Qing et al. [143] explored the application of GPT-based models in diagnosing plant diseases through image analy sis. By employing advanced image recognition technologies powered by GPT models, the system aims to provide quick and reliable diagnostics for various agricultural issues. The integration of GPT's NLP capabilities allows for the synthesis of diagnostic information and actionable recommendations. This approach is designed to enhance traditional diagnostic methods such as visual inspection and manual sampling, offering a more efficient and accurate solution for farmers and agricultural professionals to identify and manage plant diseases, ultimately improving crop health and yield. Kumar et al. [31] introduced an innovative AI-driven PA system to combat coffee leaf diseases, particularly in Karnataka. By integrating YOLOv8 for disease identification with RAG for context-aware diagnosis, the system addresses inherent challenges associated with MM-LLMs. Their methodology mitigates hallucinations in MM-LLMs, incorporating real-time monitoring and dynamic remediation strategies.

Madaan et al. [176] developed a mobile application leveraging DL models for early detection of crop diseases to aid Indian farmers. By examining MobileNetV2 and ResNet50 on a dataset of 39,131 images across 21 classes, MobileNetV2 achieved 97%-99% accuracy, proving optimal for mobile use due to its compact model size. The system not only detects diseases but also uses MM-LLMs to provide tailored recommendations, thereby enhancing crop protection and farmer profitability. Zhao et al. [141] propose an innovative approach for agricultural text classification using ChatGPT models. The study highlights the challenges in agricultural text classification, such as limited training data, poor domain transferability across languages, and the high cost of deploying large models. By using ChatGPT's advanced understanding capabilities through various prompting methods, ChatAgri demonstrates competitive performance against fine-tuned PLMs. The study finds that ChatGPT outperforms PLM fine-tuning methods in few-shot scenarios and showcases remarkable cross-linguistic transferability.

Yang et al. [177] investigate the feasibility of using MM-LLMs, such as GPT-4, in pest management. The study focuses on evaluating the quality of pest management advice generated by MM-LLMs, using GPT-4 to score the content on coherence, logical consistency, fluency, relevance, comprehensibility, and exhaustiveness. An expert system based on crop threshold data is integrated as a baseline to assess factual accuracy. Results indicate that GPT-3.5 and GPT-4 outperform Google's FLAN models in most evaluation categories. The use of instruction-based prompting with domain-specific knowledge achieves an accuracy rate of 72%, demonstrating the effectiveness of MM-LLMs in providing reliable pest management suggestions.

3) MM-LLMs for Precision Irrigation and Fertilization:

The integration of MM-LLMs in precision irrigation and fertilization practices marks a significant advancement in agricultural technology. MM-LLMs have demonstrated remarkable potential in handling complex sets of information, generating insights, and providing sophisticated reasoning for decision-making processes. Recent advancements have developed a sophisticated crop management system that integrates RL, MM-LLMs, and crop simulations using tools like DSSAT [157]. This system employs deep RL, specifically a deep Q-network, to train policies that interpret various state variables from the simulator as observations. A key innovation is translating these state variables into descriptive language, enhancing the MM-LLMs' ability to comprehend and optimize management practices. Experimental results with maize crops in Florida (US) and Zaragoza (Spain) demonstrate that the MM-LLMs not only achieve top performance across multiple metrics but also significantly boost economic profits by over 49% while reducing environmental impacts compared to traditional methods.

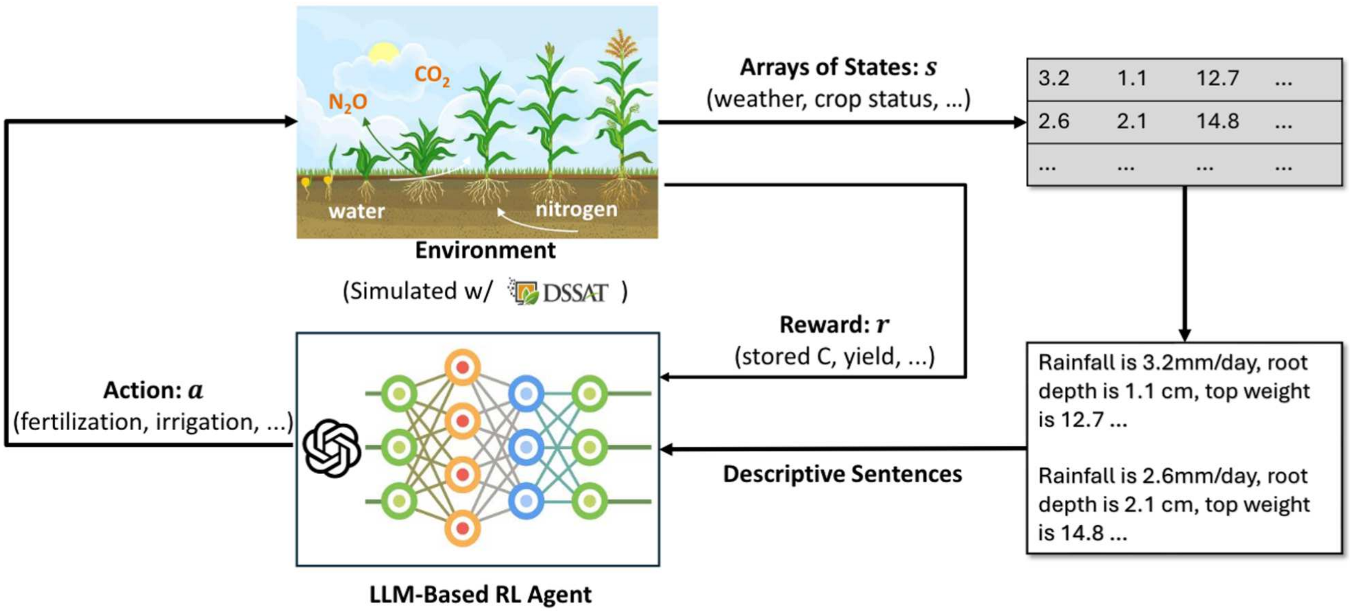

Fig. 6. LLM-based reinforcement learning framework for fertilization and irrigation [157].

The deployment of MM-LLMs as RL agents has shown promising results in various domains, including robotics and user interface interactions [178], [179], [180]. For instance, replacing traditional multi-layer perceptrons (MLPs) with a distilled and pre-trained BERT model as the RL agent has significantly improved decision-making processes [181]. In this architecture, the BERT model encodes state variable sentences into feature embeddings, which are then transformed through fully connected layers to align with the action space of the RL agent. This setup ensures that complex relationships within crop management data are effectively captured and translated into actionable insights. Another notable example is presented in a framework that incorporates a powerful MM-LLMswith crop simulations via tools such as DSSAT and Gym-DSSAT as presented in Figure 6. In their approach, traditional numerical state representations from simulation tools are transformed into descriptive sentences. This transformation allows the use of MM-LLMs, which encode these sentences into embeddings that capture a more nuanced understanding of the states. By employing the advanced cognitive capabilities of MM-LLMs, such as their reasoning and memory functions [182], [183], the RL agent within this framework can comprehend complex aspects of crop growth and simulation environments more effectively.

4) MM-LLMs for Agricultural Information and Decision Support Tools:

Zhao et al. [141] investigate the application of ChatGPT for classifying agricultural texts across different languages. The study addresses challenges such as limited training data and poor domain transferability. By employing various prompting methods, ChatAgri not only demonstrates competitive performance compared to PLM fine-tuning methods but also excels in few-shot scenarios, outperforming PLM fine-tuning with limited data.The research highlights ChatGPT's ability to handle cross-linguistic text classification effectively, making it a robust solution for organizing and accessing the vast amounts of agricultural information available online. This study underscores the potential of ChatGPT to revolutionize agricultural text processing, paving the way for more efficient and accessible knowledge dissemination in the field of sustainable, smart agriculture.

KissanAI, launched on March 15, 2023, is an AI chatbot designed to support India's agricultural sector by providing verbal interaction capabilities for farmers with varying literacy levels. It leverages OpenAI's ChatGPT technology and a proprietary knowledge base to offer advice on crop cultivation, pest management, and irrigation in multiple languages, including Hindi, English, and Gujarati. This study emphasizes its role in bridging the knowledge gap between farmers and experts, enhancing agricultural productivity and sustainability [184].

5) MM-LLMs for Agricultural Extension and Technical Advisors:

In the United States, public sector advisors associated with Land Grant universities are traditionally known as 'Extension' educators. These roles vary globally; in other countries, these professionals may be referred to as extension agents, specialists, or consultants [185]. For example, in China, there are over 600,000 agricultural extensionists providing support. Globally, extension services assist over 570 million farmers who cultivate 24% of the world's agricultural land, producing 29% of its crop food kilocalories [22]. In the U.S., Extension advisors typically operate locally, often within a specific county or geographic region, close to the agricultural producers they serve. Frequently, the Land-Grant universities that employ these advisors also have faculty research and extension specialists connected to or directly funded through Extension but who have a more significant role in leading the applied research that serves as the knowledge and information source for those local advisors [186]. In this section, we expand the conversation to include other groups who may work in roles similar to Extension advisors. These may be other public sector advisors or those who work in the private sector. Examples of agricultural technical advisers who will likely benefit from using MM-LLMs include agricultural consultants, agronomists, veterinarians, or others offering highly specialized advice and determine agricultural profitability, such as farm financial specialists or attorneys. Considering this broad spectrum of technical advisers that includes Extension advisors and others, we will refer to the collective group as a community of Technical Advisors (TAs). We envision the following several roles that MM-LLMs could play in the work of TA's. Among them are tools that have the potential to significantly enhance decision-making, empowering TA's to make more informed and confident choices.

- Enhance the TA's ability to lead, participate in, or analyze and interpret the results of applied agricultural research.
- Assist TA with knowledge translation into usable tools for stakeholders, increasing access, and aiding in content creation.
- Improve work efficiency and processes for TAs to enable higher-level focus on relationships and interpersonal communication with stakeholders

This following points will also talk about some of the caveats, cautions, risks, and pitfalls that must be considered as TAs adopt MM-LLMs as an everyday tool of practice.

a) Assist in research leadership and interpretation: Technical Advisors (TAs) in agriculture, who integrate multidisciplinary scientific knowledge, can leverage MM-LLMs to enhance their advisory roles. These models enable efficient management of data-driven tasks such as weather forecasting, pest monitoring, and crop condition analysis [144]. MM-LLMs can process and integrate multimodal data from various sources like satellites and UAVs to aid precision farming [187]. Additionally, they can facilitate on-farm applied research by using machine vision for pest identification and creating actionable recommendations for TAs to convey to producers [152]. By improving efficiency and extending their operational capacity, TAs can better analyze trends and risks, ultimately reaching a broader audience [157], [172].

b) Assist in knowledge translation and content creation: MM-LLMs significantly enhance the speed and accuracy of translating complex scientific knowledge into practical, personalized advice for agricultural producers [134], [188]. These tools excel in synthesizing insights from diverse data sources, enabling TAs to provide comprehensive management recommendations [145]. Studies in India and Nigeria show positive farmer reception towards AI for tasks like weather prediction and disease management, with AI responses often surpassing those from human TAs in terms of quality and timeliness [187], [189]. MM-LLMs offer the additional benefit of multilingual translation, facilitating immediate and specific advice tailored to individual needs, which is essential in resource-limited settings [22], [151], [152], [157].

c) Improve work efficiency and processes: Often, TAs in public and private sectors who work as consultants, Extension advisors, or in other roles encounter various administrative tasks that take away from their ability to build relationships, learn, and communicate valuable technical information. MM-LLMs can help automate program scheduling, registration, record keeping, drafting technical reports, follow-up emails, and other administrative communication [190]. Increases in time efficiency and savings will allow the TA to spend more time on critical interactions and relationship building, allowing stakeholder needs to be addressed more effectively. MM-LLMs and other AI applications create new opportunities for entrepreneurs who wish to offer high-value and innovative solutions at affordable prices, making research and data-driven practices more accessible for a broader range of production operations irrespective of their size, scale, and production type [187].

d) Potential threats and pitfalls: As we consider using MM-LLMs and other highly accessible AI tools to support and enhance the work of TAs, we must also consider possible risks, pitfalls, or threats. For example, MM-LLMs are complex and require considerable computing capabilities. Models must be continuously trained and updated with new research findings, models, and data, and the computational costs and complexity could limit their widespread application, especially on highly specialized issues where "general" GPTs are not entirely suitable [134]. To be fully trusted by both the TA and their stakeholders, transparency is critical; MM-LLMsmust be able to explain its reasoning and limit its output to areas where queries can be answered with a high degree of certainty, or there will be opportunities to misuse these tools or generate recommendations that have unintended consequences [22], [190]. To be effective, these tools will need the latest and highest quality technical information as training data. This leads to questions about data privacy, security, intellectual property concerns, and prevention of unauthorized access to the raw training data [134]. To the extent that MM-LLMs are trained on available data from geographic areas or groups that have the resources to generate that data (time, expertise, financing, etc.), we must recognize and understand that there will always be biases in recommendations, potentially affecting producers and other stakeholders if not corrected and accounted for [191]. For example, a model that is trained largely on US resources or even more geographically specific areas (e.g., Midwestern US) will represent those biases and may or may not be applicable in other regions of the world. There are also concerns that the widespread use of MM-LLMs by TAs could lead to job loss and displacement workers. Of particular note is that generative AI in agriculture could displace extension service workers or other agricultural professionals [190] all together.

A. Commercialization of MM-LLMs by Private Companies in Agriculture

The rapid adoption of MM-LLMs and AI in agriculture is transforming the industry, offering solutions to address the needs of a growing world population [192]. Various companies are developing AI chatbots, copilots, and specialized models to assist farmers, agronomists, and other stakeholders with a wide

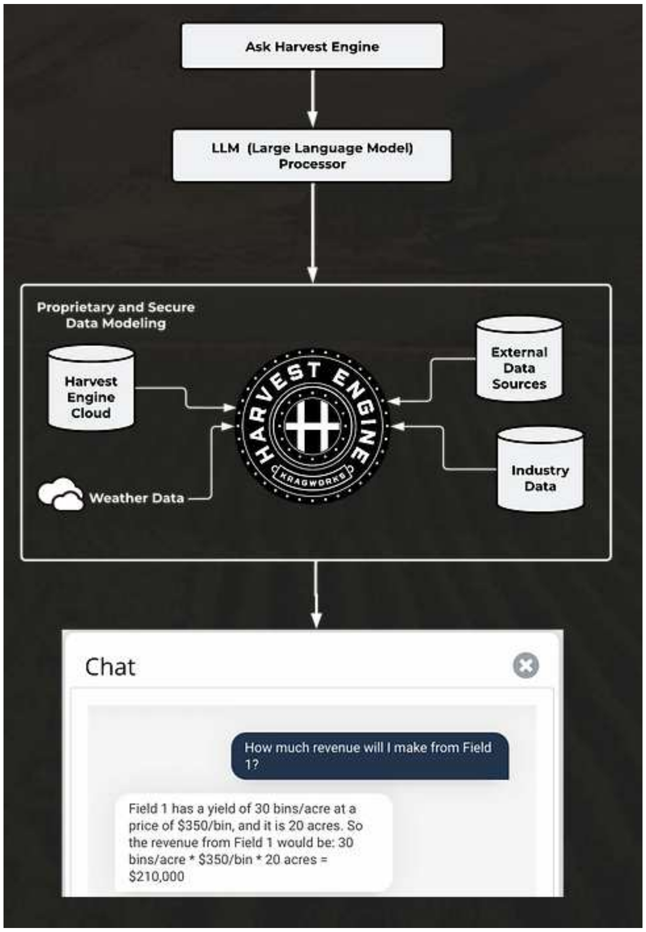

Fig. 7. Harvest Engine® (a packhouse management software based on MM-LLMs) developed by Kragworks Ag Solution (Yakima, WA) [197]. This LLM -based platform integrates diverse agricultural data to provide real-time, conversational insights.

range of tasks, from crop management and decision-making to equipment maintenance and market analysis [192], [193], [194], [195].

Several companies are at the forefront of this technological revolution. Kragworks Ag Solutions (Washington, USA) is developing Harvest Engine® (Figure 7, which is being tested on orchard blocks operated by Northwest Farm Management (NWFM) [193]. Farmers Business Network (South Dakota, USA) has launched a chatbot called Norm for general agricultural queries, while AGvisorPRO has introduced MM-LLMs for equipment dealers [193], [194]. Bayer (Bayer AG, Leverkusen, Germany) and ITC have showcased their AI innovations, with Bayer's Azure-based expert generative AI system and ITC's Krishi Mitra app serving millions of farmers in India [192]. Cropin Technology has been particularly active, launching 'akṣara', the sector's first purpose-built open-weights Micro Language Model for climate-smart agriculture, and partnering with Google's Gemini generative AI chatbot to create a tool that generates grid-based maps of agricultural land and provides localized insights [195], [196].

These AI-powered tools aim to provide quick, conversational access to information on diverse topics such as yield optimization, weather impacts, profitability, disease risks, labor needs, crop protection, and harvest windows [192], [193], [195]. They leverage data from multiple sources, including extension newsletters, research papers, farm-specific records, equipment manuals, and specialized datasets [192], [193], [194], [195], [196]. The potential benefits of MM-LLMs in agriculture are significant, including increased efficiency, improved decision-making, and enhanced sustainability. However, challenges remain, including concerns about occasional incorrect information or "hallucinated" responses by AI systems [193].

Researchers and developers are expanding these models to cover more comprehensive agronomic recommendations, incorporating data on crop health, weather forecasts, irrigation, and even generating detailed agricultural land maps [192], [195], [196]. The growth of AI in agriculture is driving a technology market predicted to reach USD13.8 billion by 2031 [192]. While industry experts are optimistic about the potential of MM-LLMs in agriculture, they emphasize the importance of a balanced approach, seeing the technology as a powerful tool to augment human expertise rather than replace it [193], [195]. The development of specialized, open-weights models like akṣara and partnerships with advanced AI systems represent significant steps towards improving access to AI in agriculture for farming communities all over the globe including those in Global South [193], [195], [196].

Multi-modal MM-LLMs enhance agricultural productivity by providing advice for timely farming decisions and optimizing farming operations [18]. One significant application is in personalized agricultural advisory services, where models like GPT-4 generate location-specific, data-driven recommendations for farmers on pest control, irrigation, and fertilizer usage. Systems such as KissanGPT in India demonstrate the effectiveness of these models, particularly for farmers with low literacy, by providing verbal advice that outperforms traditional extension services in clarity and relevance [22]. Additionally, MM-LLMs excel in crop disease diagnosis and pest management [198], [199]. Multimodal MM-LLMs integrate text and image processing to accurately identify diseases and pests, enhancing the timeliness and precision of interventions, as seen in platforms like PlantVillage Nuru [22].

MM-LLMs empower chatbots to swiftly access extensive agricultural databases by processing complex queries into actionable insights, outperforming traditional search methods. In precision farming, they merge real-time satellite and IoT data to deliver hyper-local advice [34]. For example, by analyzing soil, weather, and meteorological data, MM-LLMs offer tailored planting and irrigation recommendations that maximize yield and resource efficiency.

MM-LLMs significantly enhance agricultural operations by optimizing market analysis and supply chain logistics, advising farmers on optimal selling times and distribution to meet market demands. They also offer interactive, locally tailored education on sustainable practices and aid policymakers in modeling the impact of various sustainability strategies, ultimately boosting efficiency, profitability, and ecological alignment worldwide.

In agriculture, a variety of MM-LLMs are employed to enhance various aspects of farming by processing and analyzing vast amounts of data, and thus providing solutions tailored to agricultural needs. For instance, GPT-4 has been integrated with YOLOPC to achieve a high accuracy rate in pest detection by combining the strengths of MM-LLMs with CNNs [143]. Similarly, BERT-based Seq2Seq models are used to synthesize commands for improving human-machine interactions in agricultural operations [147]. Furthermore, specialized models like AgriBERT have been developed to address specific agricultural challenges such as semantic matching and food-nutrition mapping, demonstrating superior results in these niche areas [149]. Extended BERT models with enhancements like Latent Dirichlet Allocation are applied to improve named entity recognition (NER) within agricultural texts, significantly enhancing the accuracy of identifying key agricultural entities [150].

Moreover, the integration of MM-LLMs with real-time data systems is exemplified in models like LLM-Geo, which utilize satellite imagery and IoT sensor data to provide dynamic agricultural advice [200]. ChatGPT variants are specifically adapted to perform cross-linguistic classification tasks, aiding in the management of multilingual agricultural data [141]. These models are pivotal in transforming agricultural practices by providing scalable solutions across various domains from pest management to policy development, demonstrating the flexibility and expansive utility of MM-LLMs in enhancing both the efficiency and sustainability of agricultural ecosystems [22], [151].

IV. Discussion on Research Questions

A. Core Agriculture Research Questions

RQ1: Which sources and types of data are used to train MM-LLMs for agricultural applications, and how are they integrated? MM-LLMs for agricultural applications are trained on diverse and domain-specific datasets. For instance, genomic data from 48 plant species were utilized to train MM-LLMs for genomic predictions [201], while agricultural literature and Food ontology were integrated for food-nutrition matching tasks [149]. Additionally, GPT-4 has been integrated with YOLOPC models to enhance disease diagnostics by utilizing real-time sensor data and market trends [151]. Custom agricultural corpora, such as those containing 30,000 sentences, have been employed to improve named entity recognition (NER) models [150]. These diverse data sources, including genomic datasets, sensor data, and agricultural literature, form the foundation for training MM-LLMs, ensuring they are contextually aware and tailored to specific agricultural needs. **RQ2: What evaluation metrics are utilized to assess the performance of MM-LLMstasks in agriculture?** Evaluating the performance of MM-LLMs in agricultural applications involves various metrics tailored to specific tasks. For example, the precision, recall, and F1-score metrics were used to assess the performance of transformer-based DL models for agricultural entity extraction [202]. In genomic predictions, the accuracy of predictions was evaluated using the Area Under the Receiver Operating Characteristic (AUROC) curve, which measures the ability of the model to distinguish between classes, and the Area Under the Precision-Recall Curve (AUPRC), which assesses the model's precision and recall capabilities [201]. The effectiveness of sentiment analysis models employed for evaluating smart farming technologies, such as generative pre-trained transformer 2 (GPT-2) and domain-specific BERT, was quantified using metrics such as the macro-F1-score and accuracy, which assess overall classification performance across various sentiment categories [148]. Additionally, the effectiveness of NLP systems for climate adaptation in agriculture was evaluated using domain-specific criteria in a user study, demonstrating moderate agreement among experts [173]. These metrics ensure that MM-LLMs are rigorously tested and validated across diverse agricultural applications.

RQ3: What prompt engineering techniques are applied to improve the performance of MM-LLMs in agriculture? Prompt engineering plays a crucial role in enhancing the performance of MM-LLMs in agricultural applications. Techniques such as task-specific prompting and optimized answer alignment have been used to improve semantic understanding and reasoning accuracy in agricultural text classification [141]. Custom prompts were also developed to enhance injury surveillance reports, streamlining the data extraction process [153]. Additionally, few-shot prompting was employed for fine-tuning Multilingual BERT, enabling it to adapt to diverse agricultural datasets with minimal training [203]. In another approach, task-specific prompts were tailored to capture sentiment patterns in agricultural product news, significantly improving the accuracy of sentiment analysis [154]. These techniques underscore the importance of prompt engineering in maximizing the efficacy of MM-LLMs in various agricultural contexts.

RQ4: Which optimization strategies are applied in agriculture to improve the efficiency and effectiveness of MM-LLMs? Optimization strategies are essential for enhancing the efficiency and effectiveness of MM-LLMs in agriculture. Techniques such as parameter-efficient fine-tuning using IA3 have been applied in plant genomics, enabling broader genomic predictions with minimal computational resources [201]. In another study, a mask-based self-attention mechanism was utilized to enhance global context understanding in sequence-to-sequence learning for natural language interfaces in agriculture [147]. Additionally, real-time monitoring and collaborative dataset expansion were implemented to ensure system adaptability in diverse agricultural settings, particularly in coffee leaf disease remediation [31]. The use of transformer-based DL models with Conditional Random Field (CRF) for entity extraction has also shown improved performance in agricultural tasks such as identifying crop types, disease names, and agricultural chemicals [202]. These strategies highlight the ongoing efforts to optimize MM-LLMs for more efficient and effective agricultural applications.

RQ5: What are the challenges and limitations of using MM-LLMs in agriculture? The deployment of MM-LLMs in agriculture faces several challenges and limitations. One significant challenge is the scarcity of high-quality, domain-specific training datasets, which affects the accuracy and reliability of MM-LLMs [147], [203]. Additionally, MM-LLMs often encounter difficulties in handling diverse and unstructured agricultural data, such as injury surveillance reports or regional-specific adaptation advice for climate resilience [173], [204]. Ethical concerns, such as bias and intellectual property issues, also pose limitations in the use of MM-LLMs for agricultural safety and health [205]. Furthermore, the high computational cost and environmental impact of training large models, along with challenges in real-time deployment on mobile devices, remain critical barriers to their widespread adoption in agriculture [206]. Addressing these challenges is crucial for the continued advancement of MM-LLMs in the agricultural sector.

RQ6: What are the benefits of using MM-LLMs instead of traditional ML and DL methods in agriculture? MM-LLMs offer significant advantages over traditional ML and DL) methods in agriculture. One of the key benefits is their ability to handle complex, context-dependent tasks, such as semantic matching in food-nutrition tasks or real-time decision-making in precision farming [149], [151]. MM-LLMs also provide more accurate and context-aware solutions for tasks like agricultural entity extraction, where traditional models might struggle with domain-specific terminology [202]. Additionally, MM-LLMs can leverage zero-shot learning capabilities, enabling them to perform well on tasks with limited labeled data, which is often a challenge in agricultural applications [201]. The integration of MM-LLMs with other technologies, such as GPT-4 with YOLOPC for diagnostics, further enhances their utility by reducing model parameters while improving diagnostic accuracy [143]. Furthermore, MM-LLMs are adept at processing multimodal inputs, integrating data from diverse sources like satellite images, sensor data, and textual information, which greatly enhances the accuracy of predictions and insights in agriculture. This ability to synthesize information from multiple modalities allows for more robust and comprehensive decision-support systems. By harnessing these capabilities, MM-LLMs can provide nuanced understandings of agricultural environments, aiding in more precise and informed agricultural practices

RQ7: How do MM-LLMs compare with ML and DL-based approaches in their applications in agriculture? MM-LLMs have consistently outperformed traditional ML and DL-based approaches across various agricultural applications. For instance, the domain-specific pre-trained Agricultural BERT model achieved higher macro-F1-score and accuracy in text classification tasks compared to other transformer models [148]. Similarly, MM-LLMs demonstrated superior performance in NER tasks for agricultural data, with models like GPT-4 and Multilingual BERT leading in accuracy and adaptability [203]. In another study, MM-LLMs outperformed traditional models in sentiment analysis of soybean-related news, achieving high accuracy in predicting sentiment trends [154]. Moreover, MM-LLMs have shown greater efficiency in real-time agricultural information diagnostics, significantly improving accuracy with reduced computational resources. These examples highlight the effectiveness of MM-LLMs in surpassing traditional methods, making them invaluable tools in modern agricultural research and applications.

Refer to Table V for a detailed correlation between RQs and peer-reviewed journal publications.

B. MM-LLMs for Agricultural Physical Systems

Multimodal MM-LLMs are set to transform agricultural decision-making by leveraging their ability to process and analyze extensive, complex data sets. These models synthesize inputs such as weather patterns, soil conditions, crop health, and market trends, offering farmers dynamic, actionable insights. For example, MM-LLMs enhance decision-making by predicting optimal planting schedules, recommending crop rotations, and advising on pest management strategies using a blend of historical data, present conditions, and relevant scientific findings. Additionally, MM-LLMs facilitate a reduction in manual labor by integrating with IoT and robotics to automate tasks like irrigation, fertilization, and harvesting. An LLM could, for instance, utilize real-time data from soil sensors and weather forecasts to refine irrigation schedules, thereby conserving water and cutting labor costs. Similarly, by analyzing drone-captured imagery, MM-LLMs can detect early signs of pest infestation, directing precise interventions and reducing the need for manual inspections.

However, the deployment of MM-LLMs is not without challenges. Their dependency on vast datasets might introduce biases that skew towards industrial farming, potentially overlooking smallholder farmers' practices. The opaque nature of many MM-LLMs also raises concerns about the transparency and accountability of their decisions. To mitigate these issues, it is crucial to curate diverse training data carefully, advance explainable AI techniques, and foster collaboration among AI researchers, agricultural experts, and farmers. This collaborative approach ensures the technology's relevance and ethical application across varied farming environments, promoting beneficial outcomes for all agricultural stakeholders.

V. MAJOR CHALLENGES AND LIMITATIONS

The implementation of MM-LLMs in agriculture offers considerable potential but also faces a range of substantial challenges and limitations as illustrated in the infograph diagram in Figure 8. These limitations can be categorized, roughly, into three primary areas: Data Quality Challenges, Data Availability Challenges, and Other Challenges. These sectors encompass a variety of obstacles that must be strategically overcome to fully harness the potential of MM-LLMs within the agricultural context. A critical issue is the scarcity of domain-specific, high-quality training data, such as gold standard corpora, which significantly hampers the accuracy and reliability of MM-LLMs in real-world agricultural settings. Often, studies like those focusing on agricultural term extraction or multilingual text classification rely on silver standard corpora or limited datasets, leading to potential biases and substantial gaps in model performance [141], [207]. Moreover, MM-LLMs frequently struggle to handle the diverse and constantly evolving agricultural terminologies, including synonyms, novel terms, and context-specific jargon [207]. The complexity of agricultural data, which spans textual reports to sensor data, requires robust NLP models capable of interpreting rich contextual information, a task that remains challenging [151], [153]. Additionally, the high computational costs associated with training and fine-tuning these models, combined

TABLE V

MAPPING OF RQS TO THE APPLICATION OF MM-LLMS IN AGRICULTURE. THIS TABLE SYSTEMATICALLY CATEGORIZES THE SELECTED LITERATURE BASED ON SEVEN KEY RQS, FOCUSING ON DATA SOURCES AND INTEGRATION, EVALUATION METRICS, PROMPT ENGINEERING TECHNIQUES, OPTIMIZATION STRATEGIES, CHALLENGES AND LIMITATIONS, BENEFITS OVER TRADITIONAL ML AND DL METHODS, AND COMPARATIVE PERFORMANCE OUTCOMES

Paper Reference	**RQ1: Data Sources and Integration**	**RQ2: Evaluation Metrics**	**RQ3: Prompt Engineering Techniques**	**RQ4: Optimization Strategies**	**RQ5: Challenges and Limitations**	**RQ6: Benefits over Traditional ML/DL**	**RQ7: Outperformance over ML/DL**
ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification [141]	Multi-source data: Internet news on pests, hazards, markets	Semantic understanding, reasoning accuracy in text classification	Task-specific prompting, optimized answer alignment strategies	Cross-linguistic transfer learning, zero-shot capabilities	Limited agricultural training data, deployment costs	Cost-effective, no need for supervised training data	Outperformed PLMs in multilingual zero-shot scenarios
A foundational LLM for edible plant genomes [201]	Genomic data from 48 plant species, focusing on crops	AUROC, AUPRC for evaluating genomic predictions	Masked language modeling, fine-tuning with IA3	Parameter-efficient fine-tuning via IA3 technique	Diverse genomic data, reference genome limitations	Zero-shot learning enables broader genomic predictions	Superior performance in plant genomics, zero-shot classification
AgriBERT: Knowledge-Infused Agricultural Language Models for Matching Food and Nutrition [149]	Domain-specific corpora, agricultural literature, FoodOn ontology	Precision@1, Mean Average Precision (MAP) used for assessment	Knowledge-based fine-tuning with external entities	Fine-tuning using a domain-specific agricultural corpus	External knowledge integration, domain-specific challenges	Enhanced semantic matching in food-nutrition tasks	Exceeded BERT in precision, effective answer selection tasks
GPT-aided diagnosis on agricultural image based on a new light YOLOPC [143]	Integration of GPT-4 with YOLOPC model for diagnostics	94.5% accuracy in pest detection, 90% reasoning accuracy	GPT-4 induced logical reasoning, report generation	Enhanced accuracy via partial convolution in YOLOPC	Image recognition accuracy, data scarcity challenges	Improved diagnostic accuracy with reduced model parameters	Greater efficiency in real-time agricultural diagnostics
AGRONER: An unsupervised agriculture named entity recognition using weighted distributional semantic model [150]	Custom agricultural corpus of 30,000 sentences	F-measure for evaluation, macro-average scoring	Extended BERT with LDA for entity recognition	Weighted distributional semantics, unsupervised NER	Lack of annotated data, domain-specific challenges	Effective unsupervised entity recognition, domain specificity	Outperformed standard models in agricultural NER tasks
Building Natural Language Interfaces Using Natural Language Understanding and Generation: A Case Study on Human–Machine Interaction in Agriculture [147]	Real-time agricultural command data collected via dynamic tuple extraction framework	Accuracy, precision, recall, and F-score metrics for NL2AOM task evaluation	BERT-based Seq2Seq model with special masking for sequence-to-sequence learning	Mask-based self-attention mechanism for enhanced global context understanding	Limited high-quality training datasets, order problem in sequence generation	High generalization capability, efficient data collection, structured AOM statement generation	Outperformed traditional models in accuracy, response time, and command parsing performance
AI-assisted sustainable farming: Harnessing the power of ChatGPT in modern agricultural sciences and technology [151]	Integration of real-time data from sensors, weather forecasts, and market trends	Accuracy, data-driven decision-making effectiveness	Tailored prompts for precise decision-making	Continuous learning with real-time data	Data quality, accessibility, and digital literacy	Real-time insights, personalized recommendations	Outperformed traditional models in precision farming and data interpretation
The potential of AI and ChatGPT in improving agricultural injury and illness surveillance programming and dissemination [153]	Integration of AI with text-intensive natural language reports from injury surveillance systems	Efficiency in data analysis, report generation, and dissemination	Custom prompts for injury surveillance and prevention reports	Enhanced workflows reducing human and financial resources	Data quality, system customization, and privacy concerns	Streamlined data analysis and reporting processes	Demonstrated improved outputs in injury surveillance and prevention
Harnessing the Power of GPT-3 and LSTM for NLPin Agricultural Product News: Focus on Soybeans [154]	Integration of GPT-3 for semantic scoring of soybean-related news headlines from online agricultural news sources	Performance evaluated using Mean Squared Error (MSE) on sentiment prediction accuracy of LSTM model	Utilized GPT-3 for initial sentiment analysis and LSTM for sequence dependency	Joint use of GPT-3 and LSTM to capture sentiment patterns in domain-specific news	Challenge in handling vast and diverse soybean-related news data, requiring robust NLP models	Combined NLP techniques showed substantial improvements in sentiment analysis accuracy for agricultural product news	Outperformed traditional sentiment analysis models, achieving high accuracy in predicting sentiment trends
Investigating the effect of different fine-tuning configuration scenarios on agricultural term extraction using BERT [207]	Comparison of Agriculture-BERT with SciBERT, RoBERTa, and vanilla BERT for extracting agriculture-related terms	Evaluated performance using memorization, synonym, and novel term test sets	Tested different fine-tuning configurations, including layer freezing and updating scenarios	Investigated the impact of updating embedding and encoder layers on term extraction performance	Scarcity of agriculture-related gold standard corpora; reliance on a silver standard corpus from AGRIS database	Agriculture-BERT outperformed other BERT models, especially in identifying synonyms and novel agricultural terms	Significant improvement in automatic agricultural term extraction using specialized fine-tuning strategies

TABLE V

(Continued.) MAPPING OF RQS TO THE APPLICATION OF MM-LLMS IN AGRICULTURE. THIS TABLE SYSTEMATICALLY CATEGORIZES THE SELECTED LITERATURE BASED ON SEVEN KEY RQS, FOCUSING ON DATA SOURCES AND INTEGRATION, EVALUATION METRICS, PROMPT ENGINEERING TECHNIQUES, OPTIMIZATION STRATEGIES, CHALLENGES AND LIMITATIONS, BENEFITS OVER TRADITIONAL ML AND DL METHODS, AND COMPARATIVE PERFORMANCE OUTCOMES

Paper Reference	**RQ1: Data Sources and Integration**	**RQ2: Evaluation Metrics**	**RQ3: Prompt Engineering Techniques**	**RQ4: Optimization Strategies**	**RQ5: Challenges and Limitations**	**RQ6: Benefits over Traditional ML/DL**	**RQ7: Outperformance over ML/DL**
Price Prediction of Agriculture Commodities Using Machine Learning and NLP [208]	Forecasting agriculture commodity prices using ARIMA, SARIMA, and RNN for Kannada-speaking farmers	Evaluated performance using Mean Absolute Percentage Error (MAPE)	Developed a voice bot to provide price forecasts in Kannada	Addressed language barriers using NLP for Kannada in rural India	Limited availability of region-specific data and language processing capabilities	Enhanced decision-making for farmers through accurate price forecasting in their native language	Demonstrated improved forecasting and accessibility in regional languages for rural farmers
MyClimate Advisor: An Application of NLP in Climate Adaptation for Agriculture [173]	Developed a QA tool for climate adaptation in agriculture using peer-reviewed and grey literature; achieved comparable scientific accuracy to proprietary models	Evaluated system with domain-specific criteria in a user study; moderate agreement observed among climate and agriculture experts	Integrated open-weights generative models for data privacy; RAG technique enhanced answer accuracy, though readability slightly decreased	Addressed data trustworthiness and region-specific adaptation advice for climate resilience; small models struggled with specific regional queries	Challenges in QA evaluation metrics specific to climate adaptation; scientific accuracy showed significant annotator disagreement	Improved accessibility to climate adaptation knowledge for farmers and advisors; user study showed promising usability results	Demonstrated the first study to provide tailored climate adaptation advice using MM-LLMs in agriculture; promising potential for future improvements
Enhancing Named Entity Recognition for Agricultural Commodity Monitoring with Large Language Models [203]	Developed a solution to improve NER for agricultural data using MM-LLMs, achieving high precision and recall in entity recognition	Evaluated system using F1 score across five key entities; MM-LLMs demonstrated strong performance, particularly GPT-4 and Claude v2	Utilized few-shot prompting for MM-LLMs and fine-tuned Multilingual BERT; MM-LLMs showed adaptability with minimal domain-specific training	Addressed challenges of diverse and unstructured agricultural data; MM-LLMs excelled in recognizing new and context-dependent entities	Limited by scarcity of specialized agricultural datasets; fine-tuned models performed better on specific tasks, but MM-LLMs required less maintenance	Showcased MM-LLMs' potential for real-time data analysis and decision support in agriculture; promising results for future deployments	Demonstrated MM-LLMs' effectiveness in NER tasks for agriculture, with GPT-4 and Multilingual BERT leading in accuracy and adaptability
Investigating the Use of Large Language Models in Agricultural Injury Surveillance [204]	Explored MM-LLMs for automating data extraction in agricultural injury surveillance, specifically within the AgInjuryNews database	Evaluated MM-LLMs such as ChatGPT and fine-tuned Llama2 against human-coded data with an average accuracy of 93%	Applied fine-tuning and prompt engineering to enhance data extraction accuracy in unstructured texts	Addressed challenges of accuracy and interpretability in extracting incident and victim data from diverse sources	Limited by MM-LLMs' occasional misidentification and reliance on high-quality input data from local journalism	Demonstrated potential for scaling injury surveillance with MM-LLMs, achieving near-human performance in coding incident reports	Validated MM-LLMs as viable tools for enhancing efficiency in agricultural injury data collection and analysis
Harnessing the Power of Large Language Models in Agricultural Safety & Health [205]	Explored the potential of MM-LLMs to support agricultural safety professionals by answering technical questions and interpreting research articles	Evaluated the accuracy and completeness of LLM-generated responses in agricultural safety contexts, highlighting the importance of prompt engineering	Implemented use cases in responding to safety queries, interpreting regulatory standards, and summarizing incident reports with generative AI	Addressed challenges of bias, intellectual property, and accuracy in the use of MM-LLMs in agricultural safety	Limited by risks of inaccurate or incomplete responses and ethical concerns in the application of MM-LLMs	Demonstrated the effectiveness of MM-LLMs in enhancing efficiency and supporting decision-making in agricultural safety	Emphasized the need for responsible use and continuous improvement of MM-LLMs to address ethical and practical challenges in the field
Overcoming MM-LLMsChallenges using RAG-Driven Precision in Coffee Leaf Disease Remediation [31]	Proposed an AI-driven system using YOLOv8 for disease identification and RAG for context-aware diagnostics in coffee production	Addressed MM-LLMslimitations by integrating RAG to enhance accuracy and reduce hallucinations in context-specific disease management	Implemented real-time monitoring, collaborative dataset expansion, and organizational involvement to ensure system adaptability in diverse settings	Focused on sustainable agriculture by promoting precise disease identification and reducing reliance on pesticides through eco-friendly practices	Demonstrated successful reduction in MM-LLMsinaccuracies using RAG, achieving reliable and context-aware solutions for PA	Highlighted challenges in model training, prompt formulation, and dataset handling, stressing expert input for refining model outputs	Laid the groundwork for future research focusing on scalability, reliability, and expanding disease management capabilities in coffee agriculture
Using MM-LLMs in Cyber-Physical Systems for Agriculture - ZEMELA [27]	Proposed integrating MM-LLMs into ZEMELA platform for smart agriculture, enhancing advisory services for farmers	Focused on assessing risks in agricultural projects considering Bulgarian legislation and program requirements	Summarized feasibility analysis and discussed two potential architectural approaches for MM-LLMsintegration	Aimed to refine operational capabilities of ZEMELA through integration, enhancing project preparation and implementation for farmers	Results indicated MM-LLMs' potential in organizing agricultural data, aiding in risk assessment, and providing context-aware advice	Demonstrated the architecture's potential in improving project preparation efficiency and compliance with regulations	Planned experiments to determine the most effective architecture for the advisory service prototype in ZEMELA

TABLE V

(Continued.) MAPPING OF RQs TO THE APPLICATION OF MM-LLMs IN AGRICULTURE. THIS TABLE SYSTEMATICALLY CATEGORIZES THE SELECTED LITERATURE BASED ON SEVEN KEY RQs, FOCUSING ON DATA SOURCES AND INTEGRATION, EVALUATION METRICS, PROMPT ENGINEERING TECHNIQUES, OPTIMIZATION STRATEGIES, CHALLENGES AND LIMITATIONS, BENEFITS OVER TRADITIONAL ML AND DL METHODS, AND COMPARATIVE PERFORMANCE OUTCOMES

Paper Reference	**RQ1: Data Sources and Integration**	**RQ2: Evaluation Metrics**	**RQ3: Prompt Engineering Techniques**	**RQ4: Optimization Strategies**	**RQ5: Challenges and Limitations**	**RQ6: Benefits over Traditional ML/DL**	**RQ7: Outperformance over ML/DL**
Comparative Study of Pre-trained Language Models for Text Classification in Smart Agriculture Domain [148]	Conducted a comparative study on transformer models for sentiment analysis in smart agriculture	Evaluated the performance of GPT-2, agricultural BERT, and DistilBERT on text classification in the agricultural domain	Focused on analyzing user perception towards smart farming technologies through sentiment analysis of YouTube data	Aimed to determine the effectiveness of transformer models in classifying agricultural and technical text	Results showed the domain-specific pre-trained agricultural BERT outperformed GPT-2 and DistilBERT in macro-F1-score and accuracy	Demonstrated transformers' viability for text classification in smart agriculture, with agricultural BERT as the superior model	Contributed insights into the application of DTs and ICTs in smart farming through sentiment analysis
NLPof Social Network Data for the Evaluation of Agricultural and Rural Policies [209]	Applied NLP and sentiment analysis on Twitter data to evaluate public opinion on European Common Agricultural Policy (CAP)	Analyzed the evolution of sentiment towards CAP over five years, focusing on key issues like Brexit, environmental measures, and corruption	Investigated the ability of NLP to provide insights into social sentiment on global agricultural policy issues	Demonstrated the potential of sentiment analysis to support inclusive and responsive policymaking in agriculture	Findings revealed that sentiment analysis can enhance agricultural policy design by aligning with public opinion	Highlighted limitations of social media data in representing all agricultural stakeholders, particularly farmers	Suggested future research directions, including the use of diverse NLP techniques and analysis of longer periods
End-to-end Framework for Agricultural Entity Extraction – A Hybrid Model with Transformer [202]	Developed a comprehensive framework for extracting agricultural entities from unstructured text using a hybrid approach	Combined dictionary-based, rule-based, and transformer-based DL techniques to construct an annotated dataset for training	Proposed a BERT-based transformer model with Conditional Random Field (CRF) for improved entity extraction	Achieved high performance metrics with Precision of 88.84%, Recall of 88.9%, and F1-Score of 88.87%	Enhanced entity extraction without relying on pre-labeled datasets, addressing challenges in agricultural NLP tasks	Demonstrated the effectiveness of transfer learning and CRF in improving model performance	Provided a robust solution for agricultural entity extraction, contributing to advancements in NLP applications in agriculture
Advance Research in Agricultural Text-to-Speech: The Word Segmentation of Analytic Language and DL-Based End-to-End System [206]	Discussed the framework and key technologies of Text-to-Speech (TTS) systems in agriculture	Summarized advancements in agricultural TTS, focusing on word segmentation of analytic languages and DL-based end-to-end TTS systems	Proposed solutions to improve training speed, synthesis speed, and real-time deployment on mobile devices	Emphasized challenges such as word segmentation in analytic languages like Chinese and the need for high-quality, fast speech synthesis	Aimed to enhance the intelligibility and naturalness of synthetic speech in agricultural applications	Identified research directions including weakly-supervised learning and optimization for mobile deployment	Highlighted the potential of DL in advancing agricultural TTS technology, addressing specific challenges in the domain

with limited digital literacy among end-users, particularly in rural areas, pose significant barriers to the widespread adoption of LLM-based solutions in agriculture [151]. Recent research has aimed to enhance the capabilities of MM-LLMs in agriculture by tackling core challenges. Studies have focused on semantic understanding in multilingual contexts and genomic data prediction [141], [201]. Additionally, specialized fine-tuning strategies for agricultural term extraction have made significant strides in addressing the limitations of MM-LLMs across diverse agricultural tasks [207]. These efforts are systematically detailed in Table V, which maps the specific RQs to their corresponding challenges and studies, providing a comprehensive overview of current advancements and ongoing research efforts.

Other examples of large-scale, standardized data collection procedures include the Maize Genomes to Fields (G2F) group, which conducts comprehensive evaluations of hybrids and inbreds across multiple environments to understand genotype by environment (GxE) interactions. This initiative has created a large repository of phenotypic, genotypic, environmental, and metadata information publicly available, supporting a more comprehensive understanding of the challenges associated with maize production in various environments [210]. Similarly, the International Maize and Wheat Improvement Center (CIMMYT) conducts extensive maize, wheat, and sorghum trials, which contribute to a vast public dataset aimed at enhancing genotype predictability and stability across different climatic conditions (cimmyt.org/about/data/)

Moreover, the integration of farmer-centric datasets into LLM training is essential to reduce biases and ensure the relevance of model outputs to agricultural applications. Current MM-LLMs often neglect the practices of local farming industries and rich knowledge and experience of farmers, which may lead to predictions and recommendations that may not align with the specific environmental, economic, and cultural contexts of farmers, particularly those in resource-limited settings. The following list outlines critical strategies for integrating farmer-centric data into LLM training and fine-tuning, aimed at reducing biases and enhancing the relevance and applicability of LLM outputs to farmers in specific regions following specific farming practices:

- **Diverse Data Collection:** Gathering data from a wide range of farming contexts, including smallholder farms, family farms, and urban farms in various parts of the

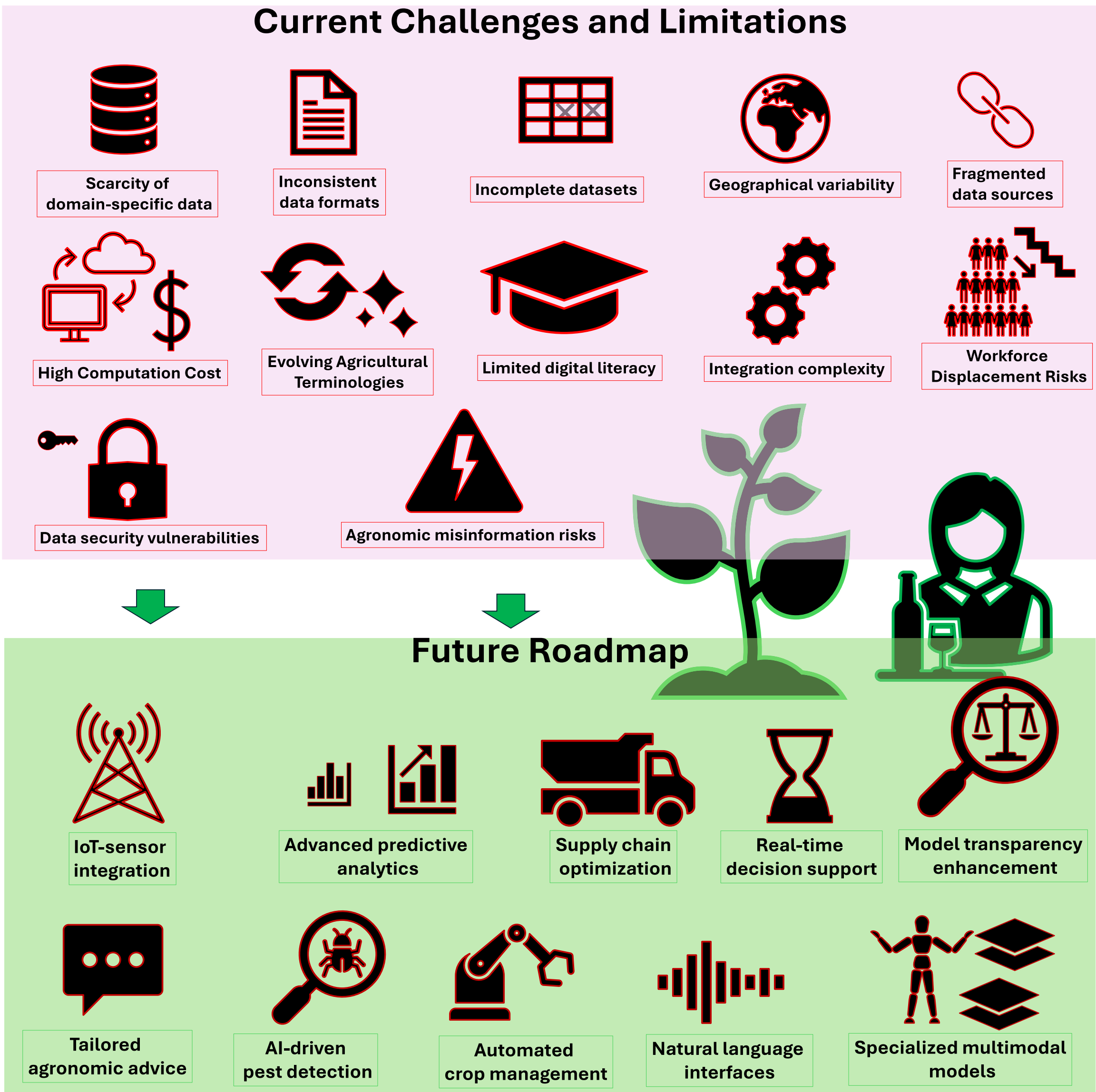

Fig. 8. Summarizing current challenges (top) and future directions (bottom) for MM-LLMs in agriculture. Key limitations include data scarcity, inconsistency, high computational costs, workforce displacement, and security risks. The roadmap emphasizes IoT-sensor synergy, predictive analytics, automated crop management, real-time decision support, multimodal LLM integration, and ethical frameworks to enhance sustainability, transparency, and equitable benefits.

world, to capture the diversity in agricultural practices [33].

- **Local Knowledge Incorporation:** Integrating traditional farming knowledge and practices from specific regions and farming industries into LLM training and finetuning to ensure cultural relevance and practical applicability.
- **Contextual Understanding:** Developing MM-LLMs that can interpret and respond to region-specific agricultural challenges, such as unique pest problems or soil conditions.
- **Adaptive Learning:** Creating mechanisms for MM-LLMs to continuously learn from farmer feedback and local agricultural outcomes, improving their recommendations over time.
- **Multilingual Capabilities:** Enhancing MM-LLMs to process and generate information in local languages and dialects, making them accessible to farmers worldwide.

A. Data Quality Challenge

The efficacy of MM-LLMs in agriculture heavily depends on the quality of data they process. Challenges in this category include:

1) Inconsistent Data Formats: Agricultural data comes from diverse sources like weather stations, satellite imagery, IoT sensors, and manual logs, each using different formats and standards. This heterogeneity in the data formats causes integration and processing challenges, making it difficult for MM-LLMs to synthesize information effectively [151].

2) Incomplete Data: Many agricultural datasets are incomplete, missing crucial information such as crop yields, nutrition status or pest prevalence. This lack of comprehensive data leads to gaps in the models' knowledge, resulting in outputs that might be imprecise, inaccurate or misleading [153].

Efforts to enhance data quality involve standardizing data collection protocols, improving data integration through advanced fusion techniques, and rigorous validation and cleaning to ensure accuracy and consistency. Collaborative efforts from across the agricultural and tech communities are essential to these endeavors [151], [153].

Improving data quality for MM-LLMs in agriculture involves several critical strategies. First, Standardization Efforts focus on creating uniform data collection protocols and formats to ensure consistency and ease data integration. This requires the joint efforts of researchers, farmers, and technology developers. Data Enrichment methods, such as interpolation and fusion, address data gaps and improve completeness by merging information from various sources and including historical or global datasets. Validation and cleaning are vital for ensuring accuracy, utilizing both automated tools and manual reviews to identify and correct errors. Finally, "Crowdsourcing" and "Collaboration" with local communities and experts can provide real-time data and refine data collection processes. These approaches collectively enhance data quality, allowing MM-LLMs to offer more precise and actionable insights for agriculture.

B. Data Availability Challenge

MM-LLMs require extensive, well-curated datasets to function optimally, yet such data are often scarce in the agricultural sector. The main issues include:

1) Geographical Variability: Agricultural data is highly variable, and region- and crop-specific, making it difficult to gather comprehensive datasets that are representative of different agricultural practices and conditions [211].

2) Fragmented Data Sources: Data collection in agriculture is often localized and fragmented, lacking the cohesion necessary for effective MM-LLMstraining. This is compounded by the lack of standardized data collection methods across different regions [26].

Addressing data availability challenges involves enhancing the infrastructure for data collection, such as deploying IoT devices to gather consistent and continuous data streams across diverse agricultural settings [212]. Crowdsourcing and collaboration is vital to address this challenge as well.

C. Other Challenges

Beyond data-related issues, several other challenges impede the adoption and effectiveness of MM-LLMs in agriculture:

1) High Computational Costs: The training and deployment of MM-LLMs require substantial computational resources, which can be a barrier, particularly in developing regions with limited access to such technologies [207].

2) Adaptation to Evolving Agricultural Practices: Agriculture is a rapidly evolving field, and MM-LLMs must continuously adapt to new terms, practices, and conditions to remain relevant and effective [201].

3) Integration Challenges: Incorporating MM-LLMs into existing agricultural systems poses significant technical challenges and requires careful consideration of both technological and human factors [141]. Educating farmers and other stakeholders including demonstration of the economic impact of adopting these technologies is crucial for wider adoption of the technologies in subsistence and commercial farming.

4) Displacement of Agricultural Workforce: As MM-LLMs become more capable of performing complex tasks, such as crop monitoring, disease diagnosis, and precision farming, the demand for human labor in these areas may decrease [33], [147]. This technological advancement may lead to the displacement of the agricultural workforce, particularly in roles traditionally reliant on manual labor and decision-making [28], [150]. This displacement raises important questions about the socio-economic impact on rural communities and the need for strategies to retrain and upskill the agricultural workforce [33], [150]. Addressing these concerns involves balancing the benefits of LLM-driven automation with the potential consequences for employment, ensuring that technological progress does not exacerbate social inequalities.

5) Personal Agronomic Data Security and Proliferation of Agronomic Misinformation: The use of AI solutions like MM-LLMs in agriculture presents notable challenges, particularly in securing personal agronomic data and controlling the spread of misinformation. MM-LLMs handle extensive datasets that can include sensitive information such as crop yields, soil health, and farming methods. It is critical to secure these datasets to prevent unauthorized access and data misuse [213]. Furthermore, MM-LLMs have the potential to generate and disseminate misleading agricultural advice, adversely affecting farming decisions and outcomes [214]. An example of this can be seen in the repercussions of Sri Lanka's transition to organic farming, where misguided policy decisions, potentially exacerbated by erroneous AI outputs, severely impacted agricultural productivity and food security [215]. Ensuring the accuracy and reliability of LLM-based recommendations, coupled with robust data security measures, is crucial to mitigate these risks and safeguard the integrity of agricultural decision-making processes.

VI. CONCLUSION AND FUTURE ROAD-MAP OF MM-LLMs IN AGRICULTURE

It is expected that the future of MM-LLMs in agriculture will be marked by advancements in several critical domains. For instance, integration of IoT and sensor data will enable precise data collection, foundational for advanced predictive analytics aimed at enhancing crop yield forecasts and operational efficiency. Automated crop management systems will revolutionize traditional farming techniques, while tailored agricultural advice and input recommendation will be customized to individual farmer needs reaching to individual plant level. AI-driven technologies will improve pest and disease detection, offering early interventions to minimize the loss while ensuring effective, integrated pest management. Furthermore, optimization of the agricultural supply chain will streamline operations from farm to market. Additionally, real-time decision support systems will facilitate informed decision-making, leveraging comprehensive data analysis facilitated by MM-LLMs.

Moreover, MM-LLMs enhance the control and operation of agricultural machinery through natural language interfaces, making it easier for farmers to manage equipment efficiently. Systems like AutoTractor, powered by LLM technology, allow farmers to command drones and tractors verbally for tasks such as pesticide spraying or field monitoring.

The integration of MM-LLMs into agriculture represents a paradigm shift in the way agricultural processes are managed and optimized. Our comprehensive survey has delved into various applications of MM-LLMs, including crop production, pest and disease detection, demonstrating their profound impact on enhancing efficiency, sustainability, and productivity in the agricultural sector.

One of the key findings of our survey is the ability of MM-LLMs to process and analyze vast amounts of agricultural data. For instance, MM-LLMs have shown remarkable effectiveness in crop management by predicting optimal planting schedules, identifying nutrient deficiencies, and recommending precise irrigation strategies. In pest and disease detection, models like GPT-4 combined with CNNs have significantly improved diagnostic accuracy while reducing computational complexity.

Moreover, the adaptability of MM-LLMs to handle diverse data types, ranging from textual descriptions to sensor data has facilitated the development of more accurate and context-aware decision support systems. These systems not only assist farmers in real, time problem-solving but also enhance the responsiveness of the agricultural supply chain, leading to better management of resources and improved market outcomes.

Despite these advancements, our survey also identifies several challenges that need to be addressed to fully realize the potential of MM-LLMs in agriculture. Data quality and availability remain critical issues, as agricultural data is often sparse, heterogeneous, and context-specific. Ensuring the reliability and accuracy of the data used to train MM-LLMs is paramount. Additionally, the interpretability of MM-LLMs outputs poses a significant challenge, as farmers and other stakeholders need to understand and trust the recommendations provided by these models.

Ethical considerations and sustainability are also crucial aspects that require attention. The deployment of MM-LLMs in agriculture must be guided by principles that ensure fair access, protect privacy, and promote environmental sustainability. Addressing these ethical concerns will be essential to gain widespread acceptance and to harness the full benefits of these technologies.

The future of multimodal MM-LLMs in agriculture is poised to take a transformative path as illustrated in lower part of Figure 8, focusing on the creation of specialized, robust models adept at handling multimodal data and delivering real-time insights, key for their effective integration within agricultural systems [216]. This advancement is anticipated to leverage technologies like IoT sensor integration, advanced predictive analytics, and automated crop management, enhancing efficiency across various farming operations. Key to this evolution is the development of MM-LLMs capable of offering tailored agronomic advice, AI-driven pest detection, and supply chain optimization. These models will benefit from real-time decision support systems and natural language interfaces that simplify complex data interactions, making advanced tools accessible to a broader range of users [217]. Additionally, specialized multimodal models will integrate diverse data types, enhancing the predictive and analytical capabilities of MM-LLMs. Transparency and interpretability also remain crucial, requiring advancements in model transparency enhancement to build trust and facilitate broader adoption in decision-making processes. Moreover, the establishment of comprehensive regulatory frameworks, developed through collaborative efforts among researchers, industry stakeholders, and policymakers, will address ethical and intellectual property concerns, ensuring responsible LLM usage. These frameworks will focus on regulatory framework development and ethical sustainability integration to manage MM-LLMs' impacts on agriculture, aiming to enhance food system resilience and prevent data exploitation. Such strategic foresight and alignment with technological advancements will enable the agricultural sector to leverage MM-LLMs responsibly, supporting innovation while ensuring equitable benefits distribution.

A. Socio-Economic and Cultural Contexts

Incorporating socio-economic and cultural factors into LLM training will enable context-aware solutions, bridging gaps in pest management, supply chain optimization, and resource allocation for diverse farming communities.

B. Automation and Multimodal Systems

Future advancements will focus on embedding MM-LLMs into IoT-enabled robotics and physical systems, automating labor-intensive tasks, and optimizing hybrid human-AI workflows to democratize precision agriculture and reduce manual labor burdens.

References

[1] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 544–551, 2011.

[2] K. Chowdhary and K. Chowdhary, "Natural language processing," in *Fundamentals of Artificial Intelligence*. New Delhi, India: Springer, 2020, pp. 603–649, doi: 10.1007/978-81-322-3972-7_19.

[3] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015.

[4] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," *Artif. Intell. Rev.*, vol. 26, no. 3, pp. 159–190, Nov. 2006.

[5] C. C. Bennett and K. Hauser, "Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach," *Artif. Intell. Med.*, vol. 57, no. 1, pp. 9–19, Jan. 2013.

[6] Y. Xu, S. Zomer, and R. G. Brereton, "Support vector machines: A recent method for classification in chemometrics," *Crit. Rev. Anal. Chem.*, vol. 36, nos. 3–4, pp. 177–188, Dec. 2006.

[7] J. K. Kruschke and T. M. Liddell, "The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective," *Psychonomic Bull. Rev.*, vol. 25, no. 1, pp. 178–206, Feb. 2018.

[8] T. Iqbal and S. Qureshi, "The survey: Text generation models in deep learning," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2515–2528, Jun. 2022.

[9] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[10] R. Dabre, C. Chu, and A. Kunchukuttan, "A survey of multilingual neural machine translation," *ACM Comput. Surveys*, vol. 53, no. 5, pp. 1–38, Sep. 2021.

[11] N. Karanikolas, E. Manga, N. Samaridi, E. Tousidou, and M. Vassilakopoulos, "Large language models versus natural language understanding and generation," in *Proc. 27th Pan-Hellenic Conf. Prog. Comput. Informat.*, Nov. 2023, pp. 278–290.

[12] H. Naveed et al., "A comprehensive overview of large language models," *ACM Trans. Intell. Syst. Technol.*, vol. 16, no. 5, pp. 1–72, 2025.

[13] R. C. Moore and W. Lewis, "Intelligent selection of language model training data," in *Proc. ACL Conf. Short Papers*, 2010, pp. 220–224.

[14] D. Saha et al., "LLM for SoC security: A paradigm shift," *IEEE Access*, vol. 12, pp. 155498–155521, 2024.

[15] M. Cascella, J. Montomoli, V. Bellini, and E. Bignami, "Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios," *J. Med. Syst.*, vol. 47, no. 1, p. 33, Mar. 2023.

[16] J. Wang et al., "Large language models for robotics: Opportunities, challenges, and perspectives," 2024, *arXiv:2401.04334*.

[17] M. Murtaza, C.-T. Cheng, M. Fard, and J. Zeleznikow, "Transforming driver education: A comparative analysis of LLM-augmented training and conventional instruction for autonomous vehicle technologies," *Int. J. Artif. Intell. Educ.*, vol. 35, no. 2, pp. 736–773, Jun. 2025.

[18] I. de Zarzà, J. de Curtò, G. Roig, and C. T. Calafate, "LLM multimodal traffic accident forecasting," *Sensors*, vol. 23, no. 22, p. 9225, Nov. 2023.

[19] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang, "A survey on large language model (LLM) security and privacy: The good, the bad, and the ugly," *High-Confidence Comput.*, vol. 4, no. 2, Jun. 2024, Art. no. 100211.

[20] M. A. K. Raiaan et al., "A review on large language models: Architectures, applications, taxonomies, open issues and challenges," *IEEE Access*, vol. 12, pp. 26839–26874, 2024.

[21] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.

[22] A. Tzachor et al., "Large language models and agricultural extension services," *Nature Food*, vol. 4, no. 11, pp. 941–948, Nov. 2023.

[23] T. B. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 1877–1901.

[24] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutiérrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nature Med.*, vol. 29, no. 8, pp. 1930–1940, 2023.

[25] X. Zhou, X. Zhao, and G. Li, "LLM-enhanced data management," 2024, *arXiv:2402.02643*.

[26] R. Szilágyi and M. Tóth, "Use of LLM for SMEs, opportunities and challenges," *J. Agricult. Informat.*, vol. 14, no. 2, pp. 1–12, Jan. 2024.

[27] S. Stoyanov, M. Kumurdjieva, V. Tabakova-Komsalova, and L. Doukovska, "Using LLMs in cyber-physical systems for agriculture–ZEMELA," in *Proc. Int. Conf. Big Data, Knowl. Control Syst. Eng.*, Nov. 2023, pp. 1–6.

[28] R. Peng, K. Liu, P. Yang, Z. Yuan, and S. Li, "Embedding-based retrieval with LLM for effective agriculture information extracting from unstructured data," 2023, *arXiv:2308.03107*.

[29] H. Zhu, S. Qin, M. Su, C. Lin, A. Li, and J. Gao, "Harnessing large vision and language models in agriculture: A review," 2024, *arXiv:2407.19679*.

[30] X. Wang, H. Kim, S. Rahman, K. Mitra, and Z. Miao, "Human-LLM collaborative annotation through effective verification of LLM labels," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2024, pp. 1–21.

[31] S. S. Kumar, A. K. M. A. Khan, I. A. Banday, M. Gada, and V. V. Shanbhag, "Overcoming LLM challenges using RAG-driven precision in coffee leaf disease remediation," in *Proc. Int. Conf. Emerg. Technol. Comput. Sci. Interdiscipl. Appl.*, Apr. 2024, pp. 1–6.

[32] L. Gong et al., "PreciseSLAM: Robust, real-time, LiDAR–inertial–ultrasonic tightly-coupled SLAM with ultraprecise positioning for plant factories," *IEEE Trans. Ind. Informat.*, vol. 20, no. 6, pp. 8818–8827, Jun. 2024.

[33] D. De Clercq, E. Nehring, H. Mayne, and A. Mahdi, "Large language models can help boost food production, but be mindful of their risks," *Frontiers Artif. Intell.*, vol. 7, Oct. 2024, Art. no. 1326153.

[34] T. Feng, H. Shen, X. Yang, J. Nianga, and Z. Wang, "Integration of large language models with IoT in smart agriculture to improve efficiency, yield, and quality," *Ind. Sci. Eng.*, vol. 1, no. 4, pp. 15–35, Apr. 2024.

[35] A. N. Hamilton, R. S. Mirmahdi, A. Ubeyitogullari, C. K. Romana, J. I. Baum, and K. E. Gibson, "From bytes to bites: Advancing the food industry with three-dimensional food printing," *Comprehensive Rev. Food Sci. Food Saf.*, vol. 23, no. 1, p. 13293, Jan. 2024.

[36] C. Wang, M. Li, and A. J. Smola, "Language models with transformers," 2019, *arXiv:1904.09408*.

[37] B. Min et al., "Recent advances in natural language processing via large pre-trained language models: A survey," *ACM Comput. Surveys*, vol. 56, no. 2, pp. 1–40, Feb. 2024.

[38] A. Pauls and D. Klein, "Faster and smaller n-gram language models," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Human Lang. Technol.*, 2011, pp. 258–267.

[39] J. Zhang, C. Wang, A. Muthu, and V. M. Varatharaju, "Computer multimedia assisted language and literature teaching using heuristic hidden Markov model and statistical language model," *Comput. Electr. Eng.*, vol. 98, Mar. 2022, Art. no. 107715.

[40] R. Rosenfeld, "A whole sentence maximum entropy language model," in *IEEE Workshop Autom. Speech Recognit. Understand. Proc.*, Aug. 1997, pp. 230–237.

[41] K. Yao, B. Peng, G. Zweig, D. Yu, X. Li, and F. Gao, "Recurrent conditional random field for language understanding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 4077–4081.

[42] H.-B. Jeon and S.-Y. Lee, "Language model adaptation based on topic probability of latent Dirichlet allocation," *ETRI J.*, vol. 38, no. 3, pp. 487–493, Mar. 2016.

[43] R. Patil and V. Gudivada, "A review of current trends, techniques, and challenges in large language models (LLMs)," *Appl. Sci.*, vol. 14, no. 5, p. 2074, Mar. 2024.

[44] N. Ding et al., "Parameter-efficient fine-tuning of large-scale pre-trained language models," *Nature Mach. Intell.*, vol. 5, no. 3, pp. 220–235, 2023.

[45] Z. Chen et al., "Exploring the potential of large language models (LLMs) in learning on graphs," *ACM SIGKDD Explorations Newslett.*, vol. 25, no. 2, pp. 42–61, Apr. 2024.

[46] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[47] A. Radford et al., "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[48] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[49] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[50] L. Coheur, "From eliza to siri and beyond," in *Proc. Inf. Process. Manage. Uncertainty Knowledge-Based Systems: 18th Int. Conf.*, Lisbon, Portugal, Jun. 2020, pp. 29–41.

[51] A. W. Lo and M. Singh et al., "From ELIZA to ChatGPT: The evolution of natural language processing and financial applications," *J. Portfolio Manage.*, vol. 49, no. 7, pp. 201–235, Jun. 2023.

[52] V. Rajaraman, "From ELIZA to ChatGPT: History of human–computer conversation," *Resonance*, vol. 28, no. 6, pp. 889–905, Jun. 2023.

[53] P. Kumar, "Large language models (LLMs): Survey, technical frameworks, and future challenges," *Artif. Intell. Rev.*, vol. 57, no. 10, p. 260, Aug. 2024.

[54] V. V. N. S. P. Chandrika, R. Verma, N. Charan, S. Ditheswar, S. Hansika, and R. Ishwariya, "POS tagging using hidden Markov models in natural language processing," in *Proc. Int. Conf. Signal Process., Comput., Electron., Power Telecommun.*, Jul. 2024, pp. 1–6.

[55] W. Feng, J. Sun, L. Zhang, C. Cao, and Q. Yang, "A support vector machine based naive Bayes algorithm for spam filtering," in *Proc. IEEE 35th Int. Perform. Comput. Commun. Conf. (IPCCC)*, Dec. 2016, pp. 1–8.

[56] S. B. S. Ahmad, M. Rafie, and S. M. Ghorabie, "Spam detection on Twitter using a support vector machine and users' features by identifying their interactions," *Multimedia Tools Appl.*, vol. 80, no. 8, pp. 11583–11605, Mar. 2021.

[57] J. Yang et al., "Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond," *ACM Trans. Knowl. Discovery from Data*, vol. 18, no. 6, pp. 1–32, Jul. 2024.

[58] Y. Chang et al., "A survey on evaluation of large language models," *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 3, pp. 1–45, 2024.

[59] A. Han et al., "Enhancing LLM conversational acuity using pragmatic measures," in *Proc. 15th Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2024, pp. 886–891.

[60] R. Yuan et al., "ChatMusician: Understanding and generating music intrinsically with LLM," 2024, *arXiv:2402.16153*.

[61] Q. Deng et al., "ComposerX: Multi-agent symbolic music composition with LLMs," 2024, *arXiv:2404.18081*.

[62] T. Chakrabarty, P. Laban, D. Agarwal, S. Muresan, and C.-S. Wu, "Art or artifice? Large language models and the false promise of creativity," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2024, pp. 1–34.

[63] A. Roush et al., "LLM as an art director (LaDi): Using LLMs to improve text-to-media generators," 2023, *arXiv:2311.03716*.

[64] P. Shailendra, R. C. Ghosh, R. Kumar, and N. Sharma, "Survey of large language models for answering questions across various fields," in *Proc. 10th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2024, pp. 520–527.

[65] S. Lankford, H. Afli, and A. Way, "AdaptMLLM: Fine-tuning multilingual language models on low-resource languages with integrated LLM playgrounds," *Information*, vol. 14, no. 12, p. 638, Nov. 2023.

[66] P. S. Ghatora, S. E. Hosseini, S. Pervez, M. J. Iqbal, and N. Shaukat, "Sentiment analysis of product reviews using machine learning and pre-trained LLM," *Big Data Cognit. Comput.*, vol. 8, no. 12, p. 199, Dec. 2024.

[67] I. Cheong, K. Xia, K. J. K. Feng, Q. Z. Chen, and A. X. Zhang, "(A)I Am not a lawyer, But...: Engaging legal experts towards responsible LLM policies for legal advice," in *Proc. ACM Conf. Fairness Accountability Transparency*, Jun. 2024, pp. 2454–2469.

[68] Z. A. Nazi and W. Peng, "Large language models in healthcare and medical domain: A review," *Informatics*, vol. 11, no. 3, p. 57, Aug. 2024.

[69] M. Shoeybi, M. Patwary, R. Puri, P. LeGresley, J. Casper, and B. Catanzaro, "Megatron-LM: Training multi-billion parameter language models using model parallelism," 2019, *arXiv:1909.08053*.

[70] Y. Zhang et al., "Meta-transformer: A unified framework for multimodal learning," 2023, *arXiv:2307.10802*.

[71] S. Abdullahi, K. U. Danyaro, A. Zakari, I. A. Aziz, N. A. W. A. Zawawi, and S. Adamu, "Time-series large language models: A systematic review of state-of-the-art," *IEEE Access*, vol. 13, pp. 30235–30261, 2025.

[72] W. Liu, J. Chen, H. Wang, Z. Fu, W. J. G. M. Peijnenburg, and H. Hong, "Perspectives on advancing multimodal learning in environmental science and engineering studies," *Environ. Sci. Technol.*, vol. 58, no. 38, pp. 16690–16703, Sep. 2024.

[73] D. Tsirmpas, I. Gkionis, G. T. Papadopoulos, and I. Mademlis, "Neural natural language processing for long texts: A survey on classification and summarization," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108231.

[74] T. A. Shaikh, T. Rasool, K. Veningston, and S. M. Yaseen, "The role of large language models in agriculture: harvesting the future with llm intelligence," *Prog. Artif. Intell.*, vol. 14, pp. 1–48, Dec. 2024.

[75] OpenAI.(2018). *Improving Language Understanding With Unsupervised Learning*. Accessed: Jun. 7, 2024. [Online]. Available: https://openai.com/index/language-unsupervised/

[76] OpenAI.(2019). *Better Language Models and Their Implications*. Accessed: Jun. 7, 2024. [Online]. Available: https://openai.com/index/better-language-models/

[77] J. Achiam et al., "GPT-4 technical report," 2023, *arXiv:2303.08774*.

[78] OPenAI.(2024). *Hello GPT-4o*. Accessed: Mar. 7, 2024. [Online]. Available: https://openai.com/index/hello-gpt-4o/

[79] N. McAleese, R. M. Pokorny, J. F. C. Uribe, E. Nitishinskaya, M. Trebacz, and J. Leike, "LLM critics help catch LLM bugs," 2024, *arXiv:2407.00215*.

[80] D. Guo et al., "DeepSeek-r1: Incentivizing reasoning capability in LLMs via reinforcement learning," 2025, *arXiv:2501.12948*.

[81] H. Fan et al., "HLAT: High-quality large language model pre-trained on AWS trainium," 2024, *arXiv:2404.10630*.

[82] Z. Zhang, A. Zhang, M. Li, H. Zhao, G. Karypis, and A. Smola, "Multimodal chain-of-thought reasoning in language models," 2023, *arXiv:2302.00923*.

[83] S. Soltan et al., "AlexaTM 20B: Few-shot learning using a large-scale multilingual Seq2Seq model," 2022, *arXiv:2208.01448*.

[84] C. Team, "Chameleon: Mixed-modal early-fusion foundation models," 2024, *arXiv:2405.09818*.

[85] (2024). *Introducing Meta Llama 3: The Most Capable Openly Available LLM to Date*. Accessed: Nov. 7, 2024. [Online]. Available: https://ai.meta.com/blog/meta-llama-3/

[86] W. Xiong et al., "Effective long-context scaling of foundation models," 2023, *arXiv:2309.16039*.

[87] C. Zhou et al., "LIMA: Less is more for alignment," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 55006–55021.

[88] S. Iyer et al., "OPT-IML: Scaling language model instruction meta learning through the lens of generalization," 2022, *arXiv:2212.12017*.

[89] J. Xu et al., "Improving open language models by learning from organic interactions," 2023, *arXiv:2306.04707*.

[90] G. Izacard et al., "Atlas: Few-shot learning with retrieval augmented language models," *J. Mach. Learn. Res.*, vol. 24, no. 251, pp. 1–43, 2022.

[91] S. Zhang et al., "OPT: Open pre-trained transformer language models," 2022, *arXiv:2205.01068*.

[92] D. Fried et al., "InCoder: A generative model for code infilling and synthesis," 2022, *arXiv:2204.05999*.

[93] R. Bachmann et al., "4M-21: An any-to-any vision model for tens of tasks and modalities," 2024, *arXiv:2406.09406*.

[94] S. Mehta et al., "OpenELM: An efficient language model family with open training and inference framework," 2024, *arXiv:2404.14619*.

[95] B. McKinzie et al., "MM1: Methods, analysis & insights from multimodal LLM pre-training," 2024, *arXiv:2403.09611*.

[96] J. Ruben Antony Moniz et al., "ReALM: Reference resolution as language modeling," 2024, *arXiv:2403.20329*.

[97] K. You et al., "Ferret-UI: Grounded mobile UI understanding with multimodal LLMs," 2024, *arXiv:2404.05719*.

[98] T.-J. Fu, W. Hu, X. Du, W. Yang Wang, Y. Yang, and Z. Gan, "Guiding instruction-based image editing via multimodal large language models," 2023, *arXiv:2309.17102*.

[99] H. You et al., "Ferret: Refer and ground anything anywhere at any granularity," 2023, *arXiv:2310.07704*.

[100] Nvidia et al., "Nemotron-4 340B technical report," 2024, *arXiv:2406.11704*.

[101] Y. Jiang et al., "VIMA: Robot manipulation with multimodal prompts," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 14975–15022.

[102] B. Wang et al., "InstructRetro: Instruction tuning post retrieval-augmented pretraining," 2023, *arXiv:2310.07713*.

[103] J. Huang, W. Ping, P. Xu, M. Shoeybi, K. Chen-Chuan Chang, and B. Catanzaro, "RAVEN: In-context learning with retrieval-augmented encoder–decoder language models," 2023, *arXiv:2308.07922*.

[104] S. Smith et al., "Using DeepSpeed and megatron to train megatron-turing NLG 530B, a large-scale generative language model," 2022, *arXiv:2201.11990*.

[105] P. Georgiev et al., "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," 2024, *arXiv:2403.05530*.

[106] K. Saab et al., "Capabilities of Gemini models in medicine," 2024, *arXiv:2404.18416*.

[107] S. De et al., "Griffin: Mixing gated linear recurrences with local attention for efficient language models," 2024, *arXiv:2402.19427*.

[108] G. Team et al., "Gemma: Open models based on Gemini research and technology," 2024, *arXiv:2403.08295*.

[109] X. Chen et al., "PaLI-3 vision language models: Smaller, faster, stronger," 2023, *arXiv:2310.09199*.

[110] A. O'Neill et al., "Open X-embodiment: Robotic learning datasets and RT-X models," 2023, *arXiv:2310.08864*.

[111] T. Tu et al., "Towards generalist biomedical AI," *NEJM AI*, vol. 1, no. 3, Feb. 2024, Art. no. 2300138.

[112] *Microsoft Readies New AI Model to Compete With Google, OpenAI, the Information Reports*, Reuter, London, U.K., May 2024.

[113] Y. Sun et al., "You only cache once: Decoder-decoder architectures for language models," 2024, *arXiv:2405.05254*.

[114] M. Abdin et al., "Phi-3 technical report: A highly capable language model locally on your phone," 2024, *arXiv:2404.14219*.

[115] *@wizardlm on Hugging Face: 'Introducing Wizardlm-2! Release Blog'*, H. FACE, Shenzhen, China, Apr. 2015.

[116] Z. Yu et al., "WaveCoder: Widespread and versatile enhancement for code large language models by instruction tuning," 2023, *arXiv:2312.14187*.

[117] A. Mitra et al., "Orca 2: Teaching small language models how to reason," 2023, *arXiv:2311.11045*.

[118] B. Xiao et al., "Florence-2: Advancing a unified representation for a variety of vision tasks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 4818–4829.

[119] J. Bai et al., "Qwen technical report," 2023, *arXiv:2309.16609*.

[120] X.-P. Nguyen et al., "SeaLLMs—Large language models for Southeast Asia," 2023, *arXiv:2312.00738*.

[121] Anthropic.(2024). *Claude 3 Technical Report*. [Online]. Available: https://www.anthropic.com/index/claude-3

[122] (2024). *Introducing Grok-1.5: Enhanced Reasoning and Vision Capabilities*. [Online]. Available: https://x.ai/blog/grok-1.5

[123] Z. Chen, Y. Deng, Y. Wu, Q. Gu, and Y. Li, "Towards understanding the mixture-of-experts layer in deep learning," in *Proc. NIPS*, 2022, pp. 23049–23062.

[124] C. Zhao et al., "Enhancing physical layer communication security through generative AI with mixture of experts," *IEEE Wireless Commun.*, vol. 32, no. 3, pp. 176–184, Jun. 2025.

[125] Y. Xie et al., "MoE-pruner: Pruning mixture-of-experts large language model using the hints from its router," 2024, *arXiv:2410.12013*.

[126] J. Li, G. Li, Y. Li, and Z. Jin, "Structured chain-of-thought prompting for code generation," *ACM Trans. Softw. Eng. Methodol.*, vol. 34, no. 2, pp. 1–23, Feb. 2025.

[127] G. Shi, S. Deng, B. Wang, C. Feng, Y. Zhuang, and X. Wang, "One for all: A unified generative framework for image emotion classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 7057–7068, Aug. 2024.

[128] A. Grattafiori et al., "The llama 3 herd of models," 2024, *arXiv:2407.21783*.

[129] H. Touvron et al., "LLaMA: Open and efficient foundation language models," 2023, *arXiv:2302.13971*.

[130] F. Jiang, "Identifying and mitigating vulnerabilities in llm-integrated applications," M.S. thesis, Dept. Elect. Comput. Eng., Univ. Washington, Seattle, WA, USA, 2024.

[131] A. Q. Jiang et al., "Mixtral of experts," 2024, *arXiv:2401.04088*.

[132] R. Anil et al., "PaLM 2 technical report," 2023, *arXiv:2305.10403*.

[133] Y. Yang et al., "Embodied multi-modal agent trained by an LLM from a parallel TextWorld," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 26275–26285.

[134] J. Li et al., "Foundation models in smart agriculture: Basics, opportunities, and challenges," *Comput. Electron. Agricult.*, vol. 222, Jul. 2024, Art. no. 109032.

[135] S. Tipirneni et al., "Context-aware clustering using large language models," 2024, *arXiv:2405.00988*.

[136] J. Zhang, J. Huang, S. Jin, and S. Lu, "Vision-language models for vision tasks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, pp. 5625–5644, Aug. 2024.

[137] J. Devlin et al., "Language models for image captioning: The quirks and what works," 2015, *arXiv:1505.01809*.

[138] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.

[139] H. Rasheed et al., "GLaMM: Pixel grounding large multimodal model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 13009–13018.

[140] H. Xu et al., "VLM: Task-agnostic video-language model pre-training for video understanding," 2021, *arXiv:2105.09996*.

[141] B. Zhao, W. Jin, J. Del Ser, and G. Yang, "ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification," *Neurocomputing*, vol. 557, Nov. 2023, Art. no. 126708.

[142] S. Rezayi et al., "Exploring new frontiers in agricultural NLP: Investigating the potential of large language models for food applications," 2023, *arXiv:2306.11892*.

[143] J. Qing, X. Deng, Y. Lan, and Z. Li, "GPT-aided diagnosis on agricultural image based on a new light YOLOPC," *Comput. Electron. Agricult.*, vol. 213, Oct. 2023, Art. no. 108168.

[144] B. Silva, L. Nunes, R. Estevão, V. Aski, and R. Chandra, "GPT-4 as an agronomist assistant? Answering agriculture exams using large language models," 2023, *arXiv:2310.06225*.

[145] A. Balaguer et al., "RAG vs fine-tuning: Pipelines, tradeoffs, and a case study on agriculture," 2024, *arXiv:2401.08406*.

[146] X. Yang, J. Gao, W. Xue, and E. Alexandersson, "PLLaMa: An open-source large language model for plant science," 2024, *arXiv:2401.01600*.

[147] Y. Zhang et al., "Building natural language interfaces using natural language understanding and generation: A case study on human–machine interaction in agriculture," *Appl. Sci.*, vol. 12, no. 22, p. 11830, Nov. 2022.

[148] S. Yadav and A. Kaushik, "Comparative study of pre-trained language models for text classification in smart agriculture domain," in *Advances in Data-Driven Computing and Intelligent Systems*, vol. 2. Cham, Switzerland: Springer, 2023, pp. 267–279.

[149] S. Rezayi et al., "AgriBERT: Knowledge-infused agricultural language models for matching food and nutrition," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, Jul. 2022, pp. 5150–5156.

[150] V. G.,V. Kanjirangat, and D. Gupta, "AGRONER: An unsupervised agriculture named entity recognition using weighted distributional semantic model," *Expert Syst. Appl.*, vol. 229, Nov. 2023, Art. no. 120440.

[151] P. P. Ray, "AI-assisted sustainable farming: Harnessing the power of ChatGPT in modern agricultural sciences and technology," *ACS Agricult. Sci. Technol.*, vol. 3, no. 6, pp. 460–462, Jun. 2023.

[152] A. Moustafa, "Smart-insect monitoring system integration and interaction via ai cloud deployment and GPT," Undergraduate Honors thesis, Dept. Comput. Sci., Comput. Eng., 2023. [Online]. Available: https://scholarworks.uark.edu/csceuht/127

[153] B. P. Weichelt, M. Pilz, R. Burke, D. Puthoff, and K. Namkoong, "The potential of AI and ChatGPT in improving agricultural injury and illness surveillance programming and dissemination," *J. Agromedicine*, vol. 29, no. 2, pp. 150–154, Apr. 2024.

[154] Z. Liang, M. Cui, R. Wu, X. Ye, and G. Su, "Harnessing the power of GPT-3 and LSTM for natural language processing in agricultural product news: Focus on soybeans," in *Proc. 4th Int. Conf. Comput. Eng. Intell. Control (ICCEIC)*, Oct. 2023, pp. 307–311.

[155] W. Ting, W. Na, C. Yunpeng, and L. Juan, "Agricultural technology knowledge intelligent question-answering system based on large language model," *Smart Agricult.*, vol. 5, no. 4, p. 105, 2023.

[156] R. Sapkota, Z. Meng, and M. Karkee, "Synthetic meets authentic: Leveraging LLM generated datasets for YOLO11 and YOLOv10-based apple detection through machine vision sensors," *Smart Agricult. Technol.*, vol. 9, Dec. 2024, Art. no. 100614.

[157] J. Wu, Z. Lai, S. Chen, R. Tao, P. Zhao, and N. Hovakimyan, "The new agronomists: Language models are experts in crop management," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 5346–5356.

[158] J. Kpodo, P. Kordjamshidi, and A. P. Nejadhashemi, "AgXQA: A benchmark for advanced agricultural extension question answering," *Comput. Electron. Agricult.*, vol. 225, Oct. 2024, Art. no. 109349.

[159] C. Dofitas, Y.-W. Kim, and Y.-C. Byun, "Advanced agricultural query resolution using ensemble-based large language models," *IEEE Access*, vol. 13, pp. 34732–34746, 2025.

[160] R. Calone et al., "Analysing the potential of ChatGPT to support plant disease risk forecasting systems," *Smart Agricult. Technol.*, vol. 10, Mar. 2025, Art. no. 100824.

[161] O. Bugaiova and K. N. J. Hansen, "A dataset dedicated to the training of large- language models for agronomic management practices and production in Norwegian agriculture," *Data Brief*, vol. 59, Apr. 2025, Art. no. 111326.

[162] D. Chen and Y. Huang, "Integrating reinforcement learning and large language models for crop production process management optimization and control through a new knowledge-based deep learning paradigm," *Comput. Electron. Agricult.*, vol. 232, May 2025, Art. no. 110028.

[163] N. Zhang et al., "Large language models for explainable decisions in dynamic digital twins," 2024, *arXiv:2405.14411*.

[164] S. Jiang, R. Angarita, S. Cormier, and F. Rousseaux, "Fine-tuning BERT-based models for plant health bulletin classification," 2021, *arXiv:2102.00838*.

[165] R. Sapkota, A. Paudel, and M. Karkee, "Zero-shot automatic annotation and instance segmentation using LLM-generated datasets: Eliminating field imaging and manual annotation for deep learning model development," 2024, *arXiv:2411.11285*.

[166] J. Wu, D. Pichler, D. Marley, D. Wilson, N. Hovakimyan, and J. Hobbs, "Extended agriculture-vision: An extension of a large aerial image dataset for agricultural pattern analysis," 2023, *arXiv:2303.02460*.

[167] B. Deforce, B. Baesens, and E. S. Asensio, "Time-series foundation models for forecasting soil moisture levels in smart agriculture," 2024, *arXiv:2405.18913*.

[168] Y. Wu, X. Hu, Z. Fu, S. Zhou, and J. Li, "GPT-4o: Visual perception performance of multimodal large language models in piglet activity understanding," 2024, *arXiv:2406.09781*.

[169] N. Darapaneni, S. Raj, V. Raghul, V. Sivaraman, S. Mohan, and A. R. Paduri, "LSTM-RASA based agri farm assistant for farmers," 2022, *arXiv:2204.09717*.

[170] J. Roberts, T. Lüddecke, R. Sheikh, K. Han, and S. Albanie, "Charting new territories: Exploring the geographic and geospatial capabilities of multimodal LLMs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2024, pp. 554–563.

[171] K. Pan, X. Zhang, and L. Chen, "Research on the training and application methods of a lightweight agricultural domain-specific large language model supporting Mandarin Chinese and Uyghur," *Appl. Sci.*, vol. 14, no. 13, p. 5764, Jul. 2024.

[172] M. T. Kuska, M. Wahabzada, and S. Paulus, "AI for crop production–where can large language models (LLMs) provide substantial value?," *Comput. Electron. Agricult.*, vol. 221, Jun. 2024, Art. no. 108924.

[173] V. Nguyen, S. Karimi, W. Hallgren, A. Harkin, and M. Prakash, "My climate advisor: An application of NLP in climate adaptation for agriculture," in *Proc. 1st Workshop Natural Lang. Process. Meets Climate Change*, 2024, pp. 27–45.

[174] Y. Lu et al., "Application of multimodal transformer model in intelligent agricultural disease detection and question-answering systems," *Plants*, vol. 13, no. 7, p. 972, Mar. 2024.

[175] Y. Hue et al., "Artificial intelligence plant doctor: Plant disease diagnosis using GPT4-vision," *Res. Plant Disease*, vol. 30, no. 1, pp. 99–102, Mar. 2024.

[176] V. Madaan et al., "Integrating language models and machine learning for crop disease detection for farmer guidance," in *Proc. Workshop Adv. Comput. Intell.*, 2024, pp. 1–16. [Online]. Available: https://ceur-ws.org/Vol-3706/Paper11.pdf

[177] S. Yang, Z. Yuan, S. Li, R. Peng, K. Liu, and P. Yang, "GPT-4 as evaluator: Evaluating large language models on pest management in agriculture," 2024, *arXiv:2403.11858*.

[178] W. Huang, P. Abbeel, D. Pathak, and I. Mordatch, "Language models as zero-shot planners: Extracting actionable knowledge for embodied agents," 2022, *arXiv:2201.07207*.

[179] S. S. Raman, V. Cohen, E. Rosen, I. Idrees, D. Paulius, and S. Tellex, "Planning with large language models via corrective re-prompting," in *Proc. NeurIPS Found. Models Decis. Making Workshop*, 2022, pp. 1–21.

[180] W. Huang et al., "Inner monologue: Embodied reasoning through planning with language models," 2022, *arXiv:2207.05608*.

[181] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," 2019, *arXiv:1910.01108*.

[182] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," 2023, *arXiv:2304.03442*.

[183] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," 2022, *arXiv:2201.11903*.

[184] (2024). *KissanAI*. Accessed: Jul. 15, 2024. [Online]. Available: https://kissan.ai/

[185] T. Wu, K. Liu, M. Cheng, Z. Gu, W. Guo, and X. Jiao, "Paddy field scale evapotranspiration estimation based on two-source energy balance model with energy flux constraints and UAV multimodal data," *Remote Sens.*, vol. 17, no. 10, p. 1662, May 2025.

[186] T. L. Marxhall, A. D. Hagerman, H. E. Shear, K. H. Burdine, and B. B. Jablonski, "Building up the next generation of extension specialists," *Appl. Econ. Teaching Resour.*, vol. 4, no. 3, pp. 1–11, 2022.

[187] P. Raina, "Assessing the attitude of farmers towards artificial intelligence in agriculture," *Asian J. Agricult. Extension*, vol. 42, no. 7, pp. 59–63, 2024.

[188] D. K. Jayasingh, A. Anand, and K. S. Das, "Artificial intelligence in agricultural extension," Univ. Agricult. Technol., Bhubaneswar, India, Tech. Rep., 2024, p. 115.

[189] A. Ibrahim, K. Senthilkumar, and K. Saito, "Evaluating responses by ChatGPT to farmers' questions on irrigated lowland Rice cultivation in Nigeria," *Sci. Rep.*, vol. 14, no. 1, p. 3407, Feb. 2024.

[190] P. Hill and L. Narine, "Ensuring responsible and transparent use of generative AI in extension," *J. Extension*, vol. 61, no. 2, p. 13, Sep. 2023.

[191] T. A. Shaikh, T. Rasool, and F. R. Lone, "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming," *Comput. Electron. Agricult.*, vol. 198, Jul. 2022, Art. no. 107119.

[192] P. Richter, "World agri-tech 2024: Pioneering agriculture resilience with AI," Microsoft Cloud Retail Agricult. Solutions, Redmond, WA, USA, Tech. Rep., Apr. 2024. [Online]. Available: https://www.microsoft.com/en-us/industry/blog/sustainability/2024/04/02/worldagri-tech-2024-pioneering-agriculture-resilience-with-ai/

[193] R. Courtney and T. Mullinax, "Digital dialogues offer orcharding advice," Good Fruit Grower, Yakima, WA, USA, Tech. Rep., Sep. 2023. [Online]. Available: https://goodfruit.com/digital-dialogues-offer-orcharding-advice/

[194] S. Thomas, "Agvisorpro launches first large language model for equipment dealers," Precision Farming Dealer, Brookfield, WI, USA, Tech. Rep., Jul. 2023.

[195] *Cropin Launches Aks.ara: A Micro Language Model (μ-LM) for Climate Smart Agriculture, Targeting Underserved Farmers in the Global Sout*, Cropin Technology, Karnataka, India, Apr. 2024.

[196] *Cropin Partners With Google's Gemini to Launch Agri Intelligence Tool*, The Economic Times, Mumbai, India, Aug. 2024.

[197] (2024). *Harvest Enginer*. Accessed: Aug. 18, 2024. [Online]. Available: https://www.kragworks.com/

[198] X. Zhao et al., "Implementation of large language models and agricultural knowledge graphs for efficient plant disease detection," *Agriculture*, vol. 14, no. 8, p. 1359, Aug. 2024.

[199] Y. Fei, J. Fan, and G. Zhou, "Extracting fruit disease knowledge from research papers based on large language models and prompt engineering," *Appl. Sci.*, vol. 15, no. 2, p. 628, Jan. 2025.

[200] R. Palma et al., "Agricultural information model," in *Information and Communication Technologies for Agriculture—Theme III: Decision*. Cham, Switzerland: Springer, 2022, pp. 3–36.

[201] J. Mendoza-Revilla et al., "A foundational large language model for edible plant genomes," *Commun. Biol.*, vol. 7, no. 1, p. 835, Jul. 2024.

[202] E. N. Mol and M. S. Kumar, "End-to-end framework for agricultural entity extraction—A hybrid model with transformer," *Comput. Electron. Agricult.*, vol. 225, Apr. 2024, Art. no. 109309.

[203] A. Chebbi, G. Kniesel, N. Abdennadher, and G. Dimarzo, "Enhancing named entity recognition for agricultural commodity monitoring with large language models," in *Proc. 4th Workshop Mach. Learn. Syst.*, Apr. 2024, pp. 208–213.

[204] J. M. Müller et al., "Investigating the use of large language models in agricultural injury surveillance," in *Proc. Annu. Int. Meeting*, 2024, pp. 1–16.

[205] J. M. Shutske, "Editorial: Harnessing the power of large language models in agricultural safety & health," *J. Agricult. Saf. Health*, vol. 29, no. 4, pp. 205–224, 2023.

[206] X. Li, D. Ma, and B. Yin, "Advance research in agricultural text-to-speech: The word segmentation of analytic language and the deep learning-based end-to-end system," *Comput. Electron. Agricult.*, vol. 180, Jan. 2021, Art. no. 105908.

[207] H. Panoutsopoulos, B. Espejo-Garcia, S. Raaijmakers, X. Wang, S. Fountas, and C. Brewster, "Investigating the effect of different fine-tuning configuration scenarios on agricultural term extraction using BERT," *Comput. Electron. Agricult.*, vol. 225, Oct. 2024, Art. no. 109268.

[208] G. Hegde, V. R. Hulipalled, and J. B. Simha, "Price prediction of agriculture commodities using machine learning and NLP," in *Proc. 2nd Int. Conf. Smart Technol. Comput., Electr. Electron.*, Dec. 2021, pp. 1–6.

[209] A. G. Domínguez, N. Roig-Tierno, N. Chaparro-Banegas, and J.-M. García-Álvarez-Coque, "Natural language processing of social network data for the evaluation of agricultural and rural policies," *J. Rural Stud.*, vol. 109, Jul. 2024, Art. no. 103341.

[210] D. C. Lima et al., "2020–2021 field seasons of maize GxE project within the genomes to fields initiative," *BMC Res. Notes*, vol. 16, no. 1, p. 219, Sep. 2023.

[211] M. Ali et al., "Tokenizer choice for LLM training: Negligible or crucial?," 2023, *arXiv:2310.08754*.

[212] Y. Sun and J. Ortiz, "An AI-based system utilizing IoT-enabled ambient sensors and LLMs for complex activity tracking," 2024, *arXiv:2407.02606*.

[213] U. Jayasinghe and D. Ghoshal, "Fertiliser ban decimates Sri Lankan crops as government popularity ebbs," London, U.K., Tech. Rep., 2022, vol. 3. [Online]. Available: https://www.reuters.com/markets/commodities/fertiliser-ban-decimates-sri-lankan-crops-government-popularity-ebbs-2022-03-03/

[214] X. Tang et al., "Risks of AI scientists: Prioritizing safeguarding over autonomy," 2024, *arXiv:2402.04247*.

[215] B. Chandra Das, M. Hadi Amini, and Y. Wu, "Security and privacy challenges of large language models: A survey," 2024, *arXiv:2402.00888*.

[216] Z. Yu, Y. Wu, N. Zhang, C. Wang, Y. Vorobeychik, and C. Xiao, "Codeipprompt: Intellectual property infringement assessment of code language models," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 40373–40389.

[217] A. George and T. Walsh, "Artificial intelligence is breaking patent law," *Nature*, vol. 605, no. 7911, pp. 616–618, May 2022.

Ranjan Sapkota (Member, IEEE) received the M.S. degree in agricultural and biosystems engineering from North Dakota State University, USA, in 2022, where he specialized in computer vision, GIS, remote sensing, agricultural machinery, and UAV applications for agriculture. He is currently pursuing the Ph.D. degree with the Department of Biological and Environmental Engineering, Cornell University. From 2022 to 2024, he was with Washington State University, advancing research in agricultural automation and intelligent robotic systems. His research interests include artificial intelligence and robotics in agriculture, with expertise spanning automation systems, machine vision, robot manipulation, multimodal large language models (MM-LLMs), deep learning, agentic AI, and generative AI technologies.

Rizwan Qureshi (Senior Member, IEEE) received the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2021. He was an Assistant Professor with the Fast School of Computing, National University of Computer and Emerging Sciences, Karachi, Pakistan. He was also a Post-Doctoral Researcher with the College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar, and The University of Texas MD Anderson Cancer Center, Houston, TX, USA, and the AI-Team Lead of the Center for Regenerative Medicine and Health, Hong Kong Institute of Science and Innovation, Chinese Academy of Sciences, Hong Kong. He is currently with the Center for Research in Computer Vision (CRCV), University of Central Florida, Orlando, USA. His research interests include AI applications in life sciences, cancer data sciences, computer vision, large language models, and machine learning. He was also shortlisted for the USERN International Prize in 2024.

Muhammad Usman Hadi (Member, IEEE) received the M.S. and Ph.D. degrees from the University of Bologna, Italy. He was a Post-Doctoral Researcher with Aalborg University, Denmark, in close collaboration with Nokia. He was a Visiting Researcher with ESIEE Paris, France, and Nokia Bell Laboratories. He is currently an Assistant Professor with the School of Engineering, Ulster University, Belfast, U.K. He has authored more than 60 journal articles and transactions. He has received the EPSRC DTNET+, Innovate U.K., British Council, Research and Impact fund, Proof of Concept RI, and Garfield Weston Trust fundings, as a Principal Investigator, and has been the Co-Investigator for EPSRC, DFE, and ISPF research grants. His research interests include machine learning for engineering applications, MM-LLMs, autonomous aerial systems, fiber wireless communication, microwave photonics, and devices for telecommunications. He has been included in the top 2% of researchers for three consecutive years since 2021.

Syed Zohaib Hassan received the B.S. degree in civil engineering from NED University, Karachi, Pakistan, in 2018, and the M.S. degree in construction automation from Yonsei University, Seoul, South Korea, in 2021. He is currently pursuing the Ph.D. degree with the Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, USA. His research interests include remote sensing, computer vision, and AI for applications in environmental monitoring, smart cities, and precision agriculture.

Ferhat Sadak received the M.Sc. degree in advanced mechanical engineering and the Ph.D. degree in medical robotics from the University of Birmingham in 2016 and 2021, respectively. He was a Post-Doctoral Researcher with Sorbonne University, Paris. He is currently an Assistant Professor with Bartin University, Türkiye. His research interests include image processing, machine learning, and vision-guided for robotic automatization.

Maged Shoman received the Ph.D. degree in transportation engineering from the University of Missouri, Columbia, USA, in 2022, focusing on computer vision and machine learning. He is currently a Post-Doctoral Researcher with UCF. Prior to joining UCF, he was a Visiting Scholar with The University of British Columbia (UBC). He has authored over 15 peer-reviewed publications in several prestigious journals and has presented at multiple conferences. His papers have been published in top venues, including IEEE CVPR, the ASCE *Journal of Transportation Engineering*, and the *Transportation Research Record*. His research interests include computer vision, transportation sensing, big data analytics, and deep learning. He received numerous awards and distinctions, including multiple placements in the NVIDIA AI City Challenge and the Outstanding Ph.D. Student Award. He has also been actively involved in academic and professional organizations, holding leadership roles at the Institute of Transportation Engineers and American Society of Civil Engineers.

Muhammad Sajjad (Member, IEEE) works on the Frontiers of interdisciplinary climate research (integrated physical science, social aspects, and technology) to explore the risk-resilience-sustainability nexus using geospatial and remote sensing data science approaches to provide practical solutions. His research interests include environmental sustainability, urban planning/design, public health, geo-AI, machine learning, location intelligence, and interdisciplinary climate science.

Fayaz Ali Dharejo (Senior Member, IEEE) received the B.E. degree in electronic engineering from QUEST, Pakistan, the M.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, and the Ph.D. degree in computer applied engineering from the Computer Network Information Center, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China. He is currently a Post-Doctoral Fellow with the University of Würzburg, Germany. Previously, he was a Post-Doctoral Fellow with Khalifa University, United Arab Emirates. He has published more than 50 articles in reputed ISI impact factor journals, including IEEE TRANSACTIONS ON FUZZY SYSTEMS, ACM TIST, *International Journal of Intelligent Systems*, IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. His research interests include computer vision, image enhancement, lightweight models, and super-resolution. He is a member of professional bodies, such as PEC, ACM, IEEE, and IEEE Societies, such as the IEEE Geoscience and Remote Sensing Society, the IEEE Computer Society, and the IEEE Signal Processing Society.

Achyut Paudel received the B.S. degree in mechanical engineering from Tribhuvan University, Nepal, and the M.S. degree in mechanical engineering and the Ph.D. degree in agricultural engineering from Washington State University. He is currently a Machine Learning/Computer Vision Research Engineer with Orchard Robotics, Seattle, WA, USA, where he focuses on automating various tasks in orchards and vineyards. His research leverages machine vision and artificial intelligence to enhance agricultural practices.

Jiajia Li (Graduate Student Member, IEEE) received the bachelor's degree from the University of Electronic Science and Technology of China, and the master's degree from Michigan State University, MI, USA (MSU), where she is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. Her research interests include smart agriculture, remote sensing, and computer vision.

Zhichao Meng received the bachelor's degree in mechanical design, manufacturing, and automation from Huzhou University, China, and the Ph.D. degree in mechanical engineering from Zhejiang Sci-Tech University, China, in 2025. He is currently a Researcher with the College of Optical and Mechanical Engineering, Zhejiang A&F University, Hangzhou, China. His research interests include agricultural robotics, machine vision, and robotic arm control.

John Shutske is currently a Professor and an Extension Specialist in biological systems engineering with the University of Wisconsin–Madison, focusing on safety engineering and technology-driven risk control in agriculture. He also studies the adoption and barriers of digital tools and connectivity among agricultural producers and rural service providers.

Manoj Karkee received the Ph.D. degree in agricultural engineering and human–computer interaction from Iowa State University. He is currently the Norman R. and Sharon R. Scott Professor of Agriculture and Life Sciences with the Department of Biological and Environmental Engineering, Cornell University. He has been the Director and a Professor with the Center for Precision and Automated Agricultural Systems, Washington State University. He leads a strong research and education program in the area of sensing, machine vision, AI, and robotics in agriculture. He has published widely in such journals as *Information Fusion*, *Artificial Intelligence Review*, *Computers and Electronics in Agriculture*, *Computers in Industry*, *Journal of Field Robotics*, and *Journal of the American Society of Agricultural and Biological Engineers* (ASABE), and has been an invited speaker at numerous national and international conferences and universities. He was awarded the 2020 Rainbird Engineering Concept of the Year by ASABE and was recognized as the 2019 Pioneer in Artificial Intelligence and IoT by Connected World Magazine. He is also the Elected Chair of CIGR (International Commission of Agricultural and Biosystems Engineering) Section III –Plant Production, and IFAC (International Federation of Automatic Control) Technical Committee 8.1—Control in Agriculture. He is also serving as the Editor-in-Chief for *Computers and Electronics in Agriculture* and an Associate Editor for *Journal of the ASABE*, and served as a Guest Editor for *Journal of Field Robotics*.