



OPEN ACCESS

EDITED BY

Lei Shu,
Nanjing Agricultural University, China

REVIEWED BY

Orly Enrique Apolo-Apolo,
KU Leuven, Belgium
Zhanhao Shi,
Shandong Agriculture and Engineering
University, China

*CORRESPONDENCE

Hongyan Zhu

✉ hyzhu-zju@foxmail.com

Junfeng Gao

✉ junfeng.gao@abdn.ac.uk

†These authors have contributed equally to this work and share first authorship

RECEIVED 19 February 2025

ACCEPTED 28 July 2025

PUBLISHED 02 September 2025

CITATION

Zhu H, Qin S, Su M, Lin C, Li A and Gao J
(2025) Harnessing large vision and
language models in agriculture:
a review.
Front. Plant Sci. 16:1579355.
doi: 10.3389/fpls.2025.1579355

COPYRIGHT

© 2025 Zhu, Qin, Su, Lin, Li and Gao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Harnessing large vision and language models in agriculture: a review

Hongyan Zhu^{1,2*†}, Shuai Qin^{1,2†}, Min Su^{1,2}, Chengzhi Lin^{1,2},
Anjie Li^{1,2} and Junfeng Gao^{3*}

¹Guangxi Key Laboratory of Brain-inspired Computing and Intelligent Chips, School of Electronic and Information Engineering, Guangxi Normal University, Guilin, China, ²Key Laboratory of Integrated Circuits and Microsystems (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region, Guilin, China, ³Department of Computer Science, University of Aberdeen, Aberdeen, United Kingdom

Introduction: Agriculture is a cornerstone of human society but faces significant challenges, including pests, diseases, and the need for increased production efficiency. Large models, encompassing large language models, large vision models, and multimodal large language models, have shown transformative potential in various domains. This review aims to explore the potential applications of these models in agriculture to address existing problems and improve production.

Methods: We conduct a systematic review of the development trajectories and key capabilities of large models. A bibliometric analysis of literature from Web of Science and arXiv is performed to quantify the current research focus and identify the gap between the potential and the application of large models in the agricultural sector.

Results: Our analysis confirms that agriculture is an emerging but currently underrepresented field for large model research. Nevertheless, we identify and categorize promising applications, including tailored models for agricultural question-answering, robotic automation, and advanced image analysis from remote sensing and spectral data. These applications demonstrate significant potential to solve complex, nuanced agricultural tasks.

Discussion: This review culminates in a pragmatic framework to guide the choice between large and traditional models, balancing data availability against deployment constraints. We also highlight critical challenges, including data acquisition, infrastructure barriers, and the significant ethical considerations for responsible deployment. We conclude that while tailored large models are poised to greatly enhance agricultural efficiency and yield, realizing this future requires a concerted effort to overcome the existing technical, infrastructural, and ethical hurdles.

KEYWORDS

large model, agriculture, natural language processing, computer vision, multimodal model

1 Introduction

The significance of agriculture in the global economy is increasing steadily, and there is growing awareness regarding its sustainability. [Ahirwar et al. \(2019\)](#) believe that it is necessary to increase global agricultural food production by a minimum of 70% to meet the needs of the increasing world population. Unfortunately, there are many factors in agriculture that make it difficult to steadily increase grain production, including 1): crop diseases caused by pathogens such as bacteria, fungi, and viruses. These diseases can spread rapidly, often leading to devastating effects on entire crops. For instance, bacterial blight in rice and late blight in potatoes can wipe out significant portions of harvests. The economic impact is staggering, as farmers face not only reduced yields but also increased costs associated with disease management; 2): poor seed quality can lead to weak plant growth, reduced yields, and greater susceptibility to both diseases and pests. Farmers who use low-quality seeds often experience crop failures, which not only jeopardizes their income but also contributes to broader food insecurity within communities. Transitioning to certified, high-quality seeds is essential for improving crop resilience and productivity; 3): many agricultural tasks remain inefficient and labor-intensive, hindering productivity. Traditional methods of weeding, planting, watering, and harvesting are often time-consuming and can lead to resource wastage. For example, manual weeding not only consumes labor but may also fail to effectively control weed populations, resulting in reduced crop yields. The adoption of mechanization and modern farming techniques, such as precision agriculture (PA), can significantly improve efficiency.

PA is an agricultural management approach that utilizes modern technology to enhance production efficiency and sustainability. It encompasses sensor technology, unmanned aerial/ground vehicles (UAVs/UGVs), remote sensing technology, automation equipment, big data, machine learning (ML), and deep learning (DL) ([Tokekar et al., 2016](#); [Khanal et al., 2020](#); [Saleem et al., 2021](#)). This enables farmers to reduce production costs and improve decision-making capabilities, providing significant economic and social benefits. For crop diseases, traditional detection methods like polymerase chain reactions based on unique deoxyribonucleic acid sequences of pathogens, enzyme-linked immunosorbent assays on the basis of pathogens proteins and hyperspectral imaging, are constrained by their operational complexity and the requirement for bulky instruments ([Yao et al., 2024](#)). For selecting high-quality seeds, quality assurance programs employ various ways to attest seed quality attributes, including germination and vigor tests ([ElMasry et al., 2019](#)). But these methods have limitations in terms of time overhead, subjectivity, and the destructive nature of assessing seed quality ([Medeiros et al., 2020](#)). For a general tasks in agriculture, the use of pesticides for weed control may have negative impacts on the environment, and Phytotoxicity reactions can lead to diminished crop quality and reduced yields ([Visentin et al., 2023](#)). And the traditional solutions to these tasks are also inefficient due to these manned implements are dreadfully slow. On the other hand, driven by growing health consciousness, the public has long

been worried about the safety and quality of food, which is linked to agricultural products. Reducing food losses and improving food safety rely significantly on the continuous monitoring of crop quality, especially the inspection of diseases during crop growth stage ([Karthikeyan et al., 2020](#)).

DL technologies in PA can effectively address the limitations of traditional methods by leveraging their powerful data processing and pattern recognition capabilities. For instance, DL can analyze vast amounts of data from sensors, drones, and satellite imagery to accurately identify crop health, soil characteristics, and potential diseases and pests ([Nasir et al., 2021](#); [Bouguettaya et al., 2022](#)). This application enables farmers to obtain real-time insights, allowing for more scientifically informed management decisions, optimized resource use, and increased crop yields. However, DL technologies also have their limitations, primarily due to the high demand for model training ([Sun et al., 2017](#)). DL models typically require large amounts of labelled data to train and often need to be retrained when faced with new agricultural environments or crop varieties ([Thenmozhi and Reddy, 2019](#)). This repetitive training process is not only time-consuming but also requires significant computational resources and expertise. The effectiveness of transfer learning lies in its ability to apply models trained in one domain to a related domain, thus reducing the need for new datasets ([Bosilj et al., 2020](#); [Paymode and Malode, 2022](#)). However, the diversity and complexity of agricultural environments can limit the effectiveness of transfer learning ([Raffel et al., 2020](#)). For example, differences in soil conditions, climate variations, and crop growth characteristics across regions can result in models trained in one area performing poorly in another. Therefore, although DL holds tremendous potential in PA, its adaptability and generalizability must be carefully considered to ensure that models remain effective in the ever-evolving agricultural field.

Large models are fundamentally distinguished from conventional DL models by their vast parameter counts (often billions) and extensive pre-training on massive, diverse datasets. By being exposed to a rich array of information, these models can better understand and adapt to various contexts, making them highly versatile tools in fields such as natural language processing (NLP), computer vision (CV), and decision-making ([Kung et al., 2023](#)). Crucially, unlike traditional DL models, large models develop “emergent abilities”—such as few-shot/zero-shot learning, complex reasoning, and strong generative abilities—that are not simply scaled-up versions of prior performance ([Bommasani et al., 2021](#); [Zhao et al., 2023c](#)). As an efficient analytical means, large model, has found extensive application in the agricultural sector ([Stella et al., 2023](#); [Yang et al., 2023c](#)). They have demonstrated excellent performance in analyzing agricultural data, pest and disease management, PA, and more. However, they still face many problems such as difficulty in obtaining agricultural data ([Lu and Young, 2020](#)), low model training efficiency, distribution shift ([Chawla et al., 2021](#)), and plant blindness ([Geitmann and Bidhendi, 2023](#)). In response to the challenges faced by traditional agriculture, we committed to conducting a comprehensive analysis of large models. First, we systematically summarized the history of large models, their applications in other fields, and their significance for agriculture. Subsequently, we introduced many applications of

large models in agriculture. Furthermore, recognizing that large models were a relatively new technological approach, we outlined some solutions from ethical and responsibility perspectives. Finally, we summarized the current challenges and future directions of large models and drew conclusions on their effectiveness in the agricultural domain.

2 Feasibility analysis of large models in agriculture

Artificial intelligence (AI), whose main purpose is to establish systems that learn and think like human (Holzinger et al., 2019), just like human language and visual abilities. At present, research on large models is also focused on NLP and CV. Next, large language model (LLM), large vision model (LVM) and multimodal large language model (MLLM) will be introduced in detail.

2.1 Evolution and key milestones of large models

2.1.1 Development trajectories of large language models

LLM is a model based on NLP with a vast number of parameters (typically billions) trained on massive datasets of text and code, and we can divide the development of it into four stages (Figure 1):

1. Statistical Language Models (SLM): SLMs use traditional statistical methods (like n-grams) to learn word probabilities. Their effectiveness relies on the amount of data and estimation algorithms (Chelba et al., 2013). While SLMs are widely used in NLP, they have three main drawbacks: Scalability: Larger n requires more memory and parameters (n represents how many preceding words the model considers when predicting the next word); Information sharing: N-grams can't share semantic information across similar words; Data sparsity: Techniques like data smoothing can help, but neural networks handle this better.
2. Neural Language Models (NLM): NLMs utilize various neural networks and are more effective than SLMs (Bengio et al., 2000; Mikolov et al., 2010; Sundermeyer et al., 2012). They address data sparsity using feedforward and recurrent neural networks (RNNs), which learn features automatically. Key developments include: Feedforward neural networks (FFNNLM): Proposed by Bengio et al. in 2003, they learn distributed word representations (Bengio et al., 2000); RNN Language Model (RNNLM): Introduced by Mikolov et al., but struggles with long-term dependencies (Mikolov et al., 2010). Long short-term memory (LSTM) networks were later added to overcome this (Sundermeyer et al., 2012).
3. Pre-trained Language Models (PLM): PLMs are categorized into feature-based and fine-tuning methods: Feature-based:

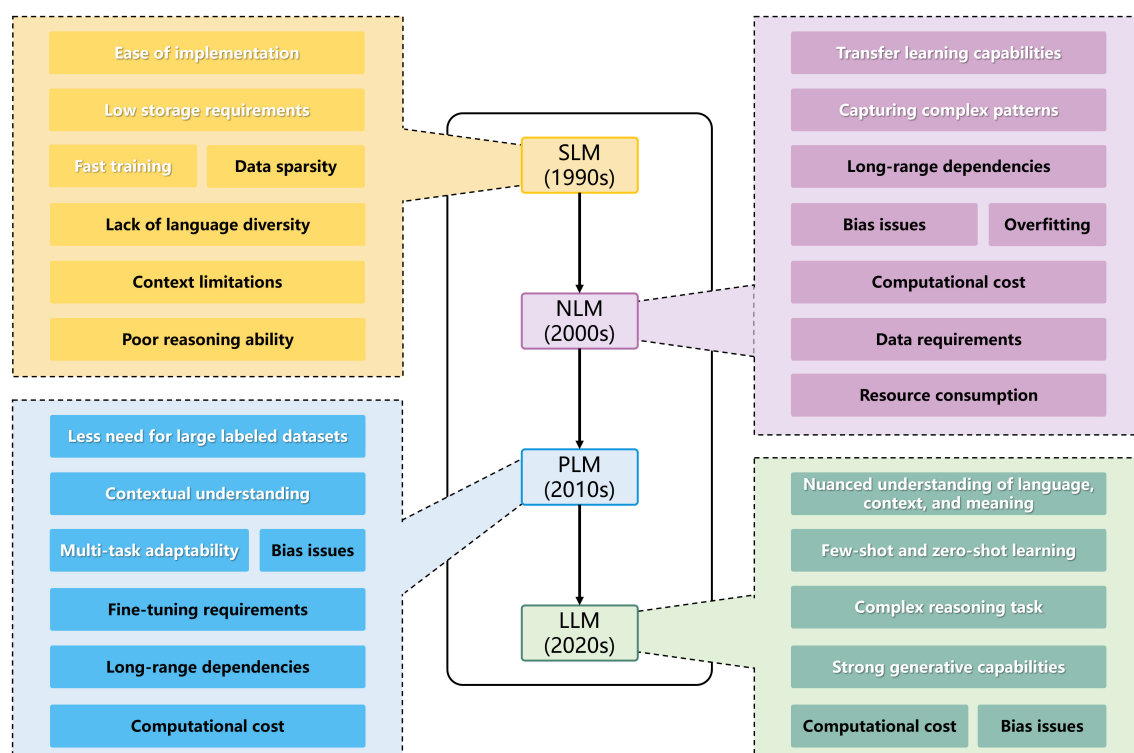


FIGURE 1

Development timeline of NLP models and their pros and cons. White characters represent advantages; black characters represent disadvantages.

Extracts features from large datasets (e.g., ElMo); Fine-tuning: Transfers entire model parameters to specific tasks, exemplified by BERT and GPT. Transformers, introduced by Google, employ a self-attention mechanism, facilitating better training and performance (Vaswani et al., 2017), GPT is fine-tuned from the Transformer. Due to the significant acceleration of model training by Transformer, it has gradually become the fundamental architecture for LLMs.

4. Large Language Models (LLM): LLMs have billions of parameters and exhibit unique capabilities, known as “emergent abilities”. Research shows that larger models perform better and are more sample-efficient. For instance, GPT-3 can generate expected outputs from input sequences without additional training, a feat beyond smaller models like GPT-2.

The transition from SLM to LLM signifies a progressive increase in model complexity, data handling abilities, and adaptability to tasks. Each new generation improves upon its predecessor to overcome limitations, fostering advancement in natural language processing. As shown in Figure 1, compared to other models, LLMs have a comprehensive understanding of language and excel at complex reasoning. Their strong few-shot, zero-shot, and generative capabilities allow them to adapt to new tasks with minimal examples. However, high computational costs and bias issues prevent them from being perfect. The high computational cost remains an unresolved challenge in today’s era of large data training. Bias issues can be mitigated through a series of review and regulatory measures, which will be detailed in section 4.

2.1.2 Key advancements and capabilities of large vision models

LVMs are a new generation of models associated with CV, characterized by their immense scale and broad pre-training. Initially, LVM might have denoted purely vision-based models

trained solely on image data. However, inspired by multimodal learning in LLM, the concept has evolved to include large models trained on both images and text, enabling rich cross-modal associations. CV models began their development in the 20th century and have continued to evolve significantly to the present day (Figure 2). Fueled by the availability of massive image datasets, the development of powerful DL architectures, and significant progress in large-scale pre-training techniques, LVMs have become one of the major development trends in CV models in recent years.

The research on CV models initially focused on shallow image feature extraction algorithms, including scale-invariant feature transform, histogram of oriented gradient, and other methods, but had significant limitations. In 2012, AlexNet (Krizhevsky et al., 2017) achieved a breakthrough success in ImageNet large scale visual recognition challenge, sparking a wave of convolutional neural networks (CNN) for vision models. With the development of DL, deep residual networks including VGGNet (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), and ResNet (He et al., 2016) were successively proposed, which improved the performance of image classification, object detection, semantic segmentation, etc. The boom of the Internet also enabled large-scale image datasets to be used for training vision models. Faster R-CNN (Ren et al., 2016), YOLO (Redmon et al., 2016), Mask R-CNN (He et al., 2017) emerged one after another.

In recent years, Transformer has been successfully applied in the domain of LVM, leading to the emergence of models like Vision Transformer (ViT) (Dosovitskiy et al., 2020) and DALL-E (Ramesh et al., 2021) which have garnered significant public attention. Unlike the traditional DL models mentioned above, LVMs such as ViT leverage transformer architectures and are typically pre-trained on significantly larger and more diverse datasets (e.g., billions of images). This foundational pre-training enables them to develop a more generalized understanding of visual concepts and emergent capabilities, allowing for superior performance on a wide range of downstream tasks, often with only limited domain-specific data.

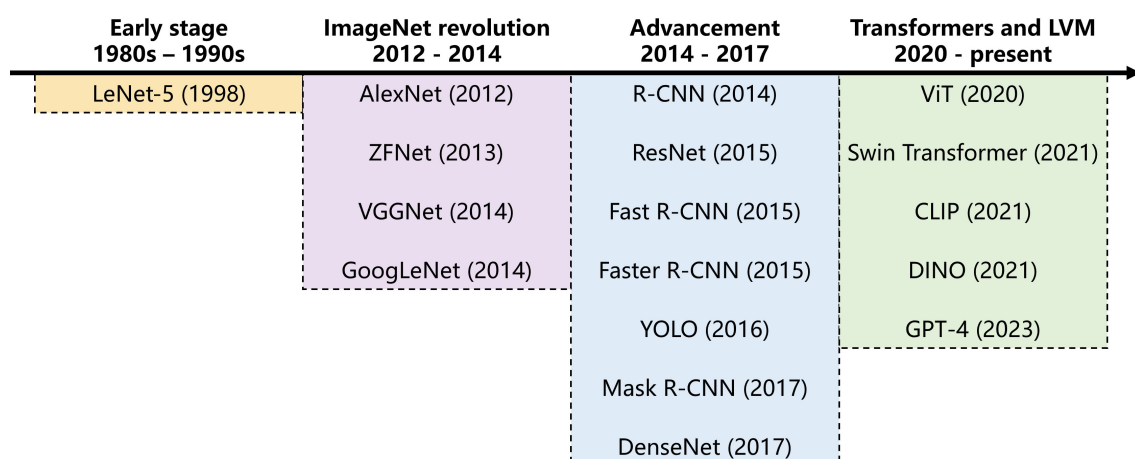


FIGURE 2
Development timeline of computer vision models.

Their ability to grasp complex visual patterns and adapt to new conditions makes them highly versatile.

As detailed in Table 1, large vision models (LVMs) and traditional vision models differ significantly in their core architecture, data requirements, and capabilities. The fundamental distinction lies in their approach to context: Transformer-based LVMs leverage global self-attention to capture broad visual context and long-range dependencies, a significant leap from the local receptive fields of traditional convolutional models (Aymen et al., 2024; Malla et al., 2024). While this architectural shift grants LVMs superior generalization, it also introduces challenges like high computational demands and data hunger. Notably, research is actively addressing these limitations. For instance, Shi et al. (2024) proposed “Scaling on Scales (S²),” which enhances performance by increasing image scales rather than model size, providing new insights for the future development of vision models.

2.1.3 The emergence of multimodal large language models

In addition to the LLMs and LVMs introduced above, MLLMs are also a research focus in the domain of AI. While LLMs perform well in text-based tasks, their capabilities alone cannot effectively reason about information presented in non-textual formats. Although LVMs perform well in the field of CV and possess some NLP abilities, researchers are not content with large models solely trained on text and images. MLLMs (Wu et al., 2023a) integrate multiple data types, such as images, text, language, audio, and more. It not only possesses the advantages of LLMs and LVMs, but also address the limitations of LLMs and LVMs by integrating multiple modalities, enabling a more comprehensive understanding of various data. It can be said that the developments in MLLMs have set up new avenues for AI, which make binary machines to understand and then process various data types (Wu

et al., 2023a). For agriculture, MLLM allows tasks to no longer be confined to just images or text; instead, it can leverage both, and even utilize multimodal inputs like audio and video, breaking the limitations of images and text.

2.2 Current applications of large models in other domains

As shown in Table 2, many LLMs are designed to develop chatbots (BLOOM (Le Scao et al., 2023), PaLM2 (Anil et al., 2023), ERNIE 4.0) or complete NLP tasks, including text classification, machine translation, and sentiment analysis [OPT (Zhang et al., 2022b)]. Similarly, LVMs are primarily engineered to interpret and process visual information. They excel at core CV tasks such as image classification, object detection, segmentation, and image generation, often forming the foundation for systems needing to understand or interact with the visual world. Models like InternImage (Wang et al., 2023) and LLaVA (Liu et al., 2023b) represent efforts to enhance performance on complex visual analysis tasks, aiming to simulate and automate human visual processes.

Although LLM and LVM satisfies some functions and takes large models a big step towards artificial general intelligence (AGI), it is not enough to achieve the goal that machines can emulate human thinking and carry out a wide range of general tasks through transfer learning and diverse other modalities without achieving the multimodality of the model (Zhao et al., 2023b). Some large models have implemented multimodality, enabling them to analyze different types of information [GPT-4 (Bubeck et al., 2023), LLaMA, Gemini (Team et al., 2023), ImageBind (Girdhar et al., 2023)] and interact with users. It is worth mentioning that most of the newer large models are MLLMs, and many models that were originally LLMs or LVMs have gradually acquired multimodal capabilities after multiple updates.

However, many current models are generic models and their training datasets are too broad, they cannot provide a satisfactory answer to knowledge in certain professional fields. As Goertzel (2014) believed, for a system to be considered AGI, it is not necessary for it to have infinite generality, adaptability, or flexibility. Therefore, some researchers have optimized and adjusted existing large models and have released some large models specifically for a single field. BloombergGPT can be used in the financial field, showcasing remarkable performance on general LLM benchmarks and surpassing comparable models on financial tasks (Wu et al., 2023c). The meteorological model in panguLM developed by Huawei can provide predictions of variables such as gravity potential, humidity, wind speed, temperature, and pressure within 1 hour to 7 days. Embedding PaLM-E into robots can achieve multiple specific tasks, like visual question answering, sequential robotic manipulation planning, and captioning (Driess et al., 2023). OceanGPT is an expert in various marine science tasks (Bi et al., 2023). It exhibits not only a higher level of knowledge expertise for oceans science tasks but also acquires preliminary embodied intelligence capabilities in ocean engineering. PMC-

TABLE 1 Comparison of large vision models and traditional vision models.

Feature/ Aspect	Large vision models	Traditional vision models
Core architecture	Transformer-based, global self-attention	Convolution-based, local receptive fields
Context & dependencies	Global context, excels at long-range dependencies	Local focus, struggles with long-range dependencies
Image handling	Processes image patches; more robust to variations	Uses sliding filters; sensitive to some variations
Data needs	Best with large-scale pre-training	Can work with smaller datasets, benefits from pre-training
Multimodal ability	Stronger, more inherent multimodal integration	Requires more specialized designs for multimodal
Parallelism	High (sequence processing)	Good for convolutions, poor for sequential tasks
Key advantages	Global understanding, long dependencies, scalability	Efficient local feature extraction
Key limitations	Can be data-hungry for pre-training	Limited global view, adversarial vulnerability

TABLE 2 The currently popular and representative large models.

Original version	Latest version	Release date (original)	Release date (latest)	Types (original → latest)	Information	References
OPT	/	May 2 rd , 2022	/	LLM	OPT promotes transparency, reproducibility, and broader community engagement and innovation in NLP research. (open source)	(Zhang et al., 2022b)
BLOOM	BLOOMZ	July 12 th , 2022	December 15 th , 2022	LLM	A decoder only model based on Transformer architecture. (open source)	(Le Scao et al., 2023)
PMC-LLaMA	/	April 27 th , 2023	/	LLM	Inject medical knowledge into existing LLM using 4.8 million biomedical academic papers. (open source)	(Wu et al., 2024)
PaLM2	/	May 11 st , 2023	/	LLM	PaLM2 was a neural network-based language model that was considered one of the most advanced language models available at the time of its release in May 2023.	(Anil et al., 2023)
BloombergGPT	/	March 30 th , 2023	/	LLM	A LLM for the financial field.	(Wu et al., 2023c)
OceanGPT-Basic-7B	OceanGPT-Basic-14B/7B/2B	October 3 rd , 2023	July 4 th , 2024	LLM	OceanGPT is the first LLM in the ocean domain. (open source)	(Bi et al., 2023)
DeepSeek LLM	DeepSeek-R1	November 29 th , 2023	January 20 th , 2025	LLM	DeepSeek-R1 excels in complex tasks such as mathematics, coding, and natural language reasoning	(Guo et al., 2025a)
Llama 2	Llama 4	July 20 th , 2023	April 6 th , 2025	LLM → MLLM	A series of large models released by Meta.	/
Qwen-7B	Qwen2.5-Omni-7B	August 3 rd , 2023	March 27 th , 2025	LLM → MLLM	A super large model launched by Alibaba Cloud. (open source)	(Bai et al., 2023)
Kimi Chat	Kimi k1.5	October 10 th , 2023	January 20 th , 2025	LLM → MLLM	Kimi k1.5 surpasses GPT-4o by 550% in mathematics, coding, and other capabilities under short-chain thinking mode.	/
Gemma	Gemma 3	February 21 st , 2024	March 12 th , 2025	LLM → MLLM	Gemma 3 is a MLLM released by Google. (open source)	(Team et al., 2024, 2025)
PaLM-E	/	March 6 th , 2023	/	LVM	PaLM-E can integrate vision and language into robot control.	(Driess et al., 2023)
InternImage	InternImage-H	November 10 th , 2022	October 4 th , 2023	LVM	A LVM based on deformable convolution. (open source)	(Wang et al., 2023)
PanguCVLM 3.0	PanguCVLM 5.0	July 7 th , 2023	June 21 th , 2024	LVM	A LVM that simulates and automates human visual processes.	/
LLaVA	LLaVA-NeXT (Stronger)	April 17 th , 2023	May 10 th , 2024	LVM	LLaVA has the ability to align and fuse the visual information of images with the semantic information of text. (open source)	(Liu et al., 2023b)
mPLUG-Owl	mPLUG-Owl3	April 27 th , 2023	August 20 th , 2024	LVM → MLLM	mPLUG-Owl is developed by Alibaba DAMO Academy. (open source)	(Ye et al., 2023, 2024)
SPARK 1.0	SPARK 4.0 Turbo	May 6 th , 2023	October 24 th , 2024	LVM → MLLM	A new generation of cognitive intelligence model with Chinese as its core.	/
Claude 3	Claude 3.7 Max	March 4 th , 2024	March 18 th , 2025	MLLM	A MLLM that primarily focuses on code processing.	/
ERNIE 4.0	ERNIE 4.5	October 17 th , 2023	March 16 th , 2025	MLLM	ERNIE is a new generation of Baidu's large model for knowledge enhancement.	/
ImageBind	/	May 9 th , 2023	/	MLLM	ImageBind is the first AI model that can bind information from six modes.	(Girdhar et al., 2023)
GPT-4	GPT-4.5	March 14 th , 2023	February 28 th , 2025	MLLM	GPT-4.5 significantly enhances its knowledge reserves and emotional intelligence by expanding unsupervised learning and reasoning capabilities.	(Bubeck et al., 2023)

(Continued)

TABLE 2 Continued

Original version	Latest version	Release date (original)	Release date (latest)	Types (original → latest)	Information	References
Skywork	Skywork 4.0	April 17 th , 2023	January 6 th , 2025	MLLM	Skywork is a series of large models developed by the Kunlun · Skywork team.	(Wei et al., 2023)
Gemini	Gemini 2.5	December 6 th , 2023	March 26 th , 2025	MLLM	Gemini is a MLLM launched by Google DeepMind.	(Team et al., 2023)
Sora	/	February 15 th , 2024	/	MLLM	Sora can create realistic and imaginative scenes from text instructions.	(Peebles and Xie, 2023)
Hunyuan-t1	Hunyuan-t1 (official)	February 17 th , 2025	March 21 st , 2025	MLLM	Hunyuan-t1 is a deep-thinking model independently developed by Tencent.	/

The arrow '→' represents the change in the model's category, from its original type to the type of its latest version.

LLaMA represents the pioneering open-source medical specific language model that demonstrates exceptional performance on diverse medical benchmarks, outperforming ChatGPT and LLaMA-2 with significantly fewer parameters (Wu et al., 2024). The success of large models across diverse fields, as highlighted in this section, underscores their potential to generalize and tackle complex problems, suggesting their applicability to the intricate tasks within agriculture.

2.3 Assessing the attention to large models within agriculture

In the past few decades, the advancement of agricultural technology has significantly improved global agricultural production efficiency. According to the forecast released by the food and agriculture organization (FAO) of the United Nations, the global grain production in 2023 was 2.84 billion tons, nearly twice that of the early 20th century. Although global agricultural production efficiency is high, the world population is also constantly growing. Continuously improving agricultural production efficiency is the lifeline of economic development and the foundation for ensuring human food, clothing, and survival needs. Hence, how to make agricultural practices advance is a crucial issue. Next, we will use bibliometric methods in conjunction with practical analysis to explain why large models are important for agriculture.

2.3.1 Bibliometric analysis and data sources

Bibliometrics is a quantitative analysis method that integrates mathematics, statistics, and bibliography, based on mathematical statistics. It focuses on the external characteristics of scientific literature to conduct statistical and visual analyses of the literature (Wang et al., 2019). Keywords encapsulate important information about the research topic. They can intuitively reflect the themes and content of the study, reveal the connections between research contents, results, and characteristics in a particular field, and demonstrate the research dynamics and emerging trends within that area (Li and Zhao, 2015). Our analysis used two methods:

1. Searching for research literature related to large models using the Science Citation Index Expanded (SCI-E) and Social Sciences Citation Index (SSCI) from Web of Science (WoS) with keywords such as “large models”, “large language models”, “large vision models”, or “foundation models”, covering the period from 2019 to 2024.
2. Collecting 3,496 papers from the arXiv in the field of artificial intelligence from 2019 to 2024 and categorizing them by discipline based on keyword searches.

Our analysis of WoS aimed to identify the established trends and peer-reviewed research regarding large models, and specifically, the frequency of agriculture-related keywords within this body of published work. This provides a view of the validated research landscape. Complementarily, we included arXiv to capture the more recent and rapidly evolving trends in artificial intelligence research. arXiv, as a leading platform for pre-prints in AI, offers valuable insights into emerging research directions and the early exploration of applying large models across various disciplines, including potential initial interest in agriculture. Pre-prints often precede formal publication, providing a timelier snapshot of the research frontier.

By analysing both published literature (WoS) and pre-prints (arXiv), we aimed to gain insights from two different perspectives: the established, peer-reviewed research landscape and the more immediate, evolving research front. This allows us to observe both the current state of validated research and the potential emerging trends and initial explorations within the field.

2.3.2 Detailed analysis and design protocol

As described in section 2.3.1, two data sources were used for the specific analysis method: (1) WoS; (2) arXiv. Next, we will elaborate on the details of using these two bibliometric analysis methods.

For the Method 1, after entering the official website of WoS, search in “Web of science Core Collection” and select Science Citation Index Expanded (SCI-EXPANDED) and Social Sciences Citation Index (SSCI) in edition. Both SCI-EXPANDED and SSCI primarily index peer-reviewed journals with established reputations within their respective fields. This ensures a certain level of quality control and scholarly rigor in the literature being analysed. Then search for topics with the keywords “large models”, “large language models”, “large

vision models”, or “foundation models”, covering articles from 2019 to 2024, and export the authors, titles, sources and abstracts of these articles in plain text file. Finally, import these plain text files into VOSviewer to draw a network of keywords in the field of large models. For the Method 2, we use the keyword “artificial intelligence” to search for articles on arXiv, and crawl the relevant articles from 2019 to 2024, including title, author, abstract and other information, to build a csv file. Then search this file according to relevant keywords. For example, in the medical field, keywords such as medical, healthcare, hospital, etc. are used to filter out relevant articles and count the number. Finally, a graph of the proportion of articles in different fields under the AI domain is constructed.

In this way, we obtained a network map of keyword through the Method 1, and a graph of the proportion of different fields in the AI domain through the Method 2. The specific results and analysis will be explained in the next section.

2.3.3 Analysis results

A total of 1,789 papers were filtered using Method 1, and a network diagram of keyword occurrences was generated using VOSviewer. As shown in Figure 3, the term ‘agriculture’ appears infrequently in these large model papers, indicating that large models have not received widespread attention in the agricultural field.

The reasons why large models have not received attention in the agricultural sector are diverse. First, large models are a relatively new technology that has emerged in recent years, and many researchers and practitioners in agriculture may not fully understand their capabilities and potential applications in the field. Second, implementing large models often requires substantial computational resources and expertise, which may not be easily accessible in many agricultural

environments. Third, agricultural tasks can be very specific and localized, leading people to prefer traditional methods over large models.

Moreover, Figure 4 illustrates a difference: the application and research of large models in agriculture are currently limited compared to other fields. This observation from our bibliometric analysis (Figures 3, 4) suggests that despite the evident potential of large models to address agricultural challenges discussed in the introduction, the field is still in the early stages of exploring and adopting this technology. Therefore, a detailed review of their potential applications, associated challenges, and responsible deployment is crucial to guide future research and accelerate their integration into agriculture.

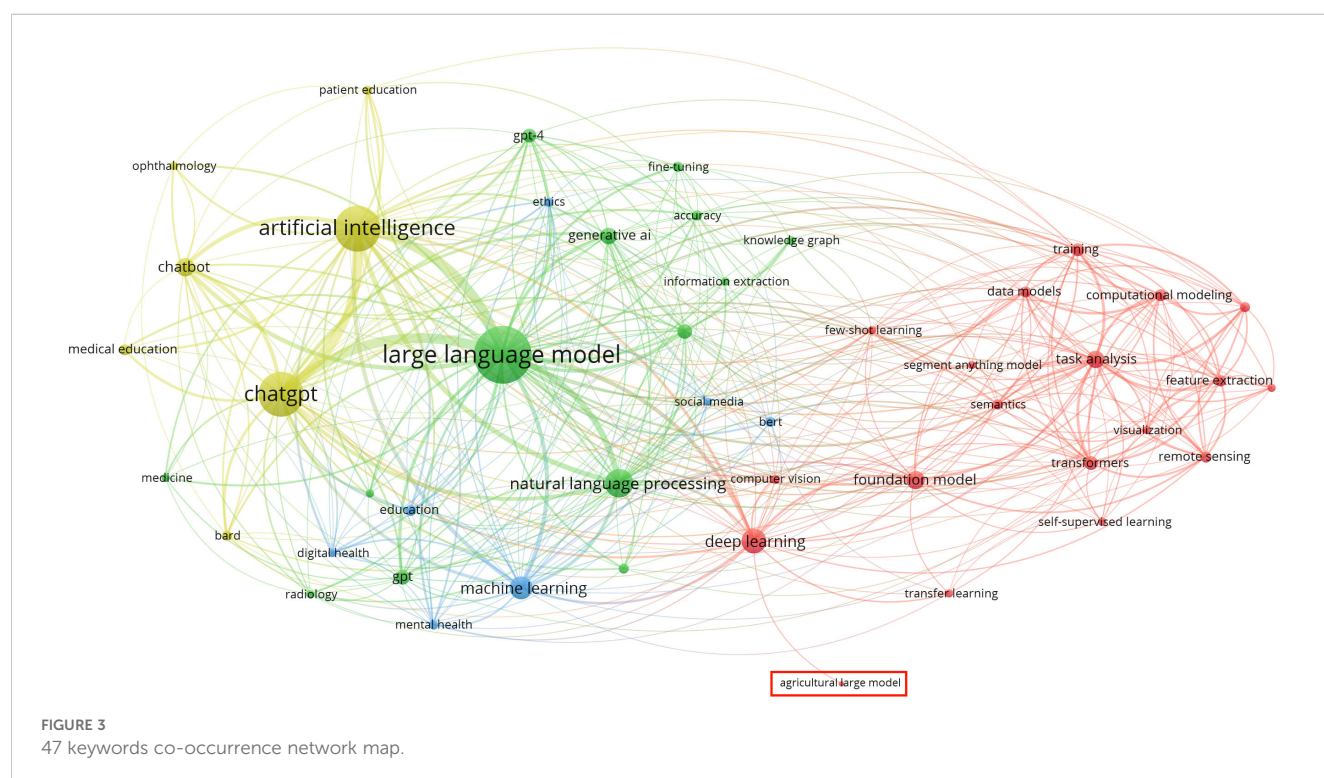
3 Large models in agricultural applications

As mentioned in the introduction, agriculture faces multiple challenges, including pests and diseases, seed quality, and crop grading. Large models have demonstrated significant potential in addressing these issues, and some researchers have already developed models specifically tailored for the agricultural domain.

3.1 Emerging potential and existing applications of large models in agriculture

3.1.1 Potential and applications

Many large models have emerged, and although they are not yet truly applied in agriculture, their problem-solving capabilities



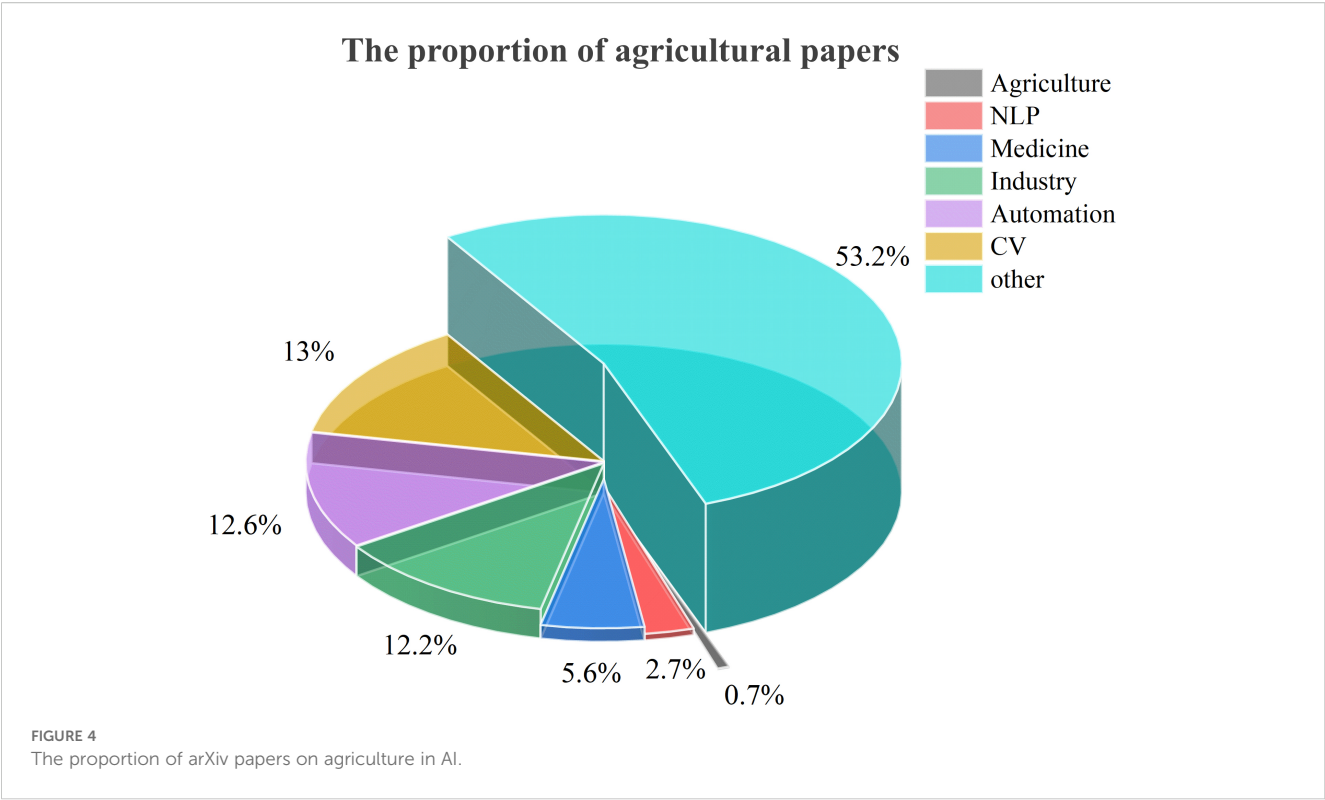


TABLE 3 Large models with agricultural potential.

Type	Based	Method	Problem	Application prospect	References
LLM	ChatGPT	GPT-3.5-turbo	Agricultural information extraction	Rapid querying of agricultural information	(Peng et al., 2023)
		AugGPT	Text data augmentation	Few-shot learning for agricultural data	(Dai et al., 2023)
	/	Aurora	Atmospheric prediction	Predicting weather in agriculture	(Bodnar et al., 2024)
LVM	/	MAE, DINO, DINOv2	Plant phenotyping tasks	Monitoring crop health	(Chen et al., 2023a)
	PaLM, ViT	PaLM-E	Robot control	Agricultural intelligent machines or robots	(Driess et al., 2023)
MLLM	SAM	TAM	Object tracking and segmentation in videos	Monitor animals in agricultural farming	(Yang et al., 2023a)

indicate potential prospects in agricultural applications. As shown in Table 3, some large models are modifications of existing models, while others are entirely original. For example, given 50 original descriptions related to “wheat rust,” AugGPT can generate 200+ expanded samples covering different growth stages and climatic conditions, thereby enhancing the robustness of disease identification models in complex environments. Aurora is a large model for weather forecasting (Bodnar et al., 2024), and if applied in agriculture, it could enable farmers to schedule activities such as planting, fertilizing, and harvesting based on accurate weather forecasts, as well as proactively mitigate losses from extreme weather events. In addition to ordinary large models, there are

also some special existences. HuggingGPT is an AI agent framework designed to orchestrate multiple specialized models, including LLMs like ChatGPT (Shen et al., 2024). It acts as a ‘model coordinator,’ integrating and managing diverse AI components to enhance decision-making in complex scenarios such as agricultural planning. This capability offers possibilities for managing a series of complex agricultural tasks, from planting to harvesting.

Notably, there are already large models applied in agriculture (Table 4). For instance, TimeGPT demonstrates its capability as a smart agriculture tool (Deforce et al., 2024), being used for predicting soil moisture, which helps farmers determine whether the soil is suitable for certain crops. FMFruit showcases the

importance of large models in agricultural detection tasks (Li et al., 2024), providing new directions and foundations for the development of robotic harvesting systems. ITLMMLP performs disease recognition on cucumbers with limited sample sizes, playing a significant role in agricultural automation and intelligence (Cao et al., 2023).

Tables 2, 3 demonstrate the feasibility and importance of large models in agriculture, where many agricultural tasks involve complex reasoning. For example, when presented with an image of a soybean field, agricultural scientists or farmers rely on large models to undertake several key steps. Firstly, the large model must identify any abnormal symptoms evident in the soybean leaves, such as leaf wrinkling. Subsequently, it must ascertain the name of the specific problem that troubles plants, such as soybean mosaic. Next, the model needs to determine the underlying cause of the disease, such as soybean mosaic virus. Finally, it must develop an appropriate treatment strategy. This multi-step, cross modal diagnostic and decision-making process is precisely the unique advantage that large models can demonstrate compared to traditional DL models with a single task.

Many question answering (QA) and dialogue systems are designed to address this type of reasoning problem (Rose Mary et al., 2021; Mostaco et al., 2018; Niranjan et al., 2019). For instance, a chatbot based on a RNN is specifically designed to handle questions related to soil testing, plant protection, and nutrient management (Rose Mary et al., 2021). Although, these QA and dialogue systems and chatbots can answer most inquiries without

the need for human interaction and with excellent accuracy, they have limited capabilities for complex problems by reason of their small model size as well as of inadequate training data. Therefore, the agricultural domain requires large models to promote the development of QA and dialogue systems and chatbots. The traditional methods for detecting crop pests and diseases mainly rely on special methods such as serology and molecular biology-based technical means, in addition to artificial visual evaluation. Although these methods can accurately determine pests and diseases to a certain extent, they often require a lot of time and money. And some methods of sampling crops often lead to crop damage, which goes against the original intention of diagnosing diseases and pests to protect crops. Therefore, image processing and analysis is an important task for large models in the field of agriculture, and another important task is to embed LVMs into robots to solve some agricultural problems (Weeding, pruning branches, harvesting, etc.) and achieve automated agriculture.

3.1.2 The advantages of agriculture-specific large models

In the field of agriculture, agriculture-specific models can offer notable advantages over general large models, particularly by effectively integrating diverse, domain-relevant data modalities such as image, text, and crucial label information. This multimodal strategy, often employing techniques like combined contrastive learning methods within a unified feature space, allows these models to address the prevalent challenge of data scarcity in

TABLE 4 Agricultural large models.

Type	Name	Achievement	Significance	References
LLM	TimeGPT	Predicting soil moisture	Contributes to sustainable agricultural practices	(Deforce et al., 2024)
	ChatGPT	Designed a tomato-picking robot	Simplify the design process of agricultural robots	(Stella et al., 2023)
	FMFruit	Identifying multiple types of fruits	Research on robotic harvesting and fruit detection	(Li et al., 2024)
	AgriGPT	Multimodal agricultural knowledge Q&A	Promote precision agriculture practices	(Liu et al., 2023a)
	ShenNong	Development of specialized large models for multiple agricultural domains	Driving agricultural intelligence and comprehensive efficiency improvement	/
	ChatAgri	Cross-linguistic classification of agricultural texts	Provide decision support for precision agriculture	(Zhao et al., 2023a)
LVM	SpectralGPT	Process spectral remote sensing data	Greatly enhanced the processing capability of agricultural spectral data	(Hong et al., 2024)
	SAM	Chicken segmentation and tracking	Facilitates segmentation and tracking tasks in agriculture	(Yang et al., 2023c)
		Agricultural field boundary delineation	Beneficial for PA, crop monitoring, and yield estimation	(Tripathy et al., 2024)
MLLM	ITLMMLP	Cucumber disease recognition	Agricultural disease recognition	(Cao et al., 2023)
	AIE-SEG	High-precision segmentation of agricultural imagery	Enables automated field monitoring and yield estimation	(Xu et al., 2023)

agriculture more effectively than models relying solely on single modalities or vast, generic pre-training datasets.

By explicitly learning and leveraging the semantic correlations between visual features (e.g., specific crop disease symptoms) and related textual descriptions or categorical labels, agriculture-specific models can develop more comprehensive, robust, and discriminative representations tailored to the nuances of the field. For example, ChatAgri excels in the specific task of agricultural visual diagnostics (Zhao et al., 2023a). A general MLLM might identify visual anomalies, and lack the specialized knowledge to accurately name the specific agricultural disease or pest, understand its lifecycle, or recommend appropriate, targeted treatments. Especially when faced with limited training samples, agriculture-specific large models may perform better compared to models with poor adaptability. Unlike general large models that often require vast datasets for pre-training and may not adapt well to fine-tuning on limited agricultural data, ITLMLP is designed to be effective with small sample sizes. It extracts richer and more discriminative features from scarce data, leading to significantly higher recognition accuracy (achieving 94.84% in their paper) compared to general large models to the same small dataset (Cao et al., 2023).

Furthermore, their focused training enables them to better identify and weigh agriculturally significant features, accurately discerning subtle but critical patterns for tasks like disease recognition while potentially mitigating the influence of irrelevant background elements, ultimately resulting in improved accuracy, reliability, and greater practical applicability within the complex agricultural environment.

3.2 Leveraging large language models for agricultural data processing, insights, and decision support

LLM can play many roles in the agricultural domain, such as processing and generating agricultural data, providing insights into agricultural production work, and supporting agricultural decision-making for farmers.

3.2.1 Large language models for processing and generating agricultural data

3.2.1.1 Information extraction

LLMs can extract structured information from unstructured agricultural text data. First, the text is divided into individual tokens and LLMs represent each token as a numerical vector called a word embedding. Then, LLMs analyse the surrounding context of each token to understand its meaning within the sentence or document, and identify and categorize named entities within the text, like names of individuals, locations, organizations, or specific agricultural terms. Finally, LLMs employ techniques like information extraction to identify and extract structured information from unstructured text (Involve identifying relationships between entities, extracting key facts, or populating knowledge graphs). LLMs extract information from data using a process known as NLP. Beyond mere extraction, modern LLM applications

increasingly employ a paradigm known as retrieval-augmented generation (RAG). In this approach, the LLM first retrieves relevant, up-to-date information from external, domain-specific knowledge bases—such as recent agronomic research, real-time market prices, or local pest outbreak databases. This retrieved context then “augments” the model’s input, enabling it to generate responses that are not only more accurate and timelier but also grounded in verifiable sources, thereby significantly mitigating the risks of data lag and factual inaccuracies in the agricultural domain (Gao et al., 2023).

3.2.1.2 Agricultural data generation

Generative AI models are a multimodal LLM, which is the MLLM mentioned above. An obstacle encountered when applying specialized CV algorithms to agricultural vision data is the insufficient availability of training data and labels (Qi et al., 2017; He et al., 2017). In addition, collecting data that encompasses the wide range of variations caused by season and weather changes is exceedingly challenging. Acquiring high-quality data requires a lot of time, and labelling them is even more costly (Zhou et al., 2017). To address these challenges, one approach is to fine-tune multimodal generative LLMs on the target agricultural data domain. This allows the models to generate massive training data and labels, thereby constructing an augmented training set that closely resembles the distribution of the original data (Dai et al., 2023). Besides, text-based generation models can generate images (Rombach et al., 2022) and videos (Ho et al., 2022) of specific scenes based on text descriptions, thereby supplementing training datasets that may lack certain visual content. This helps in expanding the training data and improving the performance of downstream models.

3.2.2 Large language models provide insights

LLMs possess the capability to analyse textual data and uncover trends in agricultural practices, market conditions, consumer preferences, and policy developments. Through analysis of agricultural text data from sources such as news articles, reports, and social media, these models can offer valuable insights into market dynamics and pricing trends (Yang et al., 2024). This provides support for farmers to understand domains outside of agriculture. Many researchers believe that the integration of LLMs into different stages of design and development for agricultural applications is also experiencing a noticeable rise (Stella et al., 2023; Lu et al., 2023). In (Stella et al., 2023) study, Stella et al. incorporated LLM into the design phase of robotic systems. They specifically focused on designing an optimized robotic gripper for tomato picking and outlined the step-by-step process. In the initial ideation phase, they leveraged LLMs like ChatGPT (Bubeck et al., 2023) to gain insights into the possible challenges and opportunities associated with the task. Building upon this knowledge, they identified the most promising and captivating pathways, engaging in ongoing discussions with the LLM to refine and narrow down the design possibilities. Throughout this process, the human collaborator harnesses the expansive knowledge of the LLM to tap into insights transcend their individual expertise. In the following stage of the design process, which emphasizes technical aspects, the

broad directions derived from the collaboration need to be transformed into a real, completely functional robot. Although LLMs do not provide comprehensive technical support, they can offer their own insights on whether the technology is feasible, helping researchers reduce the risk of failure.

Presently, LLMs lack the ability to generate comprehensive CAD models, evaluate code, or independently fabricate robots. Nevertheless, advancements in LLM research suggest that these algorithms can offer significant assistance in executing software (Chen et al., 2021), mathematical reasoning (Das et al., 2024), and even in the generation of shapes (Ramesh et al., 2022). Lu et al. specifically focused on the utilization of LLMs for organizing unstructured metadata, facilitating the conversion of metadata between different formats, and discovering potential errors in the data collection process (Lu et al., 2023). They also envisioned the next generation of LLMs as remarkably potent tools for data visualization (Bubeck et al., 2023), and anticipated that these advanced models will provide invaluable support to researchers, enabling them to extract meaningful insights from extensive volumes of phenotypic data.

Although LLMs provide insights can indirectly help farmers solve a small number of agricultural tasks, it's important to note that their insights should be used in conjunction with human judgment and domain expertise. That is to say, the insights provided by LLMs cannot be separated from human experience.

3.2.3 Large language models empower decision-making for farmers

According to a recent study, ChatGPT demonstrates the ability to comprehend natural language requests, extract valuable textual and visual information, select appropriate language and vision tasks, and effectively communicate the results to humans (Shen et al., 2024). Shen et al. proposed a system named HuggingGPT to solve AI tasks. HuggingGPT is a collaborative AI task resolution framework built on LLMs. This system connects LLM with AI models through language interface, and these AI models are derived from HuggingFace. This coordinating capability positions LLMs as the core of modern AI Agents. As the core of decision-making, LLM can be applied to agriculture to help solve the tasks proposed by farmers (Shen et al., 2024).

An AI Agent is an autonomous system that perceives its environment, reasons, plans, and acts to achieve specific goals. As shown in Figure 5, the LLM acts as the agent's "brain", performing crucial functions. When receiving a task request, LLM first divides the total task into subtasks and selects the appropriate AI model based on the needs of each subtask. For example, converting farmers' audio into text requires the use of an audio to text model [Amazon transcribe, Whisper (Radford et al., 2023)]; It is also necessary to recognize the sent image and integrate the text obtained from the audio conversion in the previous step to obtain a text-response (vit-gpt2); Considering that some farmers may have had limited access to formal education, it is necessary to further convert text-response into audio and ultimately obtain the audio-response [FastSpeech (Ren et al., 2019, 2020)]. Although LLM does not play a role in solving problems throughout the entire system, as

a "conductor", it can coordinate various AI models to complete subtasks, thereby gradually solving complex tasks and playing a core role in decision-making support.

3.3 The role of large vision model in image processing, analysis, and agricultural automation

While LLMs excel in processing textual and knowledge-based information, many agricultural tasks fundamentally rely on visual data. Using a LVM to judge crop related information can not only greatly improve the time required for judgment, but also indirectly reduce the damage caused to crops. Moreover, after crops are invaded by pests and diseases, their color, texture, spectral characteristics will undergo certain changes, all of which are related to CV.

3.3.1 Image processing and analysis

At present, there are four types of methods for obtaining crop image information: 1) ordinary channels, taking photos to obtain images; 2) obtaining remote sensing images through agricultural machinery near the ground; 3) obtaining remote sensing images through aircraft monitoring platforms (Yuan et al., 2022); 4) obtaining remote sensing images through satellites (Zhang et al., 2019). Remote sensing can provide large-scale land use and land cover information. By analysing satellite images or high-altitude images, various surface information can be identified, such as surface conditions, soil moisture, vegetation coverage, and crop growth status (Khanal et al., 2020). Classifying and segmenting from limited examples obtained from remote sensing is a significant challenge. Regarding this, Wu et al. (2023b) put forward GenCo (a generator-based two-stage approach) for few-shot classification and segmentation on remote sensing and earth observation data. Their approach presents an alternative solution for addressing the labelling challenges encountered in the domains of remote sensing and agriculture. Spectral data can provide rich insights into the composition of observed objects and materials, especially in remote sensing applications. The challenges faced in processing spectral data in agriculture include: 1) effectively processing and utilizing vast amounts of remote sensing spectral big data derived from various sources; 2) deriving significant knowledge representations from intricate spatial-spectral mixed information; 3) addressing the spectral degradation in the modelling of neighbouring spectral relevance. Hong et al.'s SpectralGPT empowers intelligent processing of spectral remote sensing big data, and this LVM has also demonstrated its excellent spectral reconstruction capabilities in agriculture (Hong et al., 2024). Due to multispectral imaging (MSI) and hyperspectral imaging (HSI) make it possible to monitor crop health in the field. The integration of remotely sensed multisource data, such as HSI and LiDAR (Light detection and ranging), enables the monitoring of changes occurring in different parts of a plant (Omia et al., 2023). By using a large visual model to analyse these spectral data, the obtained crop health information can help farmers quickly and

accurately identify diseases and treat them, reducing the loss of crop yield.

Studies suggest that the use of LVMs for image recognition and predictive analysis of crop information is often more effective than traditional ML algorithms. When farmers need to obtain crop information, four types of image acquisition methods can be used to obtain crop image information (Figure 6). Then, the image information is processed through image recognition (Divided into four tasks: image classification, object detection, semantic segmentation, instance segmentation), and the identified results need to be further predictive analytics to obtain crop information that farmers can understand.

In addition to obtaining information by analysing the phenotypic characteristics of crops, Feng et al. (2022) developed a traditional DL model called organelle segmentation network (OrgSegNet). OrgSegNet is capable of accurately capturing the actual sizes of chloroplasts, mitochondria, nuclei, and vacuoles within plant cell, further inspecting plant phenotypic at the subcellular level. They have tested two applications: 1) A thermo-sensitive rice albino leaf mutant was cultivated at cold temperature conditions. In the transmission electron microscope images (TEMs), the albinotic leaves lacked typical chloroplasts, and OrgSegNet failed to identify any chloroplast structures; 2) Young leaf chlorosis 1 (Ylc1). Young leaves of the ylc1 mutant showed lower levels of chlorophyll and lutein compared to corresponding wild type, and its TEM analysis further revealed a noticeable loose arrangement of the thylakoid lamellar structures. It can be imagined that if a large model is used to replace DL algorithms, the recognition of subcellular cells may perform better, and the recognition results can be further predictive analytics to obtain information that non plant experts can also understand.

3.3.2 Automation and robotics

Enhancing the intelligence of agricultural robots is a crucial application area for large models. Conventional agricultural robot systems, typically composed of perception, decision-making, and actuation modules, often struggle with complex visual perception and intelligent, real-time decision-making, especially in unstructured and dynamic farm environments (Yang et al., 2023b; Hamuda et al., 2016). Integrating large models is a promising approach to overcome these limitations and significantly enhance the intellectual features of agricultural robots.

Current LVMs can be used in drones to monitor crops and obtain information on their growth, disease, yield, and other factors (Ganeshkumar et al., 2023; Chin et al., 2023; Pazhanivelan et al., 2023). In addition to the above functions, ground machines that used LVMs can also be used for harvesting and classifying crops, as well as detecting pests up close. In (Yang et al., 2023c), a LVM, segment anything model (SAM) (Kirillov et al., 2023), uses infrared thermal images for chicken segmentation tasks in a zero-shot means. SAM can be used in agriculture to segment immature fruits on a fruit tree and quickly achieve yield prediction. Yang et al. (2023a) subsequently proposed the Track Anything Model (TAM) by combining SAM and video. Unfortunately, TAM places more emphasis on maintaining short-term memory rather than

long-term memory. Nevertheless, based on its capabilities, TAM still has great potential in the agricultural field. If its long-term memory ability can be improved, it can monitor early changes in crop diseases and provide early warning to farmers. Embedding LVMs such as SAM and TAM into robots can not only achieve automation in agriculture, but these LVMs themselves can help achieve automation in agricultural robot design.

Beyond perception, large models are also revolutionizing the design process of agricultural robots. As mentioned previously, Stella et al. (2023) demonstrated using LLMs like ChatGPT to assist in designing an optimized robotic gripper for tomato picking. With the latest multimodal versions like GPT-4.5, designers can now input not only text descriptions but also sketches to partially automate the robot design process. This integration of LVMs for perception and LLMs for both control logic and design automation marks a significant step towards fully autonomous agricultural systems.

3.4 Integration of multimodal models

LVMs provide powerful capabilities for visual analysis and robotic perception. However, the most complex agricultural challenges often require integrating information from multiple sources. MLLM recently has emerged as a prominent research hotspot (Figure 7), which uses powerful LLMs as a core to tackle multimodal tasks (Yin et al., 2023). In recent years, many researchers have utilized and merged diverse types of data inputs, such as text, images, audio, video (Zhang et al., 2023a), sensor data (Driess et al., 2023), depth information, point cloud (Chen et al., 2024), and more.

The agricultural community has started exploring the realm of multimodal learning in agricultural applications. By incorporating multimodal learning techniques, the agricultural community seeks to unlock new opportunities for optimizing various agricultural processes and achieving improved outcomes. As an example, Bender et al. have released an open-source multimodal dataset specifically curated for agricultural robotics (Bender et al., 2020). This dataset was collected from cauliflower and broccoli fields and aims to foster research endeavors in robotics and ML within the agricultural domain. It encompasses a diverse range of data types, including stereo color, thermal, hyperspectral imagery, as well as essential environmental information such as weather conditions and soil conditions. The availability of this comprehensive dataset uses as a precious resource for advancing the development of innovative solutions in agricultural robotics and ML. Cao et al. (2023) proposed a novel approach for cucumber disease recognition using a MLLM that incorporates image-text-label information. Their methodology effectively integrated label information from many domains by employing image-text multimodal contrastive learning and image self-supervised contrastive learning. The approach facilitated the measurement of sample distances within the common image-text-label space. The results of the experiment demonstrated the effectiveness of this innovative approach, achieving a recognition accuracy rate of 94.84% on a moderately sized multimodal cucumber disease dataset.

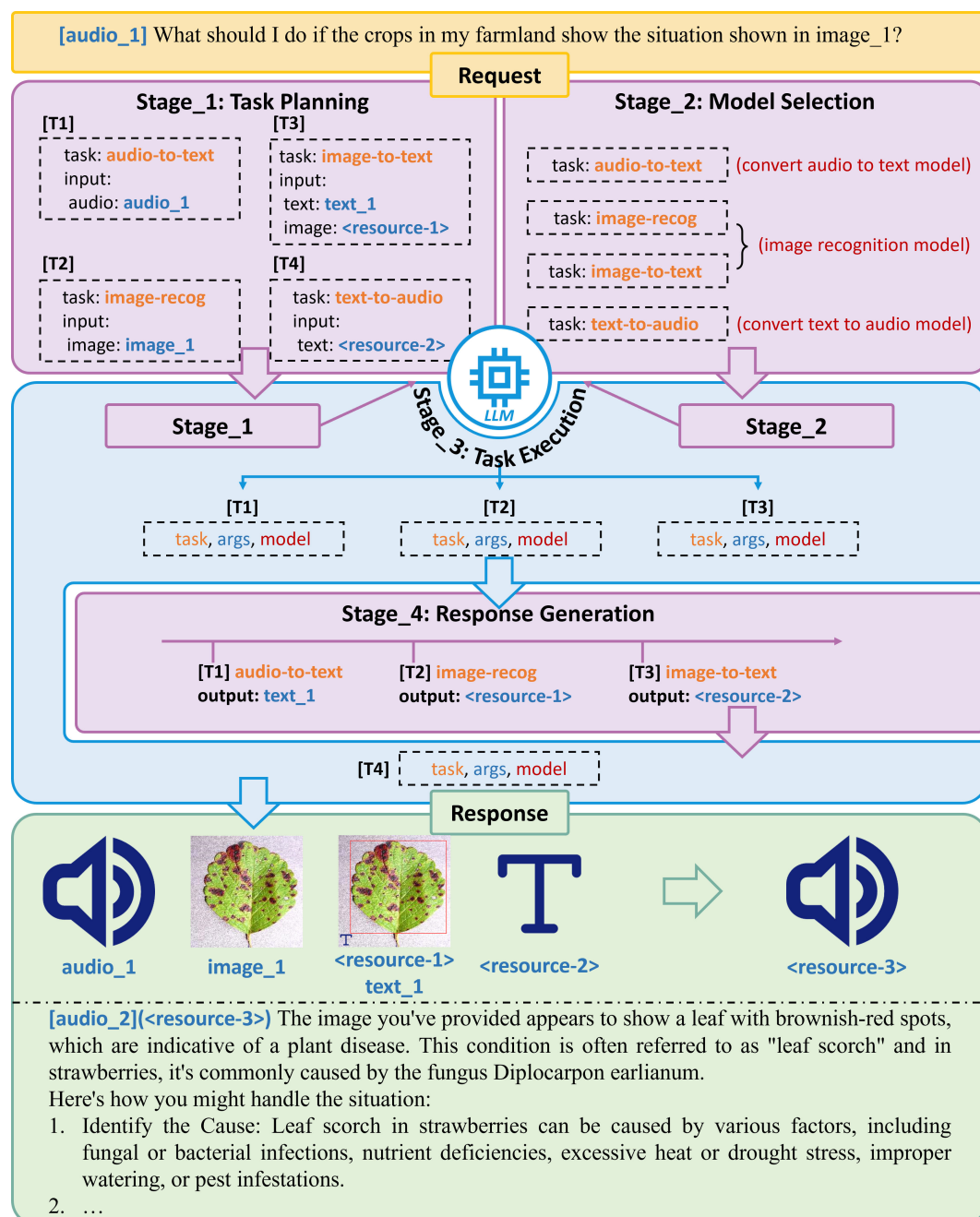


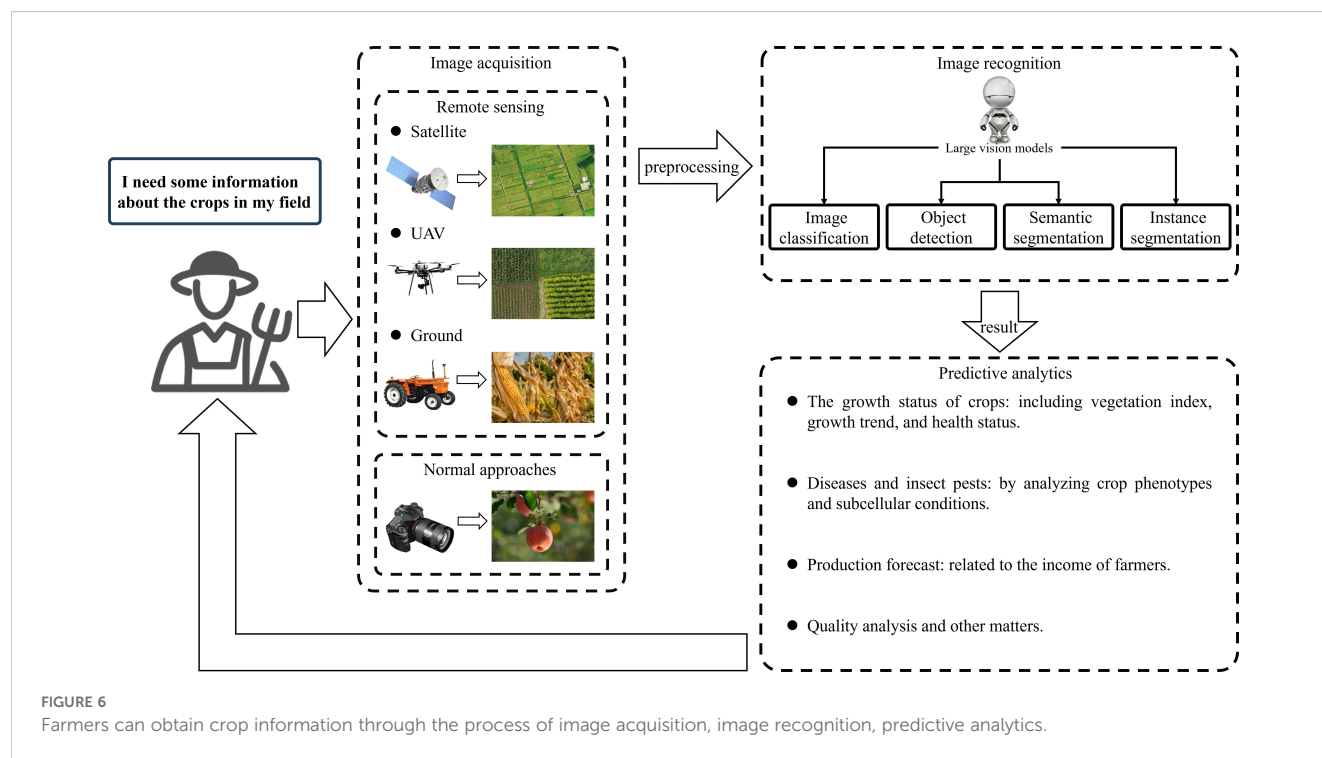
FIGURE 5
An LLM-based AI Agent architecture for agricultural decision-making support.

Nevertheless, it is important to highlight that existing models primarily rely on text-image data and are mostly limited to QA tasks. There is a noticeable lack of applications in the realm of agricultural robotics that incorporate inputs like images, text, voice (Human instructions), and depth information (From LiDAR or laser sensors). These agricultural robots, commonly deployed for tasks such as fruit picking or crop monitoring (Tao and Zhou, 2017), present a significant opportunity for the integration of multimodal data sources to enhance their capabilities. In short,

large models lacking a high degree of multimodality perform fewer tasks and lack good applicability.

3.5 The choice between large models and traditional models

The decision to implement either a large model or a traditional model in agriculture is not straightforward. It involves considering a

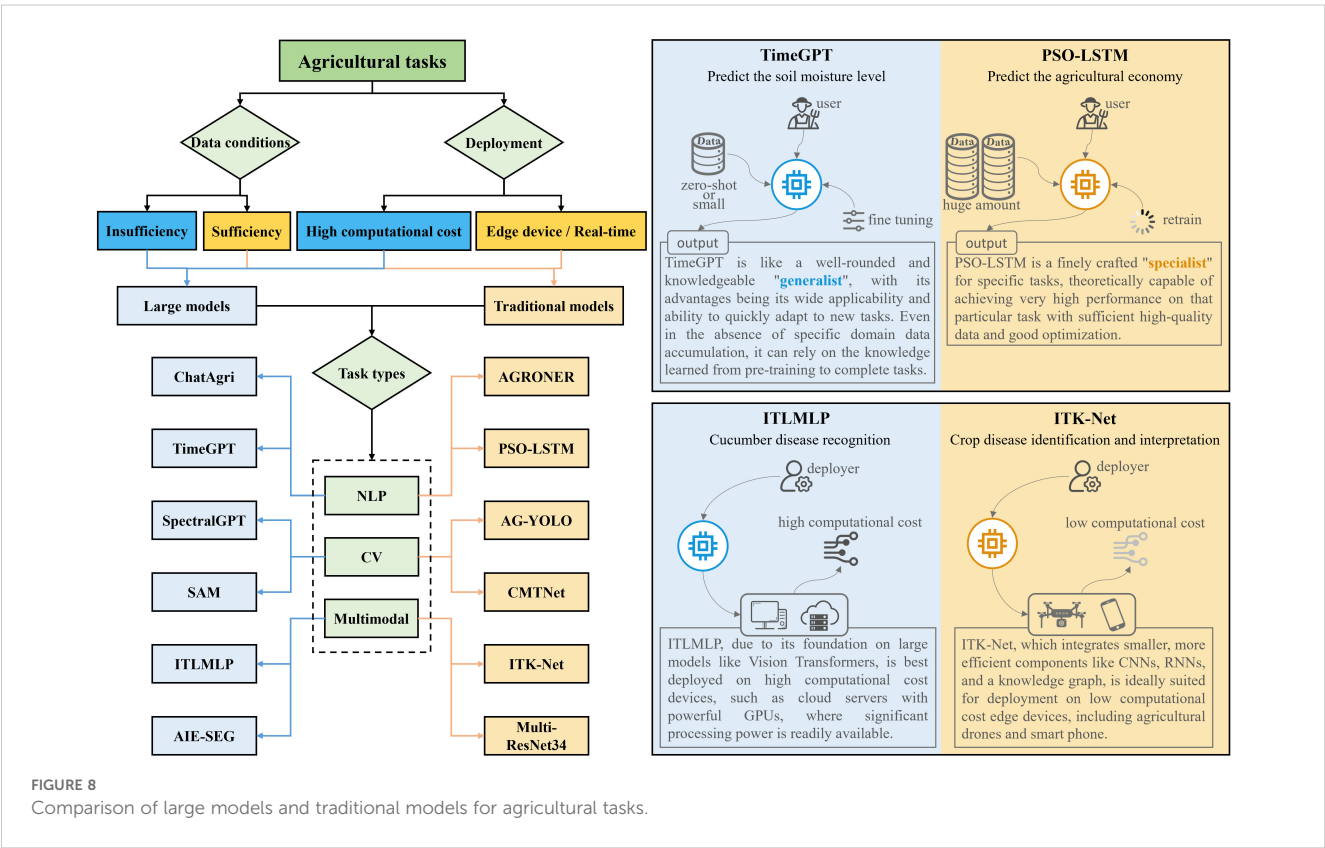
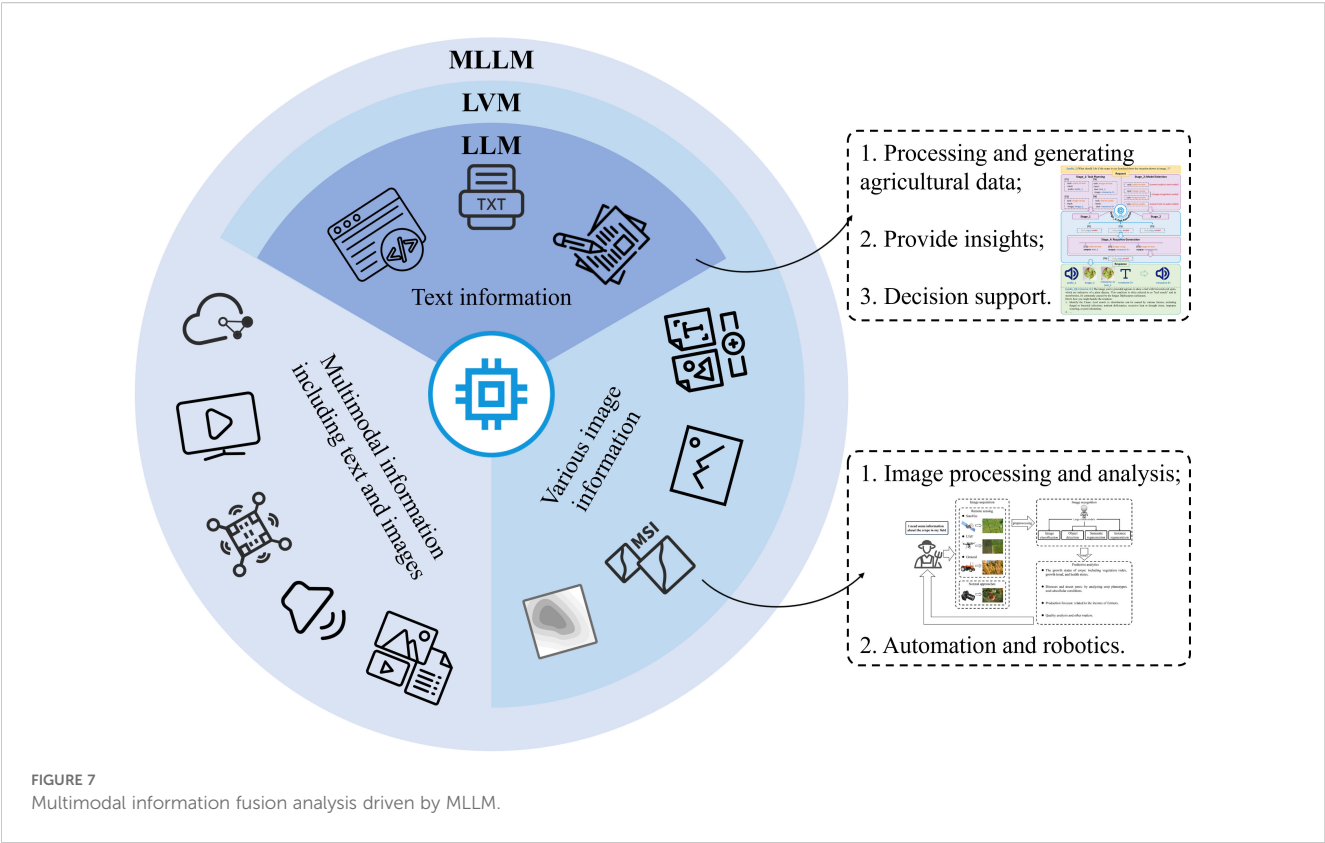


multitude of factors, such as the volume and quality of available data, the required model generalizability, and the practical limits on computational power and inference speed. However, by analyzing the studies of Deforce et al. (2024); Zhao et al. (2023a), and Cao et al. (2023), we found that these considerations can be effectively categorized under two primary factors: Data and deployment conditions. Similar to how large models are divided into LLM, LVM, and MLLM, traditional models can be classified according to the specific agricultural task, falling into the categories of NLP, CV, and multimodal. For instance, models like AGRONER and PSO-LSTM are designed to handle NLP tasks (Veena et al., 2023; Zheng and Li, 2023), AG-YOLO and CMTNet address CV tasks (Lin et al., 2024; Guo et al., 2025b), while ITK-Net and Multi-ResNet34 are tailored for multimodal applications (Zhou et al., 2021; Zhang et al., 2022a). Before selecting a model, it is best to first determine which category the agricultural task belongs to.

When approaching an agricultural task, a critical step is to assess the sufficiency of available data. If a substantial volume of high-quality, task-specific data is available, a traditional model becomes a good option. Conversely, in scenarios marked by data insufficiency, leveraging a large model is often the more suitable choice. Figure 8 presents a comparative analysis of traditional models versus large models based on data conditions and deployment constraints. PSO-LSTM can be retrained on abundant data, and it can deliver superior performance for a particular agricultural task, thus positioning this model as a “specialist”. TimeGPT, on the other hand, functions as a “generalist”, capable of handling diverse, non-specific agricultural tasks using only minimal fine-tuning or a zero-shot approach in data-scarce situations, thereby avoiding the need for complete model retraining for each new task (Deforce et al., 2024). The pre-embedded

knowledge within large models can effectively compensate for the lack of domain-specific data.

On the other hand, deployment conditions are also a crucial factor in model selection. While devices with high computational capacity can deploy both large models and traditional models, the significant computational and time costs associated with large models make them unsuitable for edge devices and systems requiring real-time response. For an agricultural task that requires the model to be deployed on an edge device with real-time detection needs, ITK-Net is the pragmatic and superior choice due to its efficiency and low resource requirements (Zhou et al., 2021). While the ITLMLP model proposed by Cao et al. (2023) also targets crop disease recognition, it is suited for deployment only on devices that can handle high computational costs. As a large model, ITLMLP’s deployment conditions are considerably more stringent than those of ITK-Net, the traditional model. However, this does not imply that ITLMLP is without its merits. Its value lies not in real-time field deployment, but in its powerful offline analysis capabilities. By batch-processing vast agricultural data stored on cloud servers, it can perform in-depth retrospective diagnostics and trend analysis. Leveraging its powerful feature extraction and generalization capabilities, acquired from pre-training on large-scale data, ITLMLP can conduct reclassification of historical disease data, compile statistics on disease occurrence frequencies across different periods, and uncover potential correlations between image features and specific environmental descriptions. By the way, by optimizing the model architecture, using efficient inference algorithms, and utilizing hardware acceleration techniques, the real-time performance of LVMs can be improved to a certain extent (Chen et al., 2023b). In addition, we have also



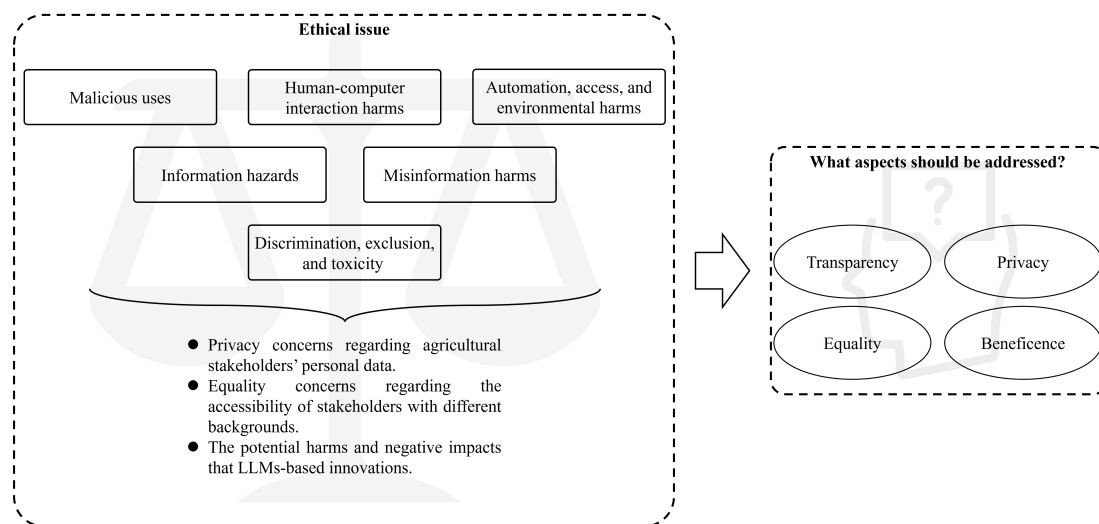


FIGURE 9

The ethical issues faced by large models.

discovered that ITLMLP could process a large dataset to generate highly accurate annotated labels, which can then be used to train smaller, more efficient models like ITK-Net. This creates a synergistic ecosystem where the power of large models enables the effectiveness of traditional models on the edge.

The choice between a large model and a traditional model for agricultural tasks is not a matter of one being definitively superior to the other, but rather a strategic decision based on a careful evaluation of trade-offs. Large models, with their powerful generalization capabilities, offer a robust solution for data-scarce environments, while traditional models excel in data-rich scenarios where their specialized nature can be fully leveraged. Similarly, the high computational cost of large models makes them suitable for offline, server-based analysis, whereas the efficiency of traditional models is indispensable for real-time, on-device deployment.

4 Ethical issues and responsible use of large vision and language models in agriculture

As large models demonstrate their powerful potential and are increasingly applied to agricultural tasks (referencing section 3), it is crucial to critically examine the ethical and societal implications of their deployment. However, there are often ethical and responsibility issues in the development and deployment process of AI today. The digital gap between those who have the resources to develop and utilize large models and those who cannot afford to do so creates an inequality in accessing large models, resulting in an unfair distribution of risks and benefits (Harfouche et al., 2024). Not only that, this divide can be exacerbated by the presence of AI biases (Dara et al., 2022; Ryan, 2023). Accordingly, to ensure ethical issues and responsible use of large models, this chapter starts from the

ethical and responsibility issues in the agricultural large models and explore corresponding measures.

4.1 Ethical considerations in the deployment of large models in agriculture

Predicting and solving ethics problems of large models in agriculture is a critical scientific and societal challenge. Although large models point the way for the future of smart agriculture, due to their characteristic of being influenced by close association, large models often learn some bad knowledge in addition to useful knowledge. Ethical issues have always been an indispensable topic of discussion in the process of technological progress (Such as the ethical issues discussed by Holmes et al. in the field of education regarding educational AI (Holmes et al., 2022)), and we also need to pay attention to ethics issues when using large models in the agricultural direction. As mentioned below, many relevant institutions and personnel have put forward their own ideas on ethical issues related to large models.

Weidinger et al. (2021) put forward six types of ethical risks (Figure 9): 1) Malicious uses, 2) Human-computer interaction harms, 3) Automation, access, and environmental harms, 4) Information hazards, 5) Misinformation harms, and 6) Discrimination, exclusion, and toxicity. Understanding these issues can help us responsibly use large models in the agricultural field.

1. Malicious uses. Prior to the release of GPT-4, OpenAI hired a team of 50 experts and scholars to conduct a six-month adversarial test on GPT-4. Andrew White, a professor of chemical engineering at the University of Rochester who was invited to participate in the test, stated that early versions of GPT-4 could assist in the manufacture of

chemical weapons and even provide a convenient manufacturing location. From the perspective of the agricultural sector, if this issue is not properly addressed, some may use large models to learn ways to destroy other people's farmland for the sake of profit, thereby allowing themselves to have a larger market. Over time, this will lead to vicious competition in the market.

2. Human-computer interaction harms. The potential harms of human-computer interaction arise when users excessively trust a large model or mistakenly treat it as human.
3. Automation, access, and environmental harms. The large model can give rise to automation, access, and environmental harms due to its potential environmental or downstream economic impacts.
4. Information hazards. Due to the involvement of information from different countries, religions, and ethnicities, model outputs leaking or inferring sensitive information often led to political violence.
5. Misinformation harms. A study discussed the potential risks of using poorly performing large models. The original intention of this study was to provide a natural language generation model in MOOC to respond to students and improve their participation rate (Li and Xing, 2021). Even so, due to the poor performance of the model, the corresponding negative results further reduced the enthusiasm of students. If a poorly performing large model is used in the agricultural field, it may mislead farmers in their judgment (Such as analysing incorrect disease types), not only causing further damage to crops in the farmland, but also making farmers increasingly distrust the large model. For this phenomenon, Angelone et al. proposed that warning labels can be applied to the content generated by the large model (Angelone et al., 2022), but this also involves the trust issue of the large model in its own generated results.
6. Discrimination, exclusion, and toxicity. Two researches have indicated that potential discrimination, exclusion, and toxicity issues may occur if adopting a model that is accurate but unfair (Sha et al., 2021; Merine and Purkayastha, 2022).

Despite Weidinger et al.'s viewpoint can provide us with a fundamental understanding of the risks associated with large models, manners of systematic ethical supervision of large models' research and innovation (R&I) are especially restricted. Coincidentally, the European Commission has officially approved comprehensive "ethics guidelines for trustworthy AI" specifically designed for R&I. These guidelines require that principal investigators recognize and tackle the ethical matters raised by their proposed research. Principal investigators are also required to adhere to ethical principles and relevant legislation in their work. In a similar vein, Stanford University's Ethics and Society Review necessitates researchers to distinguish potential societal hazards

associated with their research and incorporate mitigation measures into their research design (Bernstein et al., 2021).

Furthermore, projects with large models have a vast amount of data and often raise ethics issues. For instance, while raw plant science data itself may not inherently fall within the scope of the European Union General Data Protection Regulation (GDPR) as personal data, it can become subject to GDPR regulations when linked to identifiable individuals or specific farm locations tied to individuals, creating complex challenges concerning data ownership and privacy protection (Harfouche et al., 2024). Thus, relevant guidelines must consider code of conduct for data sharing, privacy protection, and the overall governance of datasets.

4.2 Responsible use in agriculture

With the expanding development and utilization of large models, there is a growing recognition of the need for agile and effective regulatory oversight. To address this issue, it may be necessary to use AI technology to assist in overseeing the development and deployment of large models. Regarding this aspect, the AI Act, which has been jointly agreed upon by the European Parliament and the Council of Europe, represents the first comprehensive set of harmonized rules on a global scale. It promotes responsible large model designment and development by regulating large model across various applications and contexts based on a risk-based framework. Within the framework, careful consideration must be given to the level of risk involved and how to evaluate different large models as risk-free or low-risk.

To evaluate the risk level of a large model, we focus on four aspects: transparency, privacy, equality, and beneficence. On the other hand, in addition to developing and adhere to a strong regulatory framework that guides the development, deployment, and use of large models, regulatory methods also need to be considered. Consider the potential societal impact, potential harms, and long-term implications of the technology. Firstly, due to the wide applicability of large models, we cannot make a one size fits all approach. Regulation must adapt to specific issues in different domains. The United States' food and drug administration (FDA) has tailored potential regulatory methods for AI and ML technologies used in medical devices, categorizing them into three major categories based on risk levels: Class I (Low risk), Class II (Moderate risk), and Class III (High risk). Large models in agriculture can also be regulated according to the FDA's approach, dividing them into several types of models ranging from low risk to high risk. For example, genetically modified crops may have environmental impacts, food safety issues, and ecosystem damage, so large models targeting genetically modified crops should be included in high-risk types. For large models of ordinary crops, they can be classified as low-risk types. And the regulatory methods proposed by relevant departments should be made public to ensure transparency of information. Regulators can promote fairness in the deployment of agricultural large models by enforcing the use of diverse and representative data sources, which

helps mitigate potential biases present in the training data (Meskó and Topol, 2023).

From the perspective of beneficence and privacy, privacy issues related to large models have received little attention or investigation in reviewed research (Yan et al., 2024). Specifically, if the training set used to train a large model contains some personal privacy information that has not been authorized by the information owner. The disregard for privacy concerns is especially worrisome considering that LLM-based innovations involve stakeholders' natural languages, which may contain personal and sensitive information pertaining to their private lives and identities (Brown et al., 2022). If users unintentionally learn about this information while using a large model, it may cause harm to the beneficence of the information owner. Developers of large models should ensure they gain explicit consent from individuals before collecting and utilizing these personal data. Clearly communicate the purpose and scope of data usage, and offer individuals with the choice to choose out or request data deletion. Besides, limit the amount of personal and sensitive data collected and stored. Follow the principle of data minimization, ensuring that only necessary data is collected and retained. Anonymize or aggregate data whenever possible to protect individual privacy.

In general, governance approaches that promote responsible utilization of large models and focus on the outcomes rather than the technology itself will enhance research efforts and drive more innovation. By combining governance and ethics, we can harness the powerful synergy to expedite the implementation of large models in agriculture and other domains, fostering innovation at a larger scale.

5 Challenges and future directions

Although large models can play a powerful role in the field of agriculture, they still face challenges in many aspects.

5.1 Technical and practical challenges

5.1.1 Difficulty in obtaining agricultural data

A primary and recurring obstacle highlighted throughout this review is the acquisition of suitable agricultural data. While large models' data generation capabilities can partially alleviate this, as discussed in section 3.2.1, several fundamental difficulties persist:

1. **Cost and quality:** Acquiring comprehensive, high-quality, and accurately labeled real-world data is a time-consuming, labor-intensive, and costly process, especially for supervised learning approaches (Li et al., 2023a; Lu and Young, 2020).
2. **Privacy and trust:** As mentioned in section 4.2, the private nature of farmland data raises significant privacy and trust concerns among farmers, often leading to a reluctance to share information crucial for model training.

3. **Temporal complexity:** Agricultural data is inherently temporal. The need to capture entire crop growth cycles, which are influenced by daily, seasonal, and annual variations, adds another layer of complexity to data collection efforts (Li et al., 2023b).

5.1.2 Low training efficiency

Directly related to the need for massive datasets is the challenge of low training efficiency and high computational cost. As systematically compared against traditional models in section 3.5 (Figure 8), training large agricultural models is a resource-intensive endeavor. Their massive parameter counts demand significant computational power and lengthy training times, often measured in thousands of GPU hours (Li et al., 2023b). This stands in stark contrast to the efficiency of traditional models like YOLO and Faster R-CNN, whose lower computational requirements make them a more practical and cost-effective solution for many specific, real-time agricultural tasks (Badgujar et al., 2024). This efficiency gap explains the continued prevalence of traditional models despite the emergence of more powerful large-scale architectures.

5.1.3 Distribution shift

The problem of distribution shift is a major challenge when using large models in agriculture. When the data encountered by the model during deployment is obviously different from the data used in its training phase, a distribution shift will occur. The environmental conditions for collecting data may vary greatly in different regions and climates. These changes may include differences in crop types, soil conditions, weather patterns, and agricultural practices, all of which can lead to significant changes in data distribution (Wiles et al., 2021). The distribution shift will result in the trained large model not having strong applicability and may not achieve good results in some agricultural tasks. For example, it has been proven that applying large models directly to leaf segmentation tasks in a zero-shot means led to unsatisfactory performance, which can be attributed to possible distribution shifts (Chawla et al., 2021).

5.1.4 The lag of data

After the trained large model is put into use, the data used for training has a certain timeliness for a long period of time. But after a long time, some data lags in time, and the results obtained by using a large model may deviate from the current facts (Figure 10).

5.1.5 Query formulation impacts model output

The results obtained from large models can vary significantly depending on how the query is formulated. Like Figure 11, when multiple images are spliced together for questioning, GPT-4 provides ambiguous answers; When only asking for one image, GPT-4 provides a clearer answer.

To clear these obstacles, future research and development work needs to pay attention to model optimization techniques such as model compression and efficient network structure design, reducing model size without affecting performance (Zhong et al., 2023). It is also

necessary to provide update and maintenance functions for the model to ensure its timeliness. Developers need to write relevant usage instructions to help users get started quickly. Notably, emerging frameworks like RAG offer a direct solution to the data lag and accuracy challenges by connecting LLMs to real-time, external knowledge bases. Similarly, developing more sophisticated AI Agents capable of autonomous planning and tool use will be crucial for creating robust and adaptable agricultural systems.

5.2 Infrastructure and cost barriers

Applying large models to rural areas faces significant barriers related to poor connectivity and high implementation costs. These limitations disproportionately affect small-scale farmers and regions with underdeveloped infrastructure, exacerbating existing inequalities in agricultural productivity and technological access (Da Silva et al., 2023).

Dibbern et al. (2024) found that farmers often abandon digital tools due to unreliable broadband or mobile connectivity, even when initial investments are made. Technologies like IoT, cloud-based analytics, and real-time monitoring systems remain underutilized in areas lacking stable network access. This has brought some warnings for the application of large models in rural areas. In addition, the high cost of agricultural machinery using large models—render them inaccessible to resource-limited farmers. For example, autonomous machinery and AI platforms often require upfront investments exceeding \$10,000 USD, a prohibitive sum for smallholders (Bolfé et al., 2020).

To overcome poor connectivity, investing in and expanding rural broadband and mobile infrastructure is crucial, potentially through government subsidies, public-private partnerships, and the exploration of alternative network solutions like satellite internet or mesh networks tailored to agricultural regions. To mitigate high implementation costs, promoting the development of affordable, modular agricultural machinery and large model platforms designed specifically for smallholder farmers is essential. In short, bridging the digital divide and promoting inclusive technological progress requires joint efforts among technology developers, agricultural researchers, policy makers, and local farmer organizations.

5.3 Future trends in the integration of agricultural and food sectors and large models

In the future, there will undoubtedly be agricultural large models with better performance and higher applicability. And the large models in agriculture should not be limited to text and image inputs. We believe that future multimodal agricultural models can support multimodal information such as videos (Analyzing crops in videos) and audio (Tapping watermelons, and judging maturity through the sound emitted). On the other hand, agriculture is closely related to food, and the development of large models in agriculture is likely to promote the development of large models in the food domain. Trust is indispensable for agriculture and food system technologies given food's universality and importance to people (Tzachor et al., 2022). Researchers need to navigate

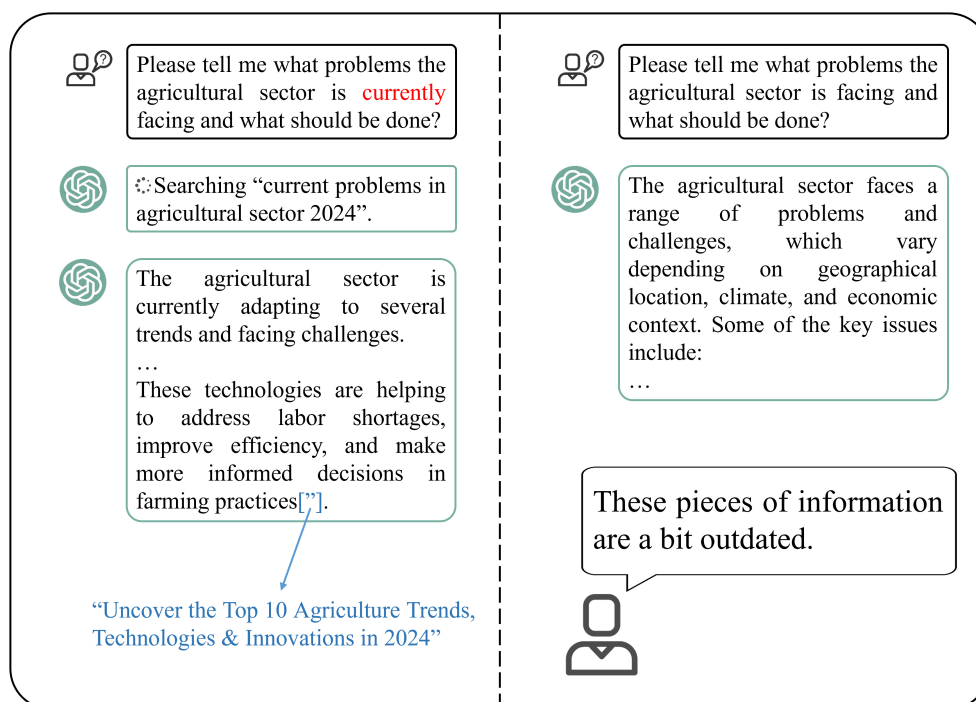


FIGURE 10
The lag of data.

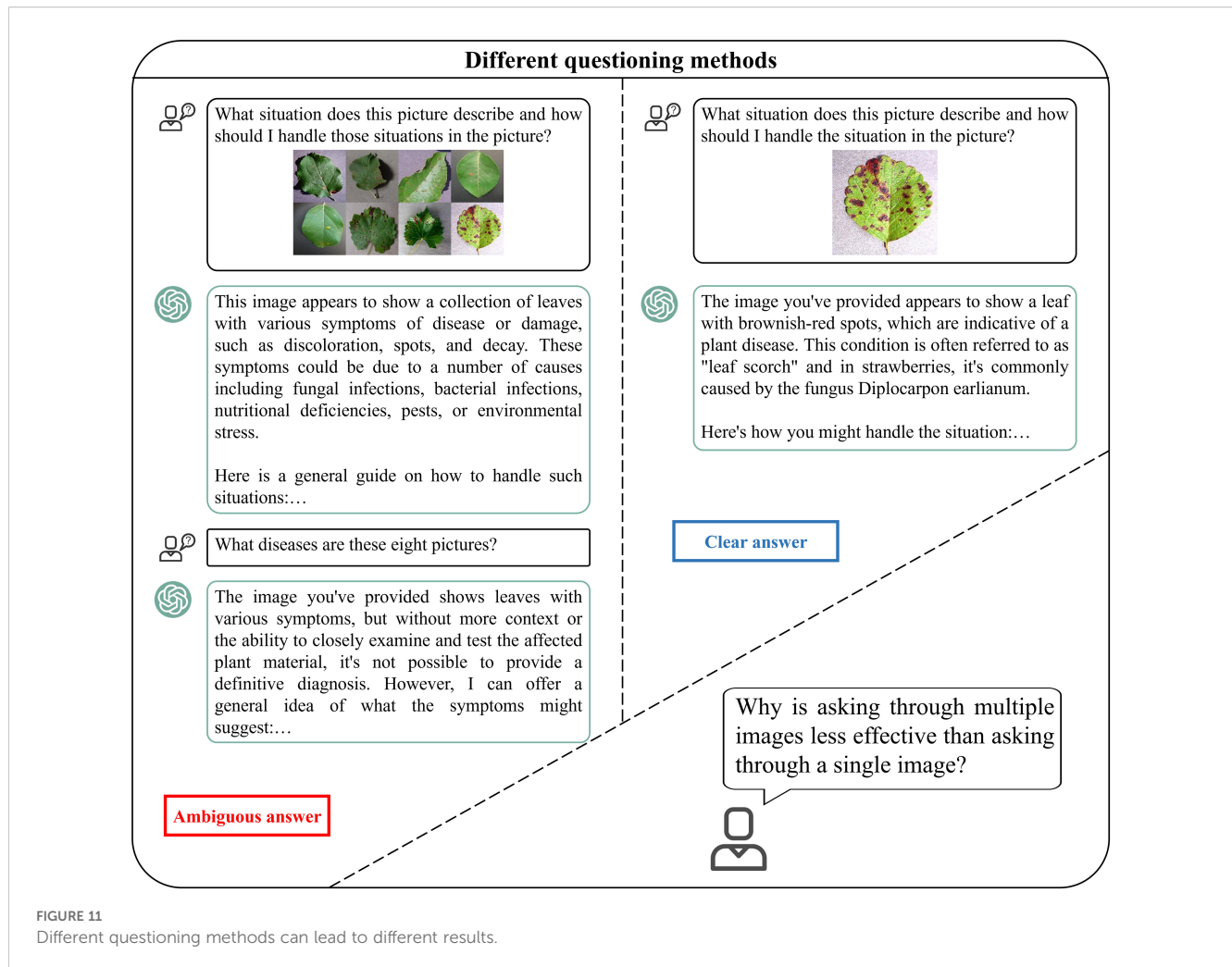


FIGURE 11
Different questioning methods can lead to different results.

complicated social, political, economic, and environmental landscapes to develop appropriate large models in the food industry. In the future food industry, researchers will strive to establish trust with governmental agencies and funders, as well as with food system partners, to provide food and products that the public trusts (Alexander et al., 2024).

Overall, although the agricultural large model still faces many challenges at present, we believe that through the joint efforts of relevant researchers in the future, these challenges can be properly addressed. And due to the close relationship between the food and agricultural domains, with the gradual development of agricultural large models, food large models will also receive further research, thereby achieving mutual positive feedback between the development of large models in these two fields.

6 Conclusion

In summary, this study investigated the application status of large models in the agricultural field. Our analysis establishes that these models offer unprecedented advantages through their capacity for complex reasoning, multimodal information processing, and the

execution of nuanced tasks ranging from pest identification to robotic automation. We further determined that the efficacy of these powerful tools is significantly amplified when they are tailored to the agricultural domain, a crucial strategy for overcoming the pervasive challenge of limited labeled data. Furthermore, this review provided a pragmatic framework for choosing between large and traditional models, emphasizing that the decision hinges on a careful trade-off between data availability and deployment constraints. While large models excel as “generalists” in data-scarce or offline analytical scenarios, efficient traditional models remain indispensable as “specialists” for real-time, on-device tasks.

However, this vast potential is tempered by critical, interconnected challenges that must be addressed. A primary hurdle is the acquisition and utilization of suitable agricultural data; issues of data scarcity, high collection costs, inherent data diversity (across crops, regions, conditions), privacy concerns associated with farmland data, and the need for time-series information create significant obstacles. Furthermore, the high computational resources required for training and deploying large models, coupled with the often-limited internet connectivity and financial resources in rural areas, creates a significant digital divide, potentially excluding smallholder farmers. Technical issues such as model susceptibility to

distribution shifts between training and deployment environments, the problem of data lag impacting real-time relevance, and sensitivity to query formulation also impact the reliability and practical applicability of current models. Finally, overarching ethical considerations, including potential biases in data or algorithms, ensuring data privacy, promoting equitable access to technology, and preventing misuse, are paramount and demand careful consideration and robust governance frameworks.

Although our study is comprehensive, there are inherent limitations to studying a rapidly developing field. To move forward, future research must directly confront the limitations and challenges identified. Developing novel techniques to mitigate data scarcity—such as advanced data augmentation and self-supervised learning tailored for agriculture—is a critical priority. Expanding multimodal capabilities to robustly incorporate inputs like video, audio, and diverse sensor data will unlock new frontiers in precision farming. Crucially, research must move beyond theoretical ethics to the practical implementation of governance structures for AI in agriculture. Furthermore, a significant opportunity lies in exploring the synergistic relationship between agricultural large models and the broader food system, addressing challenges from farm to fork.

Large models stand poised to be transformative technologies for agriculture. While significant challenges remain, the potential benefits for productivity, sustainability, and food security are immense. Addressing the technical hurdles, bridging the digital divide, and navigating the ethical landscape through collaborative, responsible innovation will be key to realizing this potential. We hope this article serves as a valuable resource and a cornerstone, stimulating further research and guiding the development of future agricultural large models that are not only powerful but also practical, efficient, and beneficial for all stakeholders in the global food system.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

Author contributions

HZ: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing. SQ: Conceptualization, Formal Analysis, Investigation, Methodology,

Resources, Software, Visualization, Writing – original draft, Writing – review & editing. MS: Funding acquisition, Writing – review & editing. CL: Software, Writing – original draft. AL: Software, Writing – original draft. JG: Conceptualization, Investigation, Project administration, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. Our research was supported by the Natural Science Foundation of Guangxi (No. 2024 GXNSFBA010381), the National Natural Science Foundation of China (No. 62361006), Guangxi Young Elite Scientist Sponsorship Program (GXYESS2025081), and the grant (No. NCOC-24-02) from Key Laboratory of Nonlinear Circuit and Optical Communications (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1579355/full#supplementary-material>

References

- Ahirwar, S., Swarnkar, R., Bhukya, S., and Namwade, G. (2019). Application of drone in agriculture. *Int. J. Curr. Microbiol. Appl. Sci.* 8, 2500–2505. doi: 10.20546/ijcmas.2019.801.264
- Alexander, C. S., Yarborough, M., and Smith, A. (2024). Who is responsible for 'responsible AI': Navigating challenges to build trust in AI agriculture and food system technology. *Precis. Agric.* 25, 146–185. doi: 10.1007/s11119-023-10063-3
- Angelone, A. M., Galassi, A., and Vittorini, P. (2022). "Improved automated classification of sentences in data science exercises," in *In Methodologies and Intelligent Systems for Technology Enhanced Learning, 11th International Conference*, Vol. 11. 12–21. doi: 10.1007/978-3-030-86618-1_2
- Anil, R., Dai, A. M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., et al. (2023). Palm 2 technical report. doi: 10.48550/arXiv.2305.10403
- Aymen, F., Monir, H., and Pester, A. (2024). "Large Vision Models: How Transformer-based Models excelled over Traditional Deep Learning Architectures in Video Processing," in *2024 5th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*. 50–54. doi: 10.1109/AIRC61399.2024.10672087
- Badgujar, C. M., Poullose, A., and Gan, H. (2024). Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. *Comput. Electron. Agric.* 223, 109090. doi: 10.1016/j.compag.2024.109090

- Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., et al. (2023). Qwen-vl: A frontier large vision-language model with versatile abilities. doi: 10.48550/arXiv.2308.12966
- Bender, A., Whelan, B., and Sukkarieh, S. (2020). A high-resolution, multimodal data set for agricultural robotics: A Ladybird's-eye view of Brassica. *J. Field Robot.* 37, 73–96. doi: 10.1002/rob.21877
- Bengio, Y., Ducharme, R., and Vincent, P. (2000). A neural probabilistic language model. *Adv. Neural Inf. Process. Syst.* 13.
- Bernstein, M. S., Levi, M., Magnus, D., Rajala, B. A., Satz, D., and Waeiss, Q. (2021). Ethics and society review: Ethics reflection as a precondition to research funding. *Proc. Natl. Acad. Sci.* 118, e2117261118. doi: 10.1073/pnas.2117261118
- Bi, Z., Zhang, N., Xue, Y., Ou, Y., Ji, D., Zheng, G., et al. (2023). OceanGPT: A large language model for ocean science tasks. doi: 10.48550/arXiv.2310.02031
- Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., et al. (2024). Aurora: A foundation model of the atmosphere. doi: 10.48550/arXiv.2405.13063
- Bolfe, ÉL, Jorge, L. A. C., Sanches, I. D., Luchiarini Júnior, A., da Costa, C. C., Victoria, D. C., et al. (2020). Precision and digital agriculture: adoption of technologies and perception of Brazilian farmers. *Agriculture* 10, 653. doi: 10.3390/agriculture10120653
- Bommasani, R., Hudson, D. A., Adeli, E., Adeli, E., Altman, R., Arora, S., et al. (2021). On the opportunities and risks of foundation models. doi: 10.48550/arXiv.2108.07258
- Bosilj, P., Aptoula, E., Duckett, T., and Cielniak, G. (2020). Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *J. Field Robot.* 37, 7–19. doi: 10.1002/rob.21869
- Bouguetaya, A., Zarzour, H., Kechida, A., and Taberkit, A. M. (2022). Deep learning techniques to classify agricultural crops through UAV imagery: A review. *Neural Computing Appl.* 34, 9511–9536. doi: 10.1007/s00521-022-07104-9
- Brown, H., Lee, K., Mireshghallah, F., Shokri, R., and Tramèr, F. (2022). “What does it mean for a language model to preserve privacy?”, in *In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2280–2292. doi: 10.1145/3531146.3534642
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. doi: 10.48550/arXiv.2303.12712
- Cao, Y., Chen, L., Yuan, Y., and Sun, G. (2023). Cucumber disease recognition with small samples using image-text-label-based multi-modal language model. *Comput. Electron. Agric.* 211, 107993. doi: 10.1016/j.compag.2023.107993
- Chawla, S., Singh, N., and Drori, I. (2021). “Quantifying and alleviating distribution shifts in foundation models on review classification,” in *NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications*. <https://openreview.net/forum?id=OG78-TuPvL>.
- Chelba, C., Mikolov, T., Schuster, M., Ge, Q., Brants, T., Koehn, P., et al. (2013). One billion word benchmark for measuring progress in statistical language modeling. doi: 10.48550/arXiv.1312.3005
- Chen, C., Du, Y., Fang, Z., Wang, Z., Luo, F., Li, P., et al. (2024). Model composition for multimodal large language models. doi: 10.48550/arXiv.2402.12750
- Chen, F., Giuffrida, M. V., and Tsafaris, S. A. (2023a). “Adapting vision foundation models for plant phenotyping,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 604–613.
- Chen, L., Zaharia, M., and Zou, J. (2023b). FrugalGPT: How to use large language models while reducing cost and improving performance. doi: 10.48550/arXiv.2305.05176
- Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P.D.O., Kaplan, J., et al. (2021). Evaluating large language models trained on code. doi: 10.48550/arXiv.2107.03374
- Chin, R., Catal, C., and Kassahun, A. (2023). Plant disease detection using drones in precision agriculture. *Precis. Agric.* 24, 1663–1682. doi: 10.1007/s11119-023-10014-y
- Dai, H., Liu, Z., Liao, W., Huang, X., Cao, Y., Wu, Z., et al. (2023). AugGPT: Leveraging chatGPT for text data augmentation. doi: 10.48550/arXiv.2302.13007
- Dara, R., Hazrati Fard, S. M., and Kaur, J. (2022). Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. *Front. Artif. Intell.* 5. doi: 10.3389/frai.2022.884192
- Das, D., Banerjee, D., Aditya, S., and Kulkarni, A. (2024). MATHSENSEI: a tool-augmented large language model for mathematical reasoning. doi: 10.48550/arXiv.2402.17231
- Da Silva, F., Da Silva, S. L. C., MaChado, F. M., Barbedo, J. G.A., and Amaral, F. G. (2023). Farmers' perception of the barriers that hinder the implementation of agriculture 4.0. *Agric. Syst.* 208, 103656. doi: 10.1016/j.agsy.2023.103656
- Deforce, B., Baesens, B., and Asensio, E. S. (2024). Time-series foundation models for forecasting soil moisture levels in smart agriculture. *arXiv preprint arXiv:2405.18913*. doi: 10.48550/arXiv.2405.18913
- Dibbern, T., Romani, L. A. S., and Massruhá, S. M. F. S. (2024). Main drivers and barriers to the adoption of Digital Agriculture technologies. *Smart Agric. Technol.* 8, 100459. doi: 10.1016/j.atech.2024.100459
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. doi: 10.48550/arXiv.2010.11929
- Driess, D., Xia, F., Sajjadi, M. S., Lynch, C., Chowdhery, A., Wahid, A., et al. (2023). Palm-e: An embodied multimodal language model. doi: 10.48550/arXiv.2303.03378
- ElMasry, G., Mandour, N., Al-Rejaie, S., Belin, E., and Rousseau, D. (2019). Recent applications of multispectral imaging in seed phenotyping and quality monitoring—An overview. *Sensors* 19, 1090. doi: 10.3390/s19051090
- Feng, X., Yu, Z., Fang, H., Jiang, H., Yang, G., Chen, L., et al. (2022). Plantorgan hunter: a deep learning-based framework for quantitative profiling plant subcellular morphology. doi: 10.21203/rs.3.rs-1811819/v1
- Ganeshkumar, C., David, A., Sankar, J. G., and Saginala, M. (2023). “Application of drone Technology in Agriculture: A predictive forecasting of Pest and disease incidence,” in *Applying drone technologies and robotics for agricultural sustainability*, IGI Global, 50–81. doi: 10.4018/978-1-6684-6413-7.ch004
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., et al. (2023). Retrieval-augmented generation for large language models: A survey. doi: 10.48550/arXiv.2312.10997
- Geitmann, A., and Bidhendi, A. J. (2023). Plant blindness and diversity in AI language models. *Trends Plant Sci.* 28, 1095–1097. doi: 10.1016/j.tplants.2023.06.016
- Girdhar, R., El-Nouby, A., Liu, Z., Singh, M., Alwala, K. V., Joulin, A., et al. (2023). “Imagebind: One embedding space to bind them all,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15180–15190. doi: 10.1109/CVPR52729.2023.01457
- Goertzel, B. (2014). Artificial general intelligence: concept, state of the art, and future prospects. *J. Artif. Gen. Intell.* 5, 1–48. doi: 10.2478/jagi-2014-0001
- Guo, X., Feng, Q., and Guo, F. (2025b). CMTNet: a hybrid CNN-transformer network for UAV-based hyperspectral crop classification in precision agriculture. *Sci. Rep.* 15, 12383. doi: 10.1038/s41598-025-97052-w
- Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., et al. (2025a). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. doi: 10.48550/arXiv.2501.12948
- Hamuda, E., Glavin, M., and Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* 125, 184–199. doi: 10.1016/j.compag.2016.04.024
- Harfouche, A. L., Petousi, V., and Jung, W. (2024). AI ethics on the road to responsible AI plant science and societal welfare. *Trends Plant Sci.* 29, 104–107. doi: 10.1016/j.tplants.2023.12.016
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*. 2961–2969. doi: 10.48550/arXiv.1703.06870
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778. doi: 10.1109/CVPR.2016.90
- Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., et al. (2022). Imagen video: High definition video generation with diffusion models. doi: 10.48550/arXiv.2210.02303
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., et al. (2022). Ethics of AI in education: Towards a community-wide framework. *Int. J. Artif. Intell. Educ.* 32 (3), 504–526. doi: 10.1007/s40593-021-00239-1
- Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip. Rev.: Data Min. Knowledge Discov.* 9, e1312. doi: 10.1002/widm.1312
- Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., et al. (2024). SpectralGPT: Spectral remote sensing foundation model. *IEEE Trans. Pattern Anal. Mach. Intelligence.* 46 (8), 5227–5244. doi: 10.1109/TPAMI.2024.3362475
- Karthikeyan, L., Chawla, I., and Mishra, A. K. (2020). A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *J. Hydrol.* 586, 124905. doi: 10.1016/j.jhydrol.2020.124905
- Khanal, S., Kc, K., Fulton, J. P., Shearer, S., and Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sens.* 12, 3783. doi: 10.3390/rs12223783
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023). “Segment anything,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4015–4026. doi: 10.48550/arXiv.2304.02643
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM* 60 (6), 84–90. doi: 10.1145/3065386
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., et al. (2023). Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digital Health* 2, e0000198. doi: 10.1371/journal.pdig.0000198
- Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., et al. (2023). Bloom: A 176b-parameter open-access multilingual language model. doi: 10.48550/arXiv.2211.05100
- Li, J., Chen, D., Qi, X., Li, Z., Huang, Y., Morris, D., et al. (2023a). Label-efficient learning in agriculture: A comprehensive review. *Comput. Electron. Agric.* 215, 108412. doi: 10.1016/j.compag.2023.108412

- Li, J., Lammers, K., Yin, X., Yin, X., He, L., Sheng, J., et al. (2024). MetaFruit meets foundation models: leveraging a comprehensive multi-fruit dataset for advancing agricultural foundation models. doi: 10.48550/arXiv.2407.04711
- Li, C., and Xing, W. (2021). Natural language generation using deep learning to support MOOC learners. *Int. J. Artif. Intell. Educ.* 31, 186–214. doi: 10.1007/s40593-020-00235-x
- Li, J., Xu, M., Xiang, L., Chen, D., Zhuang, W., Yin, X., et al. (2023b). Large language models and foundation models in smart agriculture: Basics, opportunities, and challenges. doi: 10.48550/arXiv.2308.06668
- Li, W., and Zhao, Y. (2015). Bibliometric analysis of global environmental assessment research in a 20-year period. *Environ. Impact Assess. Rev.* 50, 158–166. doi: 10.1016/j.eiar.2014.09.012
- Lin, Y., Huang, Z., Liang, Y., Liu, Y., and Jiang, W. (2024). Ag-yolo: A rapid citrus fruit detection algorithm with global context fusion. *Agriculture* 14, 114. doi: 10.3390/agriculture14010114
- Liu, H., Li, C., Wu, Q., and Lee, Y. J. (2023b). Visual instruction tuning. *Adv. Neural Inf. Process. Syst.* 36, 34892–34916.
- Liu, B., Zhao, R., Liu, J., and Wang, Q. (2023a). *AgriGPTs*. *GitHub*. Available online at: <https://github.com/AgriGPTs/AgriGPTs> (Accessed April 15, 2025).
- Lu, G., Li, S., Mai, G., Sun, J., Zhu, D., Chai, L., et al. (2023). Agi for agriculture. doi: 10.48550/arXiv.2304.06136
- Lu, Y., and Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *Comput. Electron. Agric.* 178, 105760. doi: 10.1016/j.compag.2020.105760
- Malla, A., Omwenga, M. M., and Bera, P. K. (2024). “Exploring image similarity through generative language models: A comparative study of GPT-4 with word embeddings and traditional approaches,” in *2024 IEEE International Conference on Electro Information Technology (EIT)*. 275–279. doi: 10.1109/EIT60633.2024.10609905
- Medeiros, A. D.D., Silva, L. J.D., Ribeiro, J. P.O., Ferreira, K. C., Rosas, J. T.F., Santos, A. A., et al. (2020). Machine learning for seed quality classification: An advanced approach using merger data from FT-NIR spectroscopy and X-ray imaging. *Sensors* 20, 4319. doi: 10.3390/s20154319
- Merine, R., and Purkayastha, S. (2022). “Risks and benefits of AI-generated text summarization for expert level content in graduate health informatics,” in *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. 567–574. doi: 10.1109/ICHI54592.2022.00113
- Mesko, B., and Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digital Med.* 6, 120. doi: 10.1038/s41746-023-00873-0
- Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., and Khudanpur, S. (2010). Recurrent neural network based language model. *Interspeech* 2, 1045–1048. doi: 10.21437/Interspeech.2010-343
- Mostaco, G. M., De Souza, I. R. C., Campos, L. B., Cugnasca, C. E., et al. (2018). “AgronomoBot: a smart answering Chatbot applied to agricultural sensor networks,” in *14th international conference on precision agriculture*, Vol. 24. 1–13.
- Nasir, I. M., Bibi, A., Shah, J. H., Khan, M. A., Sharif, M., Iqbal, K., et al. (2021). Deep learning-based classification of fruit diseases: An application for precision agriculture. *Comput. Mater. Contin.* 66, 1949–1962. doi: 10.32604/cmc.2020.012945
- Niranjan, P. Y., Rajpurohit, V. S., and Malgi, R. (2019). “A survey on chat-bot system for agriculture domain,” in *2019 1st International Conference on Advances in Information Technology (ICAIT)*. 99–103. doi: 10.1109/ICAIT47043.2019.8987429
- Omia, E., Bae, H., Park, E., Kim, M. S., Baek, I., Kabenge, I., et al. (2023). Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances. *Remote Sens.* 15, 354. doi: 10.3390/rs15020354
- Paymode, A. S., and Malode, V. B. (2022). Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. *Artif. Intell. Agric.* 6, 23–33. doi: 10.1016/j.iaia.2021.12.002
- Pazhanivelan, S., Kumaraperumal, R., Shanmugapriya, P., Sudarmanian, N. S., Sivamurugan, A. P., Sathesh, S., et al. (2023). Quantification of biophysical parameters and economic yield in cotton and rice using drone technology. *Agriculture* 13, 1668. doi: 10.3390/agriculture13091668
- Peebles, W., and Xie, S. (2023). “Scalable diffusion models with transformers,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4195–4205. doi: 10.48550/arXiv.2212.09748
- Peng, R., Liu, K., Yang, P., Yuan, Z., and Li, S. (2023). Embedding-based retrieval with llm for effective agriculture information extracting from unstructured data. doi: 10.48550/arXiv.2308.03107
- Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 652–660. doi: 10.1109/CVPR.2017.16
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2023). “Robust speech recognition via large-scale weak supervision,” in *International Conference on Machine Learning*. 28492–28518.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 1–67.
- Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. (2022). Hierarchical text-conditional image generation with clip latents 1, 3. doi: 10.48550/arXiv.2204.06125
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., et al. (2021). “Zero-shot text-to-image generation,” in *International conference on machine learning*. 8821–8831. doi: 10.48550/arXiv.2102.12092
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788. doi: 10.1109/CVPR.2016.91
- Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031
- Ren, Y., Hu, C., Tan, X., Qin, T., Zhao, S., Zhao, Z., et al. (2020). FastSpeech 2: Fast and high-quality end-to-end text to speech. doi: 10.48550/arXiv.2006.04558
- Ren, Y., Ruan, Y., Tan, X., Qin, T., Zhao, S., Zhao, Z., et al. (2019). FastSpeech: Fast, robust and controllable text to speech. *Adv. Neural Inf. Process. Syst.* 32. doi: 10.48550/arXiv.1905.09263
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). “High-resolution image synthesis with latent diffusion models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695. doi: 10.48550/arXiv.2112.10752
- Rose Mary, C. A., Raji Sukumar, A., and Hemalatha, N. (2021). Text based smart answering system in agriculture using RNN. *agriRxiv* 2021, 20210310498. doi: 10.31220/agriRxiv.2021.00071
- Ryan, M. (2023). The social and ethical impacts of artificial intelligence in agriculture: mapping the agricultural AI literature. *AI Soc.* 38, 2473–2485. doi: 10.1007/s00146-021-01377-9
- Saleem, M. H., Potgieter, J., and Arif, K. M. (2021). Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precis. Agric.* 22, 2053–2091. doi: 10.1007/s11119-021-09806-x
- Sha, L., Rakovic, M., Whitelock-Wainwright, A., Carroll, D., Yew, V. M., Gasevic, D., et al. (2021). “Assessing algorithmic fairness in automatic classifiers of educational forum posts,” in *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part I* 22. 381–394. doi: 10.1007/978-3-030-78292-4_31
- Shen, Y., Song, K., Tan, X., Li, D., Lu, W., and Zhuang, Y. (2024). Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face. *Adv. Neural Inf. Process. Syst.* 36. doi: 10.48550/arXiv.2303.17580
- Shi, B., Wu, Z., Mao, M., Wang, X., and Darrell, T. (2024). “When do we not need larger vision models?,” in *European Conference on Computer Vision*. 444–462 (Cham: Springer Nature Switzerland). doi: 10.1007/978-3-031-73242-3_25
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. doi: 10.48550/arXiv.1409.1556
- Stella, F., Della Santina, C., and Hughes, J. (2023). How can LLMs transform the robotic design process? *Nat. Mach. Intell.* 5, 561–564. doi: 10.1038/s42256-023-00669-7
- Sun, C., Shrivastava, A., Singh, S., and Gupta, A. (2017). “Revisiting unreasonable effectiveness of data in deep learning era,” in *Proceedings of the IEEE international conference on computer vision*. 843–852.
- Sundermeyer, M., Schlüter, R., and Ney, H. (2012). Lstm neural networks for language modeling. *Interspeech* 2012, 194–197. doi: 10.21437/Interspeech.2012-65
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9. doi: 10.1109/CVPR.2015.7298594
- Tao, Y., and Zhou, J. (2017). Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking. *Comput. Electron. Agric.* 142, 388–396. doi: 10.1016/j.compag.2017.09.019
- Team, G., Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., et al. (2023). Gemini: a family of highly capable multimodal models. doi: 10.48550/arXiv.2312.11805
- Team, G., Kamath, A., Ferret, J., Pathak, S., Vieillard, N., Merhej, R., et al. (2025). Gemma 3 technical report. doi: 10.48550/arXiv.2503.19786
- Team, G., Mesnard, T., Hardin, C., Dadashi, R., Bhupatiraju, S., Pathak, S., et al. (2024). Gemma: Open models based on gemini research and technology. doi: 10.48550/arXiv.2403.08295
- Thenmozhi, K., and Reddy, U. S. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. *Comput. Electron. Agric.* 164, 104906. doi: 10.1016/j.compag.2019.104906
- Tokekar, P., Vander Hook, J., Mulla, D., and Isler, V. (2016). Sensor planning for a symbiotic UAV and UGV system for precision agriculture. *IEEE Trans. Robotics* 32, 1498–1511. doi: 10.1109/TRO.2016.2603528
- Tripathy, P., Baylis, K., Wu, K., Watson, J., and Jiang, R. (2024). Investigating the segment anything foundation model for mapping smallholder agriculture field boundaries without training labels. doi: 10.48550/arXiv.2407.01846
- Tzachor, A., Devare, M., King, B., Avin, S., and Ó hÉigeartaigh, S. (2022). Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities. *Nat. Mach. Intell.* 4, 104–109. doi: 10.1038/s42256-022-00440-4

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30. doi: 10.48550/arXiv.1706.03762
- Veena, G., Kanjirang, V., and Gupta, D. (2023). AGRONER: An unsupervised agriculture named entity recognition using weighted distributional semantic model. *Expert Syst. Appl.* 229, 120440. doi: 10.1016/j.eswa.2023.120440
- Visentin, F., Cremasco, S., Sozzi, M., Signorini, L., Signorini, M., Marinello, F., et al. (2023). A mixed-autonomous robotic platform for intra-row and inter-row weed removal for precision agriculture. *Comput. Electron. Agric.* 214, 108270. doi: 10.1016/j.compag.2023.108270
- Wang, W., Dai, J., Chen, Z., Huang, Z., Li, Z., Zhu, X., et al. (2023). "Internimage: Exploring large-scale vision foundation models with deformable convolutions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14408–14419. doi: 10.1109/CVPR52729.2023.01385
- Wang, J., Ren, S., Zhang, Z., et al. (2019). Research progress on unmanned aerial vehicle for ecological remote sensing monitoring based on bibliometric assessment. *Trop. Geogr.* 39 (4), 616–624. doi: 10.13284/j.cnki.rddl.003157
- Wei, T., Zhao, L., Zhang, L., Zhu, B., Wang, L., Yang, H., et al. (2023). Skywork: A more open bilingual foundation model. doi: 10.48550/arXiv.2310.19341
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., et al. (2021). Ethical and social risks of harm from language models. doi: 10.48550/arXiv.2112.04359
- Wiles, O., Goyal, S., Stimberg, F., Alvisé-Rebuffi, S., Ktena, I., Dvijotham, K., et al. (2021). A fine-grained analysis on distribution shift. doi: 10.48550/arXiv.2110.11328
- Wu, J., Gan, W., Chen, Z., Wan, S., and Yu, P. S. (2023a). "Multimodal large language models: A survey," in *2023 IEEE International Conference on Big Data (BigData)*. doi: 10.1109/BigData59044.2023.10386743
- Wu, J., Hovakimyan, N., and Hobbs, J. (2023b). Genco: An auxiliary generator from contrastive learning for enhanced few-shot learning in remote sensing. doi: 10.48550/arXiv.2307.14612
- Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., et al. (2023c). Bloomberggpt: A large language model for finance. doi: 10.48550/arXiv.2303.17564
- Wu, C., Lin, W., Zhang, X., Zhang, Y., Xie, W., and Wang, Y. (2024). PMC-LLaMa: toward building open-source language models for medicine. *Journal of the American Medical Informatics Association* 31 (9), 1833–1843. doi: 10.1093/jamia/ocae045
- Xu, H., Man, Y., Yang, M., Wu, J., Zhang, Q., and Wang, J. (2023). Analytical insight of earth: a cloud-platform of intelligent computing for geospatial big data. doi: 10.48550/arXiv.2312.16385
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., et al. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *Br. J. Educ. Technol.* 55, 90–112. doi: 10.1111/bjet.13370
- Yang, X., Dai, H., Wu, Z., Bist, R., Subedi, S., Sun, J., et al. (2023c). Sam for poultry science. doi: 10.48550/arXiv.2305.10254
- Yang, Q., Du, X., Wang, Z., Meng, Z., Ma, Z., and Zhang, Q. (2023b). A review of core agricultural robot technologies for crop productions. *Comput. Electron. Agric.* 206, 107701. doi: 10.1016/j.compag.2023.107701
- Yang, J., Gao, M., Li, Z., Gao, S., Wang, F., and Zheng, F. (2023a). Track anything: Segment anything meets videos. doi: 10.48550/arXiv.2304.11968
- Yang, G., Li, Y., He, Y., Zhou, Z., Ye, L., Fang, H., et al. (2024). Multimodal large language model for wheat breeding: a new exploration of smart breeding. doi: 10.48550/arXiv.2411.15203
- Yao, S., Zhang, C., Ping, J., and Ying, Y. (2024). Recent advances in hydrogel microneedle-based biofluid extraction and detection in food and agriculture. *Biosens. Bioelectronics* 250, 116066. doi: 10.1016/j.bios.2024.116066
- Ye, J., Xu, H., Liu, H., Hu, A., Yan, M., Qian, Q., et al. (2024). mplug-owl3: Towards long image-sequence understanding in multi-modal large language models. doi: 10.48550/arXiv.2408.04840
- Ye, Q., Xu, H., Xu, G., Ye, J., Yan, M., Zhou, Y., et al. (2023). mplug-owl: Modularization empowers large language models with multimodality. doi: 10.48550/arXiv.2304.14178
- Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., et al. (2023). A survey on multimodal large language models. doi: 10.48550/arXiv.2306.13549
- Yuan, Y., Chen, L., Wu, H., and Li, L. (2022). Advanced agricultural disease image recognition technologies: A review. *Inf. Process. Agric.* 9, 48–59. doi: 10.1016/j.inpa.2021.01.003
- Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., et al. (2019). Monitoring plant diseases and pests through remote sensing technology: A review. *Comput. Electron. Agric.* 165, 104943. doi: 10.1016/j.compag.2019.104943
- Zhang, H., Li, X., and Bing, L. (2023a). Video-llama: An instruction-tuned audio-visual language model for video understanding. doi: 10.48550/arXiv.2306.02858
- Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., et al. (2022b). Opt: Open pre-trained transformer language models. doi: 10.48550/arXiv.2205.01068
- Zhang, N., Wu, H., Zhu, H., Deng, Y., and Han, X. (2022a). Tomato disease classification and identification method based on multimodal fusion deep learning. *Agriculture* 12, 2014. doi: 10.3390/agriculture12122014
- Zhao, B., Jin, W., Del Ser, J., and Yang, G. (2023a). ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification. *Neurocomputing* 557, 126708. doi: 10.1016/j.neucom.2023.126708
- Zhao, L., Zhang, L., Wu, Z., Chen, Y., Dai, H., Yu, X., et al. (2023b). When brain-inspired ai meets agi. *Meta-Radiology* 1, 100005. doi: 10.1016/j.metrad.2023.100005
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., et al. (2023c). A survey of large language models. doi: 10.48550/arXiv.2303.18223
- Zheng, C., and Li, H. (2023). The prediction of collective Economic development based on the PSO-LSTM model in smart agriculture. *PeerJ. Comput. Sci.* 9, e1304. doi: 10.7717/peerj-cs.1304
- Zhong, J., Liu, Z., and Chen, X. (2023). Transformer-based models and hardware acceleration analysis in autonomous driving: A survey. doi: 10.48550/arXiv.2304.10891
- Zhou, J., Li, J., Wang, C., Wu, H., Zhao, C., and Teng, G. (2021). Crop disease identification and interpretation method based on multimodal deep learning. *Comput. Electron. Agric.* 189, 106408. doi: 10.1016/j.compag.2021.106408
- Zhou, L., Pan, S., Wang, J., and Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing* 237, 350–361. doi: 10.1016/j.neucom.2017.01.026