Artificial Intelligence in Agriculture 15 (2025) 671–685

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/

Review Article

A review on enhancing agricultural intelligence with large language models

Hongda Li [a,b,1], Huarui Wu [b,c,d,1], Qingxue Li [b,c,d,*], Chunjiang Zhao [a,b,c,*]

[a] College of Information and Elect1rical Engineering, China Agricultural University, Beijing 100083, China
[b] National Engineering Research Center for Information Technology in Agriculture, Beijing 100101, China
[c] Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[d] Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs, Beijing 100125, China

ARTICLE INFO

Article history:
Received 14 January 2025
Received in revised form 28 May 2025
Accepted 29 May 2025
Available online 4 June 2025

Keywords:
Large language models
Agricultural knowledge
Knowledge integration
Knowledge Base
Intelligent question-answering

ABSTRACT

This paper systematically explores the application potential of large language models (LLMs) in the field of agricultural intelligence, focusing on key technologies and practical pathways. The study focuses on the adaptation of LLMs to agricultural knowledge, starting with foundational concepts such as architecture design, pre-training strategies, and fine-tuning techniques, to build a technical framework for knowledge integration in the agricultural domain. Using tools such as vector databases and knowledge graphs, the study enables the structured development of professional agricultural knowledge bases. Additionally, by combining multimodal learning and intelligent question-answering (Q&A) system design, it validates the application value of LLMs in agricultural knowledge services. Addressing core challenges in domain adaptation, including knowledge acquisition and integration, logical reasoning, multimodal data processing, agent collaboration, and dynamic knowledge updating, the paper proposes targeted solutions. The study further explores the innovative applications of LLMs in scenarios such as precision crop management and market dynamics analysis, providing theoretical support and technical pathways for the development of agricultural intelligence. Through the technological innovation of large language models and their deep integration with the agricultural sector, the intelligence level of agricultural production, decision-making, and services can be effectively enhanced.

1. Introduction

Agriculture, as a knowledge-intensive field, encompasses multidimensional expertise in crop cultivation, animal husbandry, environmental management, and more. Its production process is influenced by complex factors such as pests, weather, and soil, with knowledge being notably regional, time-sensitive, and fragmented (Wrzecińska et al., 2023; Yang et al., 2024a). Traditional agricultural knowledge dissemination relies on offline training or static databases, which struggle to address information asymmetry in remote mountainous areas. Additionally, the inefficient use of vast amounts of agricultural documents and unstructured data from farmers leads to high knowledge acquisition costs and challenges in standardization. With the increasing demand for precise decision-making and resource optimization in modern agriculture, overcoming the bottleneck of fragmented knowledge and achieving intelligent knowledge services and production support has become a key challenge in enhancing agricultural productivity and sustainability (Alemu et al., 2018; Mtega and Ngoepe, 2019).

LLMs, based on the self-attention mechanism of the Transformer architecture (Vaswani et al., 2017), are capable of semantic modeling of long texts, cross-modal information integration, and rapid knowledge emergence. This presents a novel approach to addressing challenges in agricultural knowledge management. By integrating existing agricultural knowledge and data, LLMs can provide support in areas such as agricultural technology, policies, and weather forecasts, helping agricultural producers to promote knowledge services (Tzachor et al., 2023). With the evolution from GPT-3 to GPT-4 (OpenAI, 2023), the parameter scale has expanded, further supporting multimodal processing capabilities, enhancing the efficient capture of pest and disease symptoms in images and providing control measures (Yang et al., 2024b). SkyEyeGPT, a remote sensing visual-language multimodal model, can

* Corresponding authors at: National Engineering Research Center for Information Technology in Agriculture, Beijing 100101, China.
E-mail addresses: liqx@nercita.org.cn (Q. Li), zhaocj@nercita.org.cn (C. Zhao).
[1] Co-first author.

https://doi.org/10.1016/j.aiia.2025.05.006

precisely analyze remote sensing images. In the process of agricultural development, it is crucial for accurately monitor crop growth, optimize cultivation practices, and manage agricultural resources effectively (Zhan et al., 2025). In practice, LLMs simplify agricultural technology query processes through intelligent Q&A systems, enabling farmers to interact using natural language and receive personalized recommendations, demonstrating wide application prospects in knowledge services and production support scenarios.

Although LLMs have been widely applied in fields such as healthcare (Thirunavukarasu et al., 2023), code generation (Rahmani et al., 2021; Sarsa et al., 2022), and mathematical reasoning (Luo et al., 2023), they still face multiple adaptation barriers in agricultural scenarios. The incorporation of specialized agricultural terminology and intricate knowledge systems into the LLMs pre-training framework remains a pressing concern, particularly given the models' robust general capabilities but limited effectiveness in the agricultural domain. The accuracy of inferences in cross-modal models has not yet met practical deployment standards. Agricultural artificial intelligence (AI) technology is shifting from fully autonomous systems to a human-centered AI paradigm (Holzinger et al., 2024), which emphasizes the human-machine collaboration, strengthening rather than replacing human expertise. When providing recommendations, LLMs need to balance human expert experience with legal, social, and ethical considerations, such as the EU's "Artificial Intelligence Act," which requires transparency, accountability, and human oversight.

Accordingly, this study focuses on the deep integration of LLMs within the agricultural domain, aiming to enhance key technologies and applications that drive agricultural intelligence. Section 2 lays the theoretical foundation by systematically explaining the principles of knowledge embedding in LLMs and methods for adapting them to vertical domains, with an emphasis on architectural design, pre-training tasks, and fine-tuning methods. Among these, fine-tuning and prompt engineering are critical to the adaptation of models to the agricultural domain and serve as fundamental prerequisites for achieving agricultural intelligence. Section 3 explores the integration of LLMs with agricultural knowledge bases, utilizing vector databases and knowledge graphs to enable the efficient storage and retrieval of agricultural information. The practical utility of LLMs in agricultural knowledge services is further validated through intelligent Q&A systems, multimodal learning, and the incorporation of domain-specific knowledge (Kraisnikovic et al., 2025). Section 4 addresses the challenges encountered during the adaptation of LLMs to agriculture-specific knowledge, proposing potential solutions and offering perspectives on future development, thereby comprehensively demonstrating the practical value of LLMs in advancing agricultural intelligence.

2. Large language models

In the context of the rapid development of natural language processing technologies, LLMs, with their vast parameter size and cross-scenario generalization capabilities, provide new technological solutions for the advancement of agricultural intelligence. Section 2.1 introduces the basic architecture of LLMs, examining the design principles and their influence on model performance. Section 2.2 concentrates on various pretraining approaches for LLMs, including how different pre-training tasks can capture linguistic patterns, syntax, semantics, and inter-sentence relationships. Section 2.3 reviews the fine-tuning strategies for LLMs, encompassing Section 2.3.1 on full-parameter fine-tuning, Section 2.3.2 on parameter-efficient fine-tuning, and Section 2.3.3 on prompt engineering. Each of these fine-tuning methods is discussed in terms of their role in enhancing the deep adaptation of LLMs to the agricultural domain. Through a systematic analysis of the aforementioned foundational theories, this study provides theoretical support and methodological guidance for constructing an LLM technical framework for knowledge integration in the agricultural domain.

2.1. LLMs architectures

The architectural design of LLMs is typically categorized into three primary types: Encoder-only, Encoder-Decoder, and Decoder-only. Fig. 1 illustrates the currently released LLMs, along with their open-source availability.

The Encoder-only architecture consists solely of the encoder, omitting the decoder, and is primarily employed for tasks that require understanding of input content without the need to generate an output sequence. By stacking multiple Transformer layers, each incorporating self-attention mechanisms and feedforward neural networks, the encoder can effectively capture the complex semantics of the input text. This type of architecture is commonly used for tasks such as text classification, named entity recognition (NER), and sentiment analysis.

The Encoder-Decoder architecture integrates both the encoder and decoder, making it suitable for tasks where the lengths of the input and output sequences may differ, such as machine translation and summarization. In this architecture, the encoder first converts the input sequence into a semantic embedding representation that captures its semantic features, while the decoder uses these features to generate the corresponding output sequence. Although this architecture is capable of modeling the complex relationships between input and output, its computational complexity is relatively high, and it requires more computational resources due to the involvement of both the encoder and decoder components.

The Decoder-only architecture is focused on text generation and includes only the decoder. The model receives an input sequence and generates the output iteratively, producing one word at a time. This architecture is particularly suited for tasks such as text generation, dialogue systems, and open-ended question-answering. Decoder-only models generate text in an autoregressive manner, predicting the next word based on the previously generated words. Further refinement of the Decoder-only architecture leads to two subtypes: Causal Decoder and Prefix Decoder. The Causal Decoder uses unidirectional attention, where both the input and output are processed unidirectionally, and the generation of new words depends solely on previously generated outputs without considering forthcoming words. In contrast, the Prefix Decoder employs bidirectional attention for the input portion, while using unidirectional attention for the output, allowing the generation of new outputs to take into account all previously generated content.

These architectural variations give each model distinct advantages and limitations, depending on the specific task at hand. Fig. 2 illustrates the attention patterns of these different architectures. For the Causal Decoder architecture, given a text sequence $X = (X_1, X_2, \cdots X_N)$, the model is based on autoregressive modeling. The conditional probability can be decomposed using the chain rule as follows:

$$P(x_1, x_2, \cdots, x_N) = \prod_{i=1}^{N} P(x_i | x_1, x_2, \cdots, x_{i-1}) \quad (1)$$

2.2. LLMs pretraining tasks

LLMs architectures learn linguistic knowledge and patterns from massive amounts of textual data through pre-training tasks. This section will introduce four main pre-training task methods: Masked Language Modeling (MLM), Next Sentence Prediction (NSP), Causal Language Modeling (CLM), and Permutation Language Modeling (PLM). These pre-training tasks each have their own characteristics, advantages, and limitations, and are applied to different model architectures.

MLM is a self-supervised learning approach extensively implemented in the pre-training phase of models such as BERT (Devlin et al., 2019). During training, MLM randomly masks certain words in the input text, and the objective is for the model to predict these masked words. To accomplish this task, the model must develop a deep understanding of the contextual dependencies within the sentence. MLM

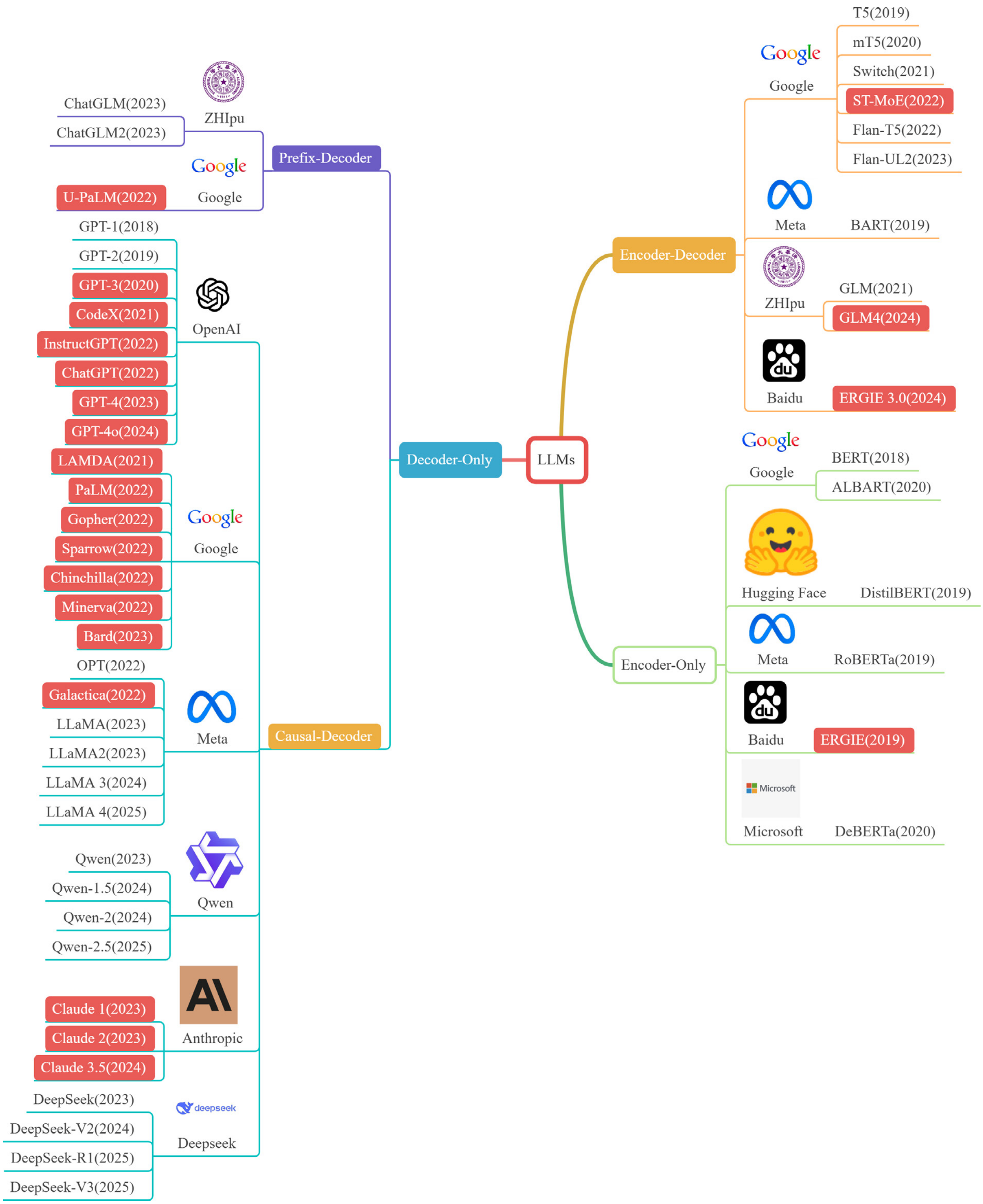

Fig. 1. The current development paths of three architectural techniques for large language models, with red annotations indicating that the model is not open-sourced.

primarily focuses on capturing lexical dependencies and long-range contextual relationships between words. However, while MLM excels at understanding and modeling such dependencies, it is less effective in generative tasks when compared to other model architectures.

NSP is one of the pre-training tasks in BERT, designed to help the model understand the relationship between sentences. During training, the model is tasked with determining whether two given sentences exhibit a natural sequential relationship. The NSP task is particularly important for understanding the relationship between questions and answers in question-answering tasks. It is especially beneficial when dealing with long texts and multi-turn dialogues, as it enhances the model's ability to maintain contextual coherence and continuity.

CLM is more oriented towards generative tasks, with the GPT (Chen et al., 2023; Ding et al., 2023) series being a prominent example. This model employs an autoregressive approach to text generation, where each word is predicted based on the preceding context. In contrast to

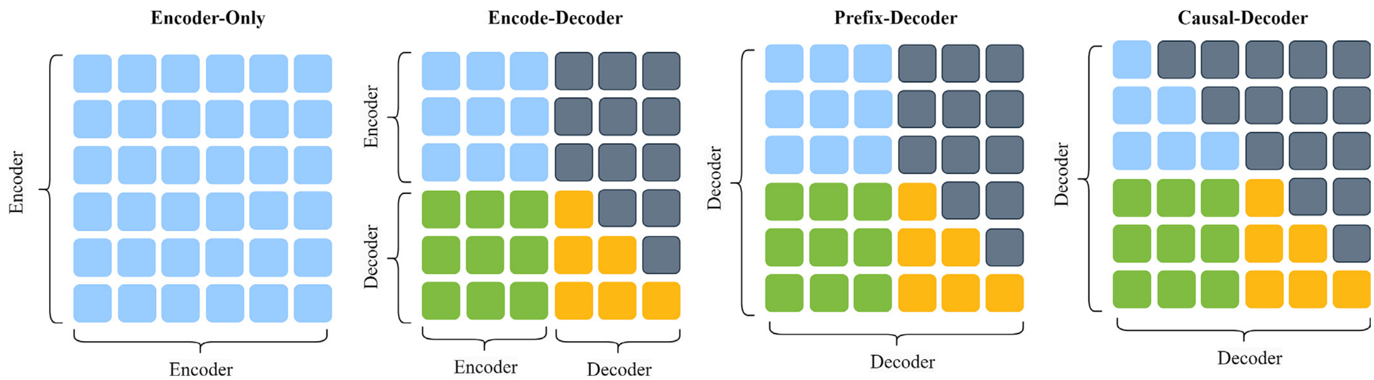

Fig. 2. The attention patterns of the architectures of four LLMs are compared, with blue, green, yellow, and gray rounded rectangles representing prefix tokens, attention between prefix and target tokens, attention between target tokens, and masked attention, respectively (Chang et al., 2023).

the bidirectional processing used in MLM, CLM generates text in a unidirectional manner. While this unidirectional generation enhances coherence and fluency in the produced text, it may result in a more superficial understanding of complex texts. As a consequence, the generated outputs may lack precision or exhibit incoherence at times.

PLM is the pre-training task employed in XLNet (Yang, 2019). During training, XLNet learns the dependencies between words by permuting the order of words in a sequence. This approach allows the model to capture a broader range of contextual information, making it particularly effective in modeling word order relationships and long-range dependencies. Compared to MLM, XLNet's permutation-based method offers greater flexibility and power, especially when dealing with complex linguistic structures.

The architecture of LLMs is closely related to their pre-training tasks, and their design together determines the performance of the model in various tasks. The above four are types of pre-training tasks rather than model architectures themselves. These pre-training tasks are used to train different model architectures, and by appropriately selecting the suitable architecture and pre-training tasks, these models can demonstrate strong capabilities in understanding and generating tasks, providing robust support for intelligent systems in agriculture.

2.3. LLMs fine-tuning

Large language models acquire rich linguistic patterns, grammar, and semantic structures during the pretraining phase. However, these pre-trained models are typically designed to serve a general purpose and may not be specifically optimized for specific tasks. Thus, to improve the performance of the model on the specific agricultural domain, fine-tuning is necessary. Fine-tuning methods can be categorized into various types depending on the training strategies, resource conditions, and application scenarios. The primary methodologies include full parameter fine-tuning, parameter-efficient fine-tuning (PEFT), and prompt engineering. The variations in model parameters resulting from different fine-tuning methods for LLMs are illustrated in Fig. 3.

2.3.1. Full parameter fine-tuning

Full parameter fine-tuning is a method that adjusts all parameters of the model to fit high-complexity tasks. Based on the existing linguistic knowledge of the pre-trained model, it is suitable for integrating large-scale, multi-domain agricultural knowledge into the model, enabling the model to comprehensively understand various complex relationships and key aspects of knowledge in agricultural production. Compared to training a large agricultural model from scratch, full

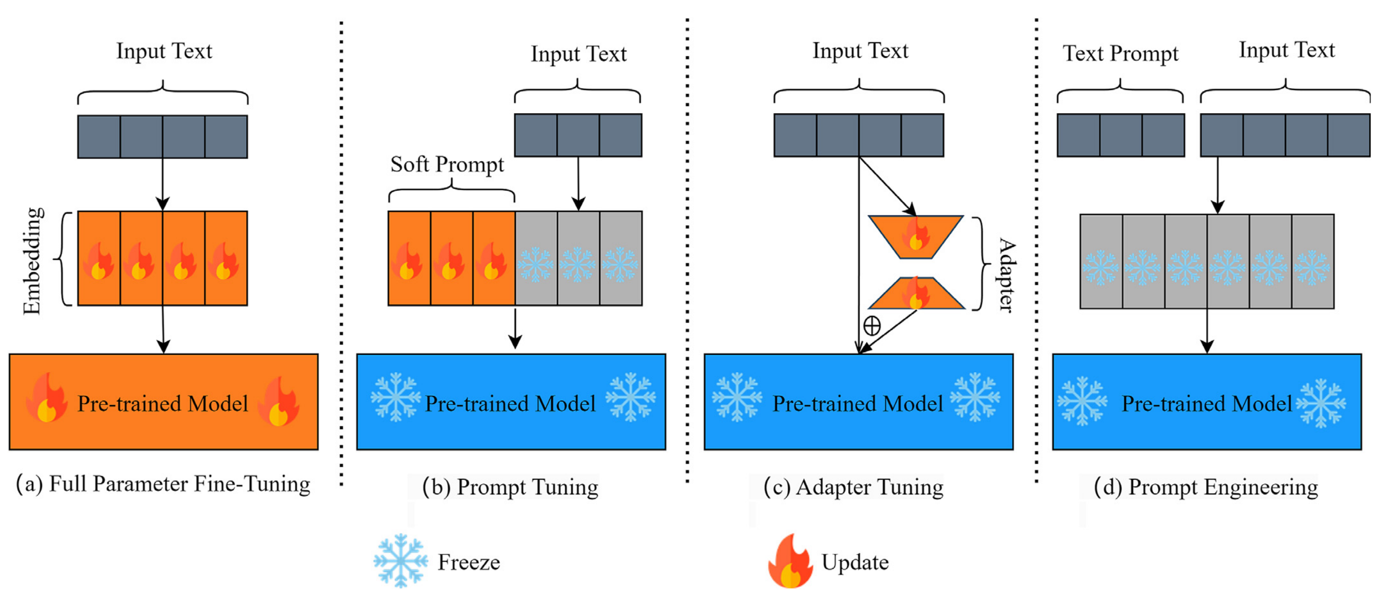

Fig. 3. Different Fine-tuning Methods for Large Language Models: (a) Full Parameter Fine-Tuning (discussed in Section 2.3.1); (b) Prompt Tuning and (c) Adapter Tuning (both are PEFT techniques discussed in Section 2.3.2); (d) Prompt Engineering (discussed in Section 2.3.3).

parameter fine-tuning requires fewer iterations of adaptive training for specific agricultural downstream tasks, significantly reducing computational resources. However, because it involves updating all parameters, the consumption of computational resources is still higher than that of some lightweight methods.

The effectiveness of this method is closely related to data quality. Its advantage lies in fully unleashing the potential of the model and is suitable for scenarios with ample annotated data and high task complexity. However, the computational cost is relatively high, and caution is needed to avoid the risk of overfitting in small datasets. In practice, a low learning rate is often adopted to retain pre-trained knowledge and prevent the model from forgetting its foundational language capabilities.

2.3.2. Parameter-efficient fine-tuning

Parameter-Efficient Fine-Tuning addresses the significant challenges of full parameter tuning of LLMs in terms of hardware and data volume (Chen et al., 2021; Hoffmann et al., 2022). By adjusting fewer than 1 % of the parameters to achieve domain adaptation, it strengthens professional performance while maintaining the basic capabilities of the model, significantly reducing the demand for computational resources. This is particularly important for agricultural applications, where limited computational resources make PEFT an efficient method to adapt general models to tasks such as pest diagnosis and planting decisions. Fig. 4 summarizes the current PEFT technical system, including key methods such as adapters and prompt tuning.

The adapter method aims to insert lightweight modules into pre-trained models, freezing the original parameters and only training the newly added modules (Buehler and Buehler, 2024; Gao et al., 2023a, 2023b; Hu et al., 2021; Hyeon-Woo et al., 2023; Kopiczko et al., 2024; Li et al., 2024a; Ponti et al., 2022; Rathnayake et al., 2022; Yeh et al., 2023; Q. Zhang et al., 2023; R. Zhang et al., 2023). MRS-Adapter (Yuan et al., 2023) achieves effective parameter transfer, utilizing contrastive learning to align multimodal data for remote sensing images and text retrieval. It can accurately identify and distinguish different crop types, providing basic data support for crop planting planning, growth status monitoring, and yield estimation.

Based on the prompt tuning method, the continuous vector representation of input prompts is optimized to guide the model to focus on the semantics of the agricultural domain (Lester et al., 2021; Li and Liang, 2021; Liu et al., 2023; Liu et al., 2022a; Wang et al., 2023a). For example, in pest and disease question answering, constraints and prompt words are embedded to construct templates (Wang et al., 2024), enhancing the relevance of the responses. Other efficient fine-tuning methods, such as LayerNorm Tuning (Zhao et al., 2023), only fine-tune the LayerNorm parameters of specific feature layers, such as attention layers and MLP. IA3 (Liu et al., 2022b) introduces learnable vectors to replace the weight matrix for Query, Key, and Value in the self-attention mechanism, further compressing the trainable parameter size. FourierFT (Gao et al., 2024) compresses adjustable weights through discrete Fourier transform, achieving better performance with fewer parameters than

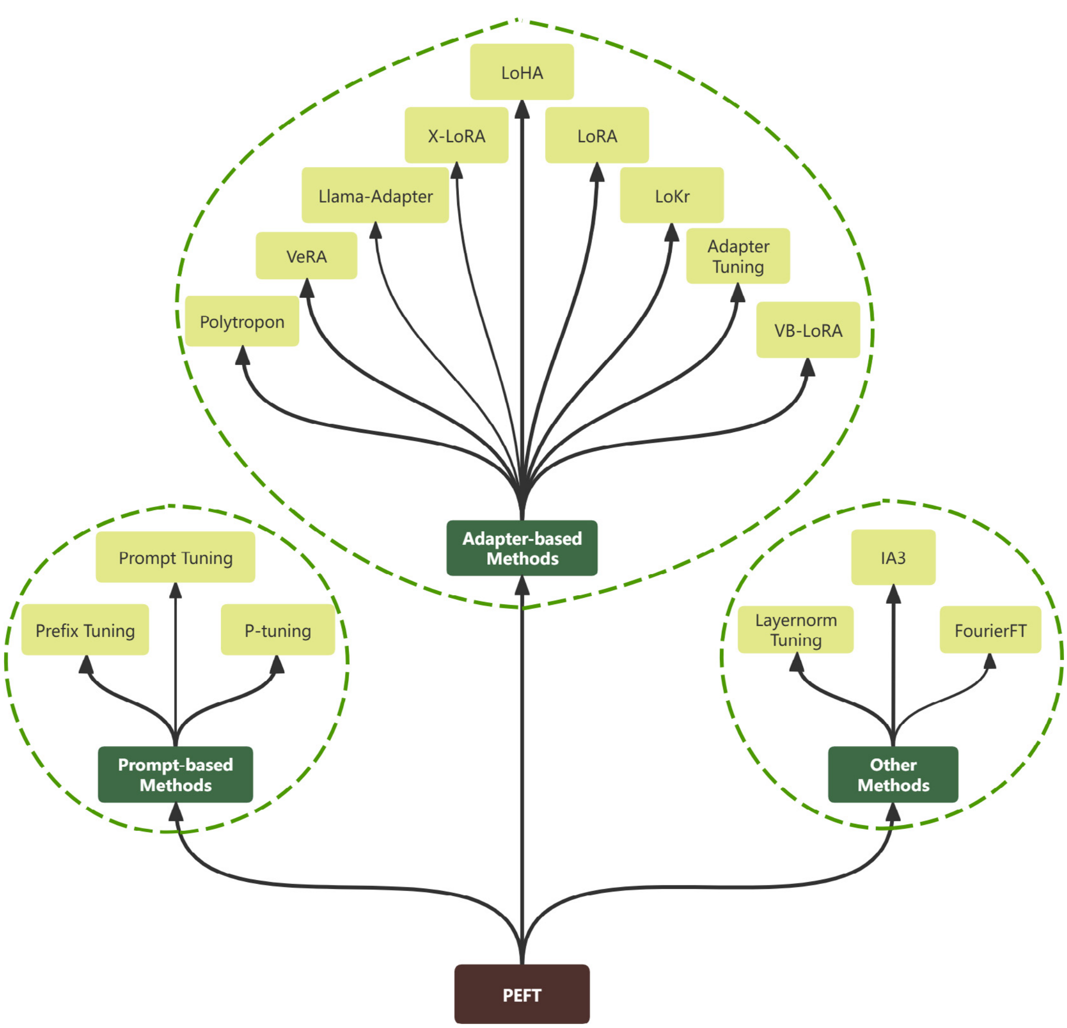

Fig. 4. Summary of efficient fine-tuning methods.

LoRA (Hu et al., 2021) in the GLUE benchmark tests and ViT classification tasks.

In conclusion, the efficient parameter fine-tuning technology adopts a more flexible strategy, adjusting only key parameters in the model. This can achieve targeted optimization towards the agricultural domain with less computational resource investment. While maintaining the basic capabilities of the model, it enhances its professional performance in the agricultural domain. The performance achieved by PEFT technology is comparable to that of full parameter fine-tuning techniques (Zhou et al., 2022).

2.3.3. Prompt engineering

Prompt engineering is a technique that guides LLMs to generate desired responses by carefully designing the way questions are posed, without modifying the model's parameters. By designing specific prompts to guide the model in generating outputs that meet the requirements, users can quickly obtain relevant knowledge and solutions when encountering various temporary and specific problems in agricultural production. Personalized prompts are designed based on the role, needs, and background knowledge of different users, aiming to achieve personalized integration and recommendation of agricultural knowledge.

Unlike parameter adjustment, prompt engineering (Zhang et al., 2023) guides the model to generate the expected output by designing structured input templates, offering a lightweight adaptation solution. Without effective prompt design and foundational technologies, the model is likely to hallucinate or fabricate answers, and more dangerous, models often produce very convincing and seemingly reasonable answers. Prompt engineering includes scenario prompts (Li et al., 2023a) that embed agricultural background information in the question.

Chain of Thought (Ge et al., 2023) involves breaking down complex agricultural problems into a sequence of sub-tasks (Ding et al., 2024; Miao et al., 2024; Mu et al., 2023), guiding the model to perform step-by-step reasoning, enhancing the logicality of multi-step decision-making. Prompt engineering primarily adopts three training paradigms: zero-shot, one-shot, and few-shot (Argüeso et al., 2020; Kojima et al., 2023), enabling quick adaptation to new tasks with a small number of examples, thus reducing the dependence on large-scale annotated data.

3. Agricultural knowledge intelligent question and answer

A theoretical framework for agricultural knowledge integration has been established, focusing on the architecture design, pre-training strategies, and fine-tuning techniques of LLMs. Both the Transformer architecture's ability to capture long-distance semantic dependencies and the domain-specific corpus injection and fine-tuning adaptations during pre-training serve as critical technological foundation for intelligent applications in agriculture. This chapter focuses on the transformation from theory to practice, exploring how to transform the general intelligence of LLMs into specialized knowledge service capabilities in the agricultural field through knowledge representation reconstruction and cross-modal capability expansion. It aims to build an intelligent question-answering system with precise knowledge services and decision-support capabilities.

The agricultural knowledge-based intelligent question-answering system aims to provide agricultural practitioners with fast and accurate access to agricultural knowledge. By enabling computers to comprehend natural language questions and effectively respond using existing knowledge or information, the system enhances the efficiency of agricultural information retrieval and strengthens decision-making capabilities (Zaib et al., 2022). Its primary role is to understand user queries and retrieve from agricultural knowledge bases, text data, or multi-source information to generate precise answers.

Based on the response method, intelligent question-answering systems can generally be divided into two types: retrieval-based question-answering (RQA) and generation-based question-answering (GQA). RQA extracts information from unstructured text, knowledge graphs, and other multi-source data through search technology, boasting high answer accuracy and fast response speed, suitable for clear factual questions. On the other hand, GQA relies on the knowledge inference and language generation capabilities of artificial intelligence models to proactively construct answers related to the questions, meeting the needs of open-ended and complex scenario question-answering. Table 1 presents an overview of the methods, question-answering modes, effectiveness, and limitations currently applied in the agricultural domain. Existing agricultural question-answering methodologies have continuously evolved, transitioning from traditional LSTM-based approaches to multimodal Transformers and other innovative techniques. Although these question-answering systems can provide responses within specific contexts, they generally suffer from incomplete knowledge coverage and imbalanced domain-specific data. As a result, the scope of these systems is constrained by the size of their knowledge bases. The traditional knowledge representation methods are not compatible with the processing logic of LLMs, which restricts the level of intelligent interaction in the system. This situation highlights the necessity of constructing knowledge carriers that are compatible with the processing logic of LLMs, and breaking through performance bottlenecks with structured knowledge representation and multimodal capabilities.

3.1. Construction of agricultural knowledge base

In the agricultural domain, there is a large amount of unstructured data from various sources such as literature, databases, social media, and other sources, and their inconsistent formats make them complex to process. When constructing an agricultural knowledge base, it is necessary to transform scattered information into structured or queryable forms through preprocessing steps, such as cleaning, denoising, and standardization. This process is fundamental to the construction of the knowledge base and provides domain knowledge support for applications such as intelligent question answering and decision support.

3.1.1. Vector database

The vector database is an essential technology for efficient storage and retrieval of agricultural information, suitable for high-dimensional data management of embeddings from text, images, etc. Through a preprocessing pipeline, agricultural knowledge is transformed into vector representations and stored. The main challenge lies in constructing efficient query mechanisms that can deliver rapid and accurate responses to complex semantic queries in big data and deep learning scenarios.

Early information retrieval techniques relied primarily on rule-based methods (Thorat and Jadhav, 2020) and shallow learning techniques, such as Support Vector Machines (Suzuki et al., 2002) and the Maximum Entropy Model (Figueroa and Atkinson, 2011; Sun et al., 2006), which were limited in handling semantics and reasoning, with relatively weak generalization capabilities.These methods were limited in handling complex semantics and reasoning and exhibited relatively weak generalization capabilities. Traditional information retrieval models often utilized statistical methods like BM25 (Gulati et al., 2023) and TF-IDF (Chamorro-Padial et al., 2024), which rely on keyword matching for text retrieval. However, these approaches often lack precision when applied to specialized agricultural knowledge bases. With the rise of deep learning techniques, especially the use of vector databases, the accuracy and efficiency of information retrieval have seen substantial improvements (Chamorro-Padial et al., 2024). Vector databases convert text into high-dimensional vector representations, enabling queries to be based not merely on keyword matching but on semantic similarity calculations, thereby improving the system's understanding of user query intent.

In practical applications, vector databases compare the semantic vectors of query statements with those stored in the database to quickly identify the most relevant agricultural knowledge. Advanced

Table 1
Agricultural intelligent question answering system.

Paper	Agriculture domain	Method	Q&A Category	Performance, Contribution	Limit	Dataset
(Tang et al., 2024)	Soil type, crop pests and diseases, and agricultural product trade information	BE-BILSTM	Information extraction	Precision 95.19 %	Data Imbalance in the Dataset	The self-made dataset, and the data sources include various agricultural websites in Malaysia
(Lu et al., 2024)	Agricultural disease detection	Multimodal Transformer	disease detection, image captioning and object detection	Average Precision 92 %	The text data cannot cover the Agriculture disease field.	The self-made dataset, and the data sources include the databases of various agricultural research institutions, agricultural science and technology websites, professional forums, scientific research paper databases, etc.
(Lan et al., 2023)	Fruit tree diseases	Multimodal bilinear decomposition pooling models	Visual Question Answering	Accuracy 86.36 %	Inaccurate keyword positioning leads to incorrect predicted answers.	Private dataset
(Huang et al., 2023)	Agricultural scenarios	Bert for semi-supervised contrastive learning	Answer Generate	Applied to Machine Reading Comprehension and Open-Ended Question Answering Tasks in Agricultural Scenarios	The scope of question-and-answer knowledge in the knowledge base is limited.	The self-made dataset, and the data sources include the Agricultural Information Database of China Agricultural University and the database of the Shandong Farm Manager App.
(Kung et al., 2021)	Intelligent pig farming	LSTM and Bi-GRU	Answer retrieval	The intelligent pig farming knowledge question and answer system has good response effects.	Lack of comprehensive evaluation.	Private dataset
(Kpodo et al., 2024)	Agricultural extension	LLM AgRoBERTa	Answer Generate	Question-Answering Task Benchmark Dataset AgXQA	Relies on the quality and quantity of training data	The self-made dataset, the Agricultural Question - Answering AgXQA dataset.
(García-García et al., 2021)	Grape pests	Expert System	Answer retrieval	A web application was developed	The matrix similarity method does not fully distinguish features.	Private dataset
(Hao et al., 2023)	Agricultural pests and diseases	Intent Detection and Slot Filling (IDSF)	Answer Generate	The first IDSF agricultural dataset AGIS is collected and annotated	Relies on large amounts of annotation data	Private dataset
(Guo et al., 2024)	Questions and answers on agriculture	BERT-DPCNN	Answer Generate	Accuracy 99.07 %	The lack of comparison with other mainstream text classification models.	The self-made dataset, and the data sources include China Agricultural Technology Extension Information Platform, China Agricultural Information Network, and the First Agricultural Economy Network.
(Qian et al., 2024)	Multimodal Question Answering	CROMIC-QA	Answer Generate	The image semantic interaction method enhances the question-answering capability.	The evaluation form is singular, but it is friendly to Chinese question answering.	The self-made dataset, and the data sources include online Q&A communities in the agricultural field.
(Wang and Zhao, 2024)	Agricultural Knowledge	Knowledge Graph	Knowledge Base	Agricultural Knowledge Graph (AGKG) automatically identifies agricultural knowledge entities.	The data construction is not comprehensive, relying solely on entity retrieval and question-answer evaluation.	Private dataset

technologies, such as Milvus and FAISS, provide efficient similarity retrieval functionalities, enabling the database to return relevant data that closely matches the query semantics at high speeds. This data can then be integrated into the reasoning processes of LLMs to generate more accurate and reliable answers. Additionally, systems like LangChain-ChatChat (Liu and Chilton, 2022) adopt text segmentation and embedding vector databases, breaking down text into sentence chunks for semantic matching, thereby extracting the most relevant information. This approach improves retrieval accuracy and ensures the system can generate high-quality answers to complex queries.

Building upon this, knowledge augmentation techniques are introduced to optimize both the retrieval and answer generation processes (Liang et al., 2023). Knowledge augmentation refers to adjusting and reconstructing the original text by integrating external knowledge or contextual information to make it more contextually relevant and enriched. This process aids in enhancing the comprehensibility and practicality of generated results (Lai et al., 2023). In agricultural knowledge bases, reorganizing and supplementing relevant background information according to different user needs and contexts can significantly improve the quality of the system responses.

In conclusion, building an efficient agricultural knowledge base requires addressing challenges in data collection, preprocessing, and storage, as well as designing effective query mechanisms. By leveraging vector databases and deep learning technologies, retrieval accuracy and response speed can be significantly improved. The integration of modern intelligent retrieval techniques, semantic understanding, and knowledge augmentation can greatly enhance the accuracy and practicality of agricultural knowledge bases in real-world applications, providing robust support for agricultural decision-making and problem-solving.

3.1.2. Agricultural knowledge graph

To further enhance the intelligence of agricultural knowledge bases, it is crucial to construct semantic structures that are compatible with large language models. Knowledge graphs serve as an effective tool by organizing entities, attributes, and their interrelationships in the agricultural domain through graph-based representations (Chen et al., 2020). Entity recognition is employed to extract significant entities from text or data in the agricultural field, such as crop names, soil types, meteorological conditions, and pests and diseases, transforming unstructured information into structured knowledge units, thereby facilitating downstream storage, querying, and analysis. Relationship extraction is used to uncover semantic connections between entities, thereby establishing the fundamental framework of an agricultural knowledge graph. Establishing these relationships supports graph-based reasoning and enables precise agricultural question answering. When constructing agricultural knowledge graphs, open-source graph databases like Neo4j and ArangoDB can provide users with convenient graph data storage and querying capabilities.

Traditional rule-based NER methods (Biswas et al., 2019; Nismi Mol and Santosh Kumar, 2024) utilize domain-specific dictionaries and regular expression matching to identify specific terms. While these approaches are simple and effective for recognizing fixed terminology within a domain, they struggle with generalization, particularly in handling complex contexts and newly emerging terms. In contrast, machine learning-based NER approaches using Conditional Random Fields (Qian et al., 2023), a machine learning technique, train the model through supervised learning in the agricultural domain, yet it relies heavily on high-quality annotated data. Addressing the dynamic adaptability of agricultural terminology in modern agriculture, along with challenges in recognition accuracy and efficiency, remains a significant hurdle. Agricultural texts are often intricately linked to cross-domain information, such as climate, economy, and policy, making it difficult for traditional methods to generalize across multiple domains.

Semantic-level deep learning techniques enable the extraction of implicit entity information from contextual semantic relationships, overcoming the limitations of traditional methods that depend on single domains or fixed rules, and significantly enhancing the system's ability to adapt to and recognize complex agricultural texts. Veena et al. (2023) proposed a novel weighted distributional semantic model that integrates an extended BERT model with Latent Dirichlet Allocation (LDA) topic modeling, referred to as ExBERT_LDA+. This method is applied to unsupervised NER in the agricultural domain. It focuses on identifying six primary entities: diseases, soils, pathogens, pesticides, crops, and locations. A corpus containing 30,000 sentences was created for evaluation. Experimental results indicate strong performance in unsupervised NER for the agricultural domain, achieving a macro-average F1 score of 80.43 %. LLMs can further enrich and refine knowledge graphs by assisting with entity recognition through prompt engineering to guide LLMs in identifying agricultural entities within texts. Yao et al. (2024) constructed and annotated a large-scale Chinese agricultural pest and disease NER corpus, named AGCNER, which includes 13 categories, 206,992 entities, 66,553 samples, and 3,909,293 characters. Additionally, they fine-tuned an agricultural language model, Agbert, to achieve high-accuracy named entity recognition. Zhang et al. (2021) addressed several challenges in Chinese NER for apple pest and disease domains, such as the variety of entity categories, the presence of aliases or abbreviations, and difficulties in recognizing rare entities. They proposed APD-CA, a character-based Chinese NER model tailored for the apple pest and disease domain. This model combines dictionaries and similarity-enhanced word representations, and experimental validation on the APDCner corpus shows high accuracy, recall, and F1-score performance. Optimizing relation extraction through LLMs involves integrating knowledge embeddings with pre-trained language models, thereby enhancing the accuracy of relation extraction (Pan et al., 2024).

3.2. Generative question answering

GQA utilizes generative models to produce answers directly, offering greater flexibility and diversity in the generated content. The generative decoder is the core component of this technology. Common architectures include LSTM (J and Nidamanuri, 2024), GRU and Transformer (Zheng et al., 2024), among others. It is capable of generating coherent and relevant responses based on context. This method leverages the natural language generation capabilities of LLMs, breaking through the limitations of RQA, which depends on pre-set answer databases. It can generate logically coherent answers to open-ended and complex situational questions.

Table 2
A comparison of RQA and GQA systems.

Feature	RQA	GQA
Principle	Fetches from a predefined database	Synthesizes responses using pre-trained models
Flexibility	Low	High
Accuracy	High, relies on existing information	May generate errors
Resource Dependency	Relies on efficient retrieval algorithms	Requires large-scale model computation
Scalability	Scales well with efficient indexing	Computationally expensive for large datasets
Consistency	High, similar phrasing retrieves the same answer	May vary depending on prompt
Response Speed	Faster, depends on retrieval algorithm	Slower due to model inference time
Context Handling	generate context-aware responses	Relies on context provided in the source
Use Cases	Factual Q&A, domain-specific queries	Open-ended questions, creative applications

In contrast, RQA systems provide answers directly obtained from established databases, ensuring the reliability and reproducibility of the responses. By optimizing retrieval algorithms, RQA can efficiently handle large datasets, relying on robust search mechanisms for quicker response times. However, RQA can only answer-questions within the scope of the database and struggles with creative or open-ended inquiries. Table 2 compares the advantages and limitations of these two question answering models, highlighting that RQA excels in information accuracy and knowledge coverage, while GQA is better suited for handling complex or ambiguously phrased questions. In practical applications, an integrated approach combining both systems, as illustrated in the structure depicted in Fig.5, this approach significantly enhances the overall performance of the Q&A system, thereby providing agricultural practitioners with precise and comprehensive knowledge services.

3.3. Large language models in agriculture

LLMs are gradually becoming important tools for advancing intelligence in the agricultural sector. The applications of LLMs can be divided into two categories: domain-specific LLMs developed specifically for agriculture and general-purpose LLMs adapted to agricultural scenarios through prompt engineering or dedicated interfaces.

Agricultural Large Language Models (AgLLMs) integrate the characteristics of the agricultural industry in logical reasoning, adopting specific analytical methods. For instance, AgroNT serves as a specialized LLM focusing on crop gene expression and genetic variation, capable of providing accurate genotype-to-phenotype predictions for plant breeding (Mendoza-Revilla et al., 2024). Additionally, the food-oriented LLM, FoodS, based on a Chinese cuisine corpus, understands food data through perception and reasoning, outperforming general-purpose LLMs in chef-related and diet-related examinations (Zhou et al., 2024). This logical reasoning method based on agricultural features can significantly improve the practicality and accuracy of intelligent question and answer systems. Numerous enterprises and academic institutions are also investing substantial resources in developing specialized agricultural LLMs. Shanxi University, Shanxi Agricultural University, and Zhongke Feichou have fine-tuned models based on massive supervised agricultural data, possessing extensive agricultural domain knowledge and intelligent analysis capabilities. The Houji AfriMa China agriculture model supports agricultural domain crop-related issues. The Tiangong Kaifeng agricultural Q&A model from Harbin Institute of Technology integrates agricultural knowledge graph construction, crop growth models, and agricultural Q&A models to achieve crop growth prediction. The Shennong 1.0 and Shennong 2.0 models from China Agricultural University introduce technologies

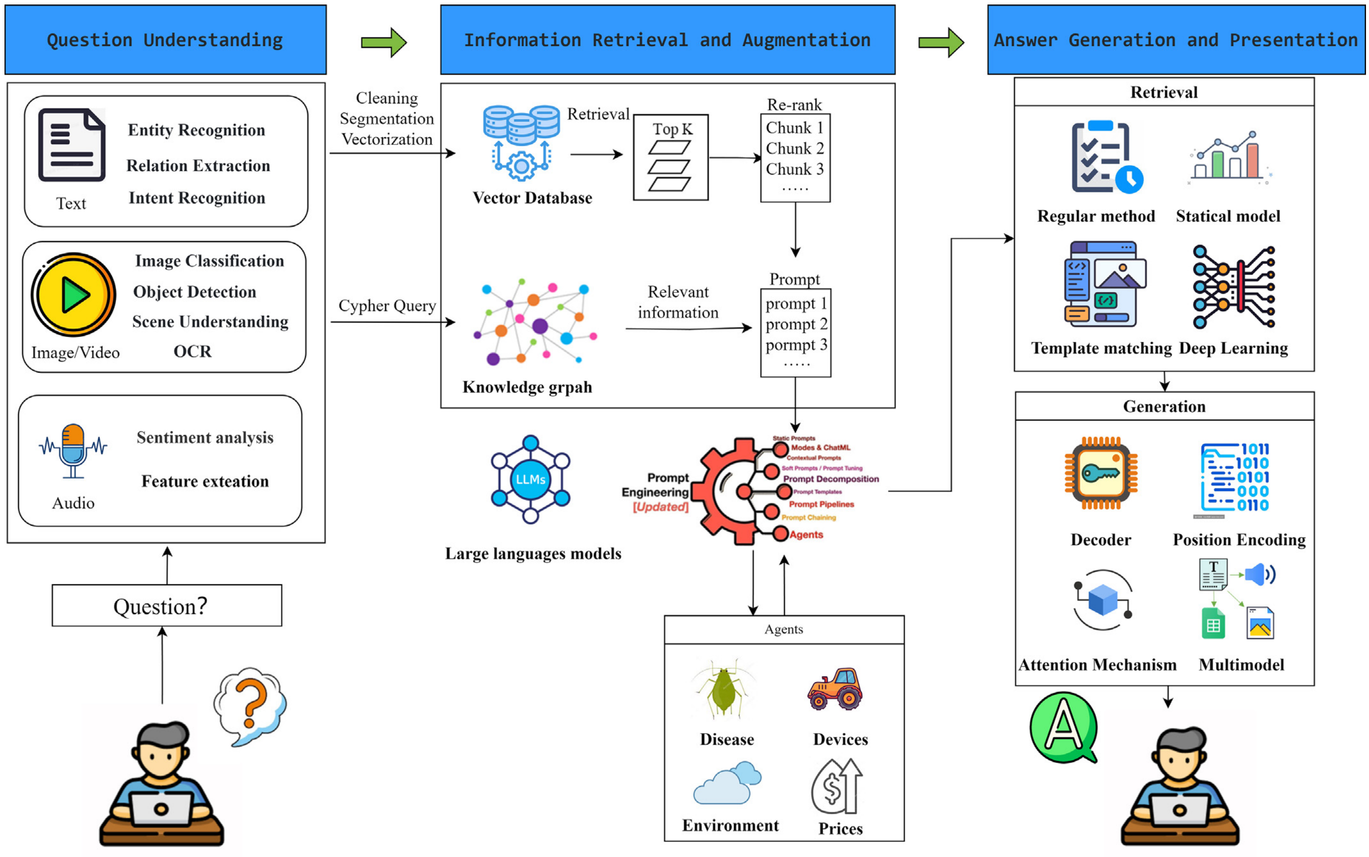

Fig. 5. Intelligent question answering architecture.

like knowledge graphs and vector databases, supporting multimodal interactions including images, voice, video, and documents, with capabilities such as agricultural knowledge Q&A, text semantic understanding, text summarization, and agricultural production decision reasoning. The Shangtang Technology Zhizhi Decision model combines knowledge from crop cultivation, biology, meteorology, and policy laws to provide comprehensive and reasonable agricultural suggestions.

In addition to specialized agricultural LLMs, some studies have focused on adapting general-purpose LLMs for agricultural scenarios. By utilizing prompt-based instructions, general-purpose LLMs have been shown to effectively address agricultural decision-making challenges, such as pest management (Yang et al., 2024b), and can accurately simulate vegetable crop growth, identifying growth stages and predicting crop development (Chunjiang et al., 2024). Through the integration of agricultural domain knowledge, general-purpose LLMs, when combined with Retrieval-Augmented Generation (RAG) techniques, can overcome the limitations of training on agricultural knowledge databases, further enhancing their performance in agricultural decision support systems (Vizniuk et al., 2025). With multimodal extensions, the GPT-4 large language model has incorporated a vision module, enabling the diagnosis and management of potato diseases (Zhu et al., 2025). The combination of these technologies not only broadens the application scope of LLMs in agriculture but also accelerates the rapid development of intelligent decision-making in the agricultural sector.

AgLLMs demonstrate significant differences from general-purpose LLMs across multiple dimensions. In terms of architectural design, AgLLMs incorporate a dedicated agricultural agent module (Yang et al., 2025), allowing for direct integration with external tools, whereas general-purpose LLMs require plugins to call tools indirectly. Regarding data foundation, AgLLMs are trained based on domain-specific datasets (Didwania et al., 2024), while general-purpose LLMs learn from texts across the entire web. In their information processing mechanisms, AgLLMs leverage the RAG framework to dynamically access the latest agricultural research findings. In contrast, general-purpose LLMs depend on static pre-trained knowledge bases, which are prone to information latency. From the perspective of reasoning capabilities, AgLLMs utilize decision tree visualizations (Maksimovich et al., 2024) and knowledge graph path tracing, providing clear and accurate methodologies for addressing agricultural issues. On the other hand, reasoning in general-purpose LLMs is largely based on shallow semantic associations (McCoy et al., 2023).

These applications demonstrate the diversity of domain-specific and universal large models in solving various agricultural problems, where the integration of multimodal learning and knowledge graphs has become two major key research directions for further expanding the application boundaries of LLMs in agriculture.

3.3.1. Multimodal learning

In the context of intelligent decision-making in agriculture, multimodal learning integrates heterogeneous information such as text, images, videos, audio, and sensor data (Lahat et al., 2015), constructing a cognitive framework that more closely resembles the real agricultural environment. Compared to traditional AI-driven decision support systems, such as rule-based expert systems (Herindra et al., 2023), single modality machine learning models, and statistical models (Haval and Mishra, 2024; Sitokonstantinou et al., 2024), multimodal learning demonstrates significant technical advantages in data processing paradigms and decision-making logic. Traditional expert systems rely on manually defined rules and struggle to adapt to the dynamic changes in agricultural environments; single modality models are constrained by the one-sidedness of data types. For example, relying only on text data makes it difficult to accurately identify subtle visual features of crop diseases, or relying solely on image data cannot combine background knowledge like meteorology and soil for comprehensive inference.

The advantage of multimodal learning lies in its ability to achieve cross-modal semantic alignment and fusion. By leveraging pre-trained models such as CLIP (Radford et al., 2021) and BLIP-2 (Li et al., 2023b) on massive image–text pairs, the models can establish mapping relationships between different modal data. At the technical implementation level, strategies such as feature concatenation, dynamic weighting of attention mechanisms, and multimodal result voting can effectively integrate complementary information from different modalities, significantly enhancing the comprehensiveness and reliability of decision-making.

The deep integration of multimodal LLMs with agricultural scenarios is currently fostering the development of novel application models. By combining the visual-language model GroundDino (Koh et al., 2023) with the general segmentation model Segment Anything (Kirillov et al., 2023), a visual-language collaborative framework is established, enabling precise crop identification and dimensional quantification (Barakat et al., 2023). This framework provides granular data support for yield estimation and resource allocation. Microphone arrays are deployed to collect animal sounds, extract acoustic features, and input them into LLMs for pattern recognition and health status classification (Neethirajan, 2024). Meanwhile, technologies such as Mono3DVG (Zhan et al., 2023), which combine monocular imagery with linguistic descriptions, facilitate the localization and tracking of 3D objects within agricultural fields, driving the intelligent advancement of tasks such as crop growth monitoring and livestock distribution in precision agriculture. Multimodal LLMs, by fusing images, sounds, and text data, significantly enhance the perceptual depth and decision-making intelligence of agricultural scenarios.

3.3.2. Integration of knowledge graphs and large language models

Agricultural knowledge graphs are a way of structuring and visualizing agriculture-related knowledge, data, and information, designed to support agricultural research, production, management, and decision-making. Knowledge graphs provide powerful semantic context for question-answering systems (Bao et al., 2016; Huang et al., 2019; Zhang et al., 2018). Researchers leverage entities and relationships within the knowledge graph to understand queries, retrieve relevant knowledge, and generate accurate answers. Knowledge graph-based question-answering methods include template-based question matching (Zheng et al., 2018) and graph-based path searching (Yasunaga et al., 2021; Zhang et al., 2023).

In the agricultural intelligent question and answer system, specific raw data is crucial for improving the system's accuracy and practicality. This includes historical records of crop growth cycles and climatic conditions, soil characteristic data from different regions, and monitoring records of agricultural pests and diseases, and their prevention and control measures. Xia et al. (2020) proposed a knowledge-based question-answering approach for the agricultural pest and disease domain using entity linking, external knowledge, and similarity calculations. This method uses entity linking to filter out inappropriate entities and relationships while utilizing external knowledge to improve recall. The questions are represented by word vectors, and entities within the knowledge graph transformed into vectors. By calculating the similarity between the question and potential answers, the system generates the final answer.

Structured data knowledge graphs can complement LLMs. Knowledge graphs are especially valuable for providing multimodal knowledge beyond vector databases, enabling the incorporation of image-text knowledge. Injecting the structured data of knowledge graph triples into large language models (Moiseev et al., 2022) can enhance the generation of coherent responses and provide semantic context cues (Li et al., 2023c). By adopting multimodal conceptual descriptions and integrating multimodal LLMs, knowledge-based visual question answering (VQA) can be enhanced, leading to more precise answers (Zha et al., 2023).

The future structure of question-answering systems is expected to integrate knowledge retrieval with pre-trained models (Abbasiantaeb and Aliannejadi, 2024) in a RAG approach. This combined paradigm leverages the high accuracy of retrieval-based methods alongside the reliability of generative models (Sui et al., 2024), effectively addressing the limitations of both methods and providing optimal solutions for complex scenarios. The integration of data allows agricultural intelligent question and answer systems to be more targeted when providing advice, thus forming a significant difference from the output of general-purpose LLMs.

The integration of agricultural knowledge graphs and LLMs demonstrates a strong synergistic effect in practical applications. LLMs and knowledge graphs mutually enhance each other: the knowledge graph provides accurate domain knowledge and reasoning paths, while LLMs offer natural language understanding, context integration, and personalized expression capabilities. Ultimately, this leads to intelligent agricultural services that are both professional and user-friendly.

3.4. Intelligent question answering systems

The field of agricultural intelligent Q&A systems has witnessed extensive research on technologies such as named entity recognition, multimodal learning, and knowledge graph integration. These advancements have accumulated valuable practical experience and innovative ideas for system development. However, existing studies still lack systematic architectural integration tailored to agricultural application scenarios. Yi et al. (2025) proposed an agricultural standard generation framework based on LLMs, by integrating vector databases and knowledge graph technologies, enable efficient information storage and retrieval, as well as intuitive representation of the structured relationships in agricultural knowledge, thus providing crucial technical support for the standardization process in agriculture. Dhavale et al. (2024) developed a crop disease recognition chatbot, leveraging large models combined with Generative Adversarial Networks to enhance image quality, CNNs for precise disease detection, and LLMs to provide professional diagnostic advice and solutions for farmers. Tzachor et al. (2023) provide a macro-level analysis of the current application status of LLMs in agricultural extension services, comprehensively discussing the challenges and opportunities they face, thereby laying a solid theoretical foundation for future research and pointing out clear research directions.

Following a comprehensive review of the previously mentioned studies, we can refine a general framework for intelligent Q&A systems. Typically, this framework consists of three stages: question understanding, information retrieval and enhancement, and answer generation and presentation. In the question understanding stage, the system receives multimodal inputs such as voice, text, images, and videos, and analyzes the information from these different modalities to accurately interpret the user's intent. During the information retrieval and enhancement stage, the system searches an agricultural knowledge database based on the query. One approach uses a vector-based knowledge repository, which stores cleaned, segmented, and vectorized text for high-speed querying, followed by retrieval and re-ranking to obtain the most relevant data. Another approach utilizes a structured agricultural knowledge graph, where information is retrieved using Cypher queries. By integrating interfaces of agricultural machinery and sensors, the system enhances its data acquisition and tool execution capabilities, improving the usefulness of the answers.

The answer generation and presentation stage is crucial. The retrieved content from the agricultural knowledge base serves as prompt words to organize information, which is input together with the original query into the LLMs. The LLMs perform joint reasoning based on these inputs to generate responses. The output answers undergo standardization of agricultural professional terms and formatting, highlighting important information to align with agricultural industry presentation habits. Text outputs are also transformed into multimodal forms such as charts, to enhance the efficiency of information delivery.

This architecture combines the advantages of both retrieval-based and generative question-answering systems, leveraging agricultural expertise alongside the capabilities of large language models, as shown in Fig.5. It constructs an intelligent, accurate, and scalable agricultural question-answering system that addressing the needs of agricultural production practices.

4. Challenges and future directions

4.1. Collection and annotation of agricultural data

The acquisition and annotation of agricultural data are crucial for the application of LLMs in agricultural knowledge-based intelligent Q&A systems. Data collection serves as the foundation for building intelligent Q&A systems, ensuring that models have broad and diverse knowledge coverage. Agricultural data sources are varied, including meteorological records, soil and crop information, satellite imagery of farmland, and market dynamics. This diversity allows the model to better understand and analyze agricultural contexts. However, in practice, the data are often fragmented, incomplete, and restricted by privacy concerns and access limitations. To integrate these heterogeneous data sources into a unified format for model training, specialized data cleaning and standardization strategies are required to eliminate noise and enhance data quality. Additionally, establishing cross-institutional and cross-regional data-sharing mechanisms and creating unified agricultural data standards and collection protocols are crucial for reducing data fragmentation and improving data accessibility.

Data annotation plays a crucial role in improving the accuracy and usefulness of large language models. In the agricultural sector, especially for tasks like pest and disease identification and crop growth diagnostics, the process often demands significant domain expertise and hands-on experience. Consequently, the scarcity and high cost of annotated data represent significant bottlenecks to model performance. To address this challenge, techniques such as transfer learning or semi-supervised learning can be employed. These approaches involve pre-training model on large amounts of unannotated data or data from similar domains, followed by fine-tuning them with a small set of precisely annotated examples, thereby improving the model's learning efficiency in new tasks and contexts. Additionally, leveraging self-supervised pre-training or generative adversarial networks (GANs) allows the extraction of valuable information from vast amounts of unannotated or weakly annotated data, thereby providing richer learning resources and further enhancing model performance.

4.2. Key technical challenges and breakthroughs of large language models

The complexity and specificity of agricultural scenarios pose unique challenges for the technical adaptation of LLMs, requiring targeted breakthroughs that integrate agricultural domain knowledge, data characteristics, and practical application needs.

4.2.1. Deep integration of domain knowledge and logical reasoning

Agricultural knowledge involves multiple disciplines, has a complex structure, and the knowledge formats from different sources are not unified, making it difficult to effectively integrate into models. The inference mechanism of traditional LLMs struggles to handle the complex causal relationships in agricultural problems, leading to inaccurate inference results. To address this, there is a need to develop a unified agricultural knowledge representation framework to standardize knowledge in different formats. Knowledge graph embedding technology should be employed to integrate structured agricultural knowledge into the pre-training process of LLMs. Additionally, a causal inference enhancement module should be introduced, combined with the causal logic rules in the agricultural domain, to optimize the model's inference process and improve its accuracy.

4.2.2. Multimodal data fusion and real-time decision making

In agricultural scenarios, images, sensor data, and other multimodal data are subject to noisy interference. The data volume is vast and high-dimensional, increasing the difficulty of data fusion. Cross-modal semantic mapping is complex, and traditional fusion methods struggle to achieve precise alignment, affecting decision-making efficiency and accuracy. It is necessary to design a multimodal data denoising algorithm to preprocess input data. By adopting attention mechanisms and multimodal feature fusion networks, the correlation learning between different modal data can be enhanced, improving the accuracy of semantic mapping. A collaborative architecture combining edge computing and cloud computing should be constructed. Preliminary data processing and feature extraction are carried out on edge devices, reducing data transmission volume and enhancing real-time decision-making capabilities.

4.2.3. Communication and interaction of agents

There are various types of agricultural agents with different data formats and communication protocols, leading to difficulties in information exchange. During multi-agent collaboration, conflicts and coordination issues may arise due to task allocation and inconsistent goals. It is necessary to establish unified communication standards and data interface specifications for agricultural agents to ensure smooth interoperability. By introducing a multi-agent collaboration task allocation algorithm, tasks can be efficiently distributed based on each agent's capabilities and task requirements. By establishing an agent conflict resolution mechanism, collaboration conflicts can be resolved through negotiation, arbitration, and other methods.

The complexity and opacity of advanced AI systems pose significant challenges for human oversight, which also applies to the agricultural domain. By combining domain-specific agricultural expertise with LLMs, human experts can offer feedback and guidance to the agent group. In the future, agriculture must develop a human-machine symbiosis framework, implement policy reforms, and deploy scalable oversight tools to ensure effective human supervision and foster the robust development of agricultural intelligence systems (Holzinger et al., 2025).

4.2.4. Dynamic knowledge update and continuous learning

Agricultural knowledge is frequently updated, rendering traditional full-parameter fine-tuning methods inefficient and prone to catastrophic forgetting. The acquisition and integration of new knowledge currently lack automated means, hindering the timeliness and responsiveness of updates. Parameter-efficient fine-tuning techniques, such as adapter tuning and prompt optimization, allow selective updating of key parameters, thereby reducing update costs and mitigating the risk of forgetting previously learned knowledge. An automated system for agricultural knowledge acquisition and updating should be developed to monitor authoritative data sources, such as agricultural research institutions and government portals in real time, automatically extract new knowledge, and update both models and knowledge graphs accordingly.

4.3. Future applications in agriculture

4.3.1. Crop management and optimization

Large language models have demonstrated significant potential in crop management and optimization. By integrating multi-source agricultural data with NLP techniques, LLMs offer farmers precise decision support, enhancing crop yield and quality while optimizing resource use and promoting sustainable agricultural practices. Key application areas include crop health management, planting optimization, and resource allocation. In crop health management, LLMs assists in identifying pests and diseases by analyzing user-provided symptoms, images, and environmental parameters. They provide scientifically grounded prevention and control recommendations, helping to reduce pesticide usage and minimize environmental pollution. For planting optimization, LLMs deliver personalized planting recommendations tailored to crop growth cycles and local environmental conditions, such as optimizing sowing time, fertilization, and irrigation to enhance efficiency and reduce resource waste. Additionally, LLMs support efficient water and fertilizer management. Through analyzing soil moisture and crop water requirements, they recommend optimal irrigation schedules and provide precise fertilization strategies, further advancing sustainable agriculture.

4.3.2. Market analysis and decision making

Market analysis in agriculture involves processing vast amount of diverse data, including fluctuations in agricultural product prices, supply chain changes, consumer demand, policy regulations, and global trends. LLMs, leveraging their powerful NLP capabilities, can extract crucial information from both structured and unstructured data, offering accurate market insights. They help farmers and agricultural enterprises understand real-time market dynamics and predict future trends, enabling more informed production and marketing decisions.

In decision support, LLMs provides personalized advice based on market prices, trend predictions, and climatic factors, assisting in optimizing planting and harvesting schedules, adjusting product mixes, and refining pricing strategies. Additionally, LLMs can analyze potential market risks by identifying key risk factors using historical data and real-time information, such as weather fluctuations, policy changes, or international trade shifts, and offer coping strategies to maintain the stability of agricultural production.

4.3.3. Supply chain and intelligent production management

The application of large language models in agricultural supply chains and intelligent production management is highly significant, driving the transformation of modern agriculture. By integrating real-time data analysis and predictive capabilities, LLMs can optimize agricultural supply chains, improve inventory management, and strengthen supplier relationships. Specific applications include real-time tracking of agricultural product demand, automated performance evaluations, and contract management, among others.

In smart production management, LLM-driven systems enable greenhouse automation, integrating Internet of Things (IoT) sensors for climate, irrigation, and ventilation control to achieve precision farming. By continuously monitoring crop health, LLMs can provide tailored management recommendations, optimizing the use of water, fertilizers, and energy. Additionally, LLMs can integrate with disease management systems, enabling early detection and personalized treatment via image analysis and sensor data, thereby improving disease prevention and management efficiency.

Future research should focus on enhancing model architectures specifically for agriculture, such as developing specialized transformer models for agricultural data, hybrid models combining symbolic agricultural knowledge with neural networks, lightweight variants for resource-constrained environments, and integrated attention mechanisms for crop and soil pattern recognition. Advances in training methods are necessary, including investigating domain-specific pre-training tasks, developing effective fine-tuning methods for limited agricultural datasets, implementing continual learning frameworks to adapt to evolving agricultural knowledge, and designing multi-task learning approaches to meet diverse agricultural goals. Collaboration mechanisms should be established, such as creating a cross-institutional agricultural AI research network, setting standardized agricultural AI benchmarks and evaluation metrics, and developing platforms for shared agricultural data and model resources. Agricultural AI systems require effective validation and deployment strategies, including developing systematic validation protocols, establishing performance metrics for agricultural applications, and designing monitoring systems to ensure long-term model effectiveness.

The key technical challenge of LLMs in agriculture lies in balancing general AI capabilities with domain specificity, requiring breakthroughs in four areas: knowledge fusion reasoning, multimodal processing, agent collaboration, and dynamic learning. By integrating agricultural expertise, optimizing cross-modal reasoning mechanisms, adapting agricultural information and equipment scheduling, and establishing efficient knowledge update systems, LLMs can drive intelligent agricultural questioning, advancing precision agriculture and intelligent decision-making.

5. Conclusion

This paper provides a detailed review of the applications of large language models in enhancing agricultural intelligence, focusing on the technical framework and practical applications in intelligent Q&A systems. The study explores how LLMs can adapt and effectively preserve, and utilize agricultural domain knowledge through architectural design, pre-training tasks, and fine-tuning strategies. By constructing structured knowledge bases and processing unstructured data to provide domain-specific expertise for LLMs, this paper demonstrates their integration with agricultural Q&A systems. It explores methods such as retrieval-based and generative Q&A to enhance professional question-answering capabilities. Moreover, by integrating text, images, sensor data through multimodal learning and knowledge graph technologies, it expands the application boundaries of LLMs, achieving more comprehensive and context-aware responses in agricultural scenarios. The paper discusses the challenges of adapting large models to agricultural knowledge, including knowledge acquisition, fusion, inference, agent collaboration, and dynamic knowledge updates, and proposes targeted solutions. It outlines the future prospects of LLMs in the agricultural field, focusing mainly on their applications in crop management, market analysis, and supply chain optimization. These advancements are expected to promote the development of precision agriculture by enabling data-driven decision-making, resource optimization, and sustainable practices.

Through an in-depth analysis of technological advancements, application modes, and challenges, this paper not only provides a theoretical framework for the intelligence of agriculture but also points out directions for future research. This will help accelerate the application of large language models in the agricultural sector, enhancing agricultural productivity, sustainability, and market decision-making levels. While the application of large language models in agricultural knowledge question and answer can significantly improve production efficiency and decision-making levels, it may also have a series of impacts on agricultural production modes and farmers' decisions, and face challenges related to data privacy and security. Therefore, it is necessary to introduce systematic governance and review mechanisms in model design, deployment, and supervision processes to fully leverage the advantages of large language models while minimizing potential negative impacts on farmers and the agricultural ecosystem.

CRediT authorship contribution statement

Hongda Li: Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Huarui Wu:** Supervision, Funding acquisition. **Qingxue Li:** Writing – review & editing, Supervision. **Chunjiang Zhao:** Validation, Project administration, Funding acquisition.

Funding

This work was supported by Agriculture Research System of China of MOF and MARA projects (Project No.CARS-23-D07) and National Key

Research and Development Program of China projects (Project No.2022YFD1600602).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbasiantaeb, Z., Aliannejadi, M., 2024. Generate then retrieve: conversational response retrieval using LLMs as answer and query generators. arXiv e-prints arXiv-2403.

Alemu, D., Jennex, M., Assefa, T., 2018. Agricultural Knowledge Management System Development for Knowledge Integration. doi:10.24251/HICSS.2018.544.

Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M.G., Bereciartua, A., Alvarez-Gila, A., 2020. Few-shot learning approach for plant disease classification using images taken in the field. Comput. Electron. Agric. 175, 105542. doi:10.1016/j.compag.2020.105542.

Bao, J., Duan, N., Yan, Z., Zhou, M., Zhao, T., 2016. Constraint-based question answering with knowledge graph. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 2503–2514.

Barakat, M., Chung, G.C., Lee, I.E., Pang, W.L., Chan, K.Y., 2023. Detection and sizing of durian using zero-shot deep learning models. Int. J. Technol. 14, 1206–1215. doi:10.14716/ijtech.v14i6.6640.

Biswas, P., Sharan, A., Kumar, A., 2019. Context pattern based agricultural named entity recognition. Res. Comput. Sci. 148, 383–399. doi:10.13053/rcs-148-10-32.

Buehler, E.L., Buehler, M.J., 2024. X-LoRA: Mixture of Low-Rank Adapter Experts, a Flexible Framework for Large Language Models with Applications in Protein Mechanics and Molecular Design. doi:10.48550/arXiv.2402.07148.

Chamorro-Padial, J., Rodrigo-Gines, F.-J., Rodriguez-Sanchez, R., 2024. Finding answers to COVID-19-specific questions: an information retrieval system based on latent keywords and adapted TF-IDF. J. Inf. Sci. 50, 935–951. doi:10.1177/01655515221110995.

Chang, Yupeng, Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Yi, Yu, P.S., Yang, Q., Xie, X., 2023. A survey on evaluation of large language models. ACM Trans. Intell. Syst. Technol. 15, 1–45. doi:10.1145/3641289.

Chen, X., Jia, S., Xiang, Y., 2020. A review: knowledge reasoning over knowledge graph. Expert Syst. Appl. 141, 112948. doi:10.1016/j.eswa.2019.112948.

Chen, M., Tworek, J., Jun, H., Yuan, Q., de Pinto, H.P.O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., 2021. Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374.

Chen, J., Qi, Y., Wang, Y., Pan, G., 2023. MindGPT: Interpreting What you See with Non-invasive Brain Recordings.

Chunjiang, Z., Jingchen, L.I., Huarui, W.U., Yusen, Y., 2024. Vegetable crop growth modeling in digital twin platform based on large language model inference. Smart Agriculture 6, 63.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.

Dhavale, C., Pawar, T., Singh, A., Pole, S., Sabat, K., 2024. Revolutionizing farming: Gan-enhanced imaging, cnn disease detection, and llm farmer assistant. 2024 2nd International Conference on Computer, Communication and Control (IC4). IEEE, pp. 1–6 doi:10.1109/IC457434.2024.10486501.

Didwania, K., Seth, P., Kasliwal, A., Agarwal, A., 2024. AgriLLM: Harnessing Transformers for Farmer Queries. doi:10.48550/arXiv.2407.04721.

Ding, N., Tang, Y., Fu, Z., Xu, C., Han, K., Wang, Y., 2023. GPT4Image: Can Large Pre-Trained Models Help Vision Models on Perception Tasks?

Ding, D., Fu, X., Peng, X., Fan, X., Huang, H., Zhang, B., 2024. Leveraging chain-of-thought to enhance stance detection with prompt-tuning. Mathematics 12, 568. doi:10.3390/math12040568.

Figueroa, A., Atkinson, J., 2011. Maximum entropy context models for ranking biographical answers to open-domain definition questions. Proceedings of the AAAI Conference on Artificial Intelligence., pp. 1173–1179

Gao, P., Geng, S., Zhang, R., Ma, T., Fang, R., Zhang, Y., Li, H., Qiao, Y., 2023a. CLIP-adapter: better vision-language models with feature adapters. Int. J. Comput. Vis. doi:10.1007/s11263-023-01891-x.

Gao, P., Han, J., Zhang, R., Lin, Z., Geng, S., Zhou, A., Zhang, W., Lu, P., He, C., Yue, X., Li, H., Qiao, Y., 2023b. LLaMA-adapter V2: parameter-efficient visual instruction model. arXiv preprint arXiv:2304.15010.

Gao, Z., Wang, Q., Chen, A., Liu, Z., Wu, B., Chen, L., Li, J., 2024. Parameter-efficient fine-tuning with discrete Fourier transform. arXiv e-prints arXiv:2304.15010.

García-García, J.I., Marín-Aragón, D., Maciá, H., Jiménez-Cantizano, A., 2021. Viñamecum: a computer-aided method for diagnoses of pests and diseases in the vineyard. Appl. Sci. 11, 4704. doi:10.3390/app11104704.

Ge, J., Luo, H., Qian, S., Gan, Y., Fu, J., Zhang, S., 2023. Chain of Thought Prompt Tuning in Vision Language Models.

Gulati, V., Kumar, D., Popescu, D.E., Hemanth, J.D., 2023. Extractive article summarization using integrated TextRank and BM25+ algorithm. Electronics 12, 372. doi:10.3390/electronics12020372.

Guo, X., Wang, J., Gao, G., Zhou, J., Li, Y., Cheng, Z., Miao, G., 2024. Efficient agricultural question classification with a BERT-enhanced DPCNN model. IEEE Access 12, 109255–109268. doi:10.1109/ACCESS.2024.3438848.

Hao, X., Wang, L., Zhu, H., Guo, X., 2023. Joint agricultural intent detection and slot filling based on enhanced heterogeneous attention mechanism. Comput. Electron. Agric. 207, 107756. doi:10.1016/j.compag.2023.107756.

Haval, A.M., Mishra, A., 2024. Real time machine learning based voltage regulation model for smart agriculture. E3S Web of Conferences. EDP Sciences, p. 13005.

Herindra, L.D., Syafei, W.A., Wibowo, A., 2023. Agricultural irrigation information system in internet of things (IoT) based screen house using rule based expert system algorithm. AIP Conference Proceedings. AIP Publishing doi:10.1063/5.0140193.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., de Casas, D.L., Hendricks, L.A., Welbl, J., Clark, A., 2022. Training compute-optimal large language models. arXiv preprint arXiv:2203.15556.

Holzinger, A., Fister Jr., I., Fister Sr., I., Kaul, H.-P., Asseng, S., 2024. Human-Centered AI in smart farming: toward agriculture 5.0. IEEE Access 12, 62199–62214. doi:10.1109/ACCESS.2024.3395532.

Holzinger, A., Zatloukal, K., Mueller, H., 2025. Is human oversight to AI systems still possible? New Biotechnol. 85, 59–62. doi:10.1016/j.nbt.2024.12.003.

Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., 2021. LoRA: low-rank adaptation of large language models. ICLR 1, 3.

Huang, X., Zhang, J., Li, D., Li, P., 2019. Knowledge graph embedding based question answering. Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 105–113 doi:10.1145/3289600.3290956.

Huang, Y., Liu, J., Lv, C., 2023. Chains-BERT: a high-performance semi-supervised and contrastive learning-based automatic question-and-answering model for agricultural scenarios. Appl. Sci.-Basel 13, 2924. doi:10.3390/app13052924.

Hyeon-Woo, N., Ye-Bin, M., Oh, T.-H., 2023. FedPara: low-rank Hadamard product for communication-efficient federated learning. arXiv preprint arXiv:2108.06098.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment anything. Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026.

Koh, J.Y., Salakhutdinov, R., Fried, D., 2023. Grounding Language Models to Images for Multimodal Inputs and Outputs.

Kojima, T., Gu, S.S., Reid, M., Matsuo, Y., Iwasawa, Y., 2023. Large Language Models are Zero-Shot Reasoners. doi:10.48550/arXiv.2205.11916.

Kopiczko, D.J., Blankevoort, T., Asano, Y.M., 2024. VeRA: Vector-Based Random Matrix Adaptation.

Kpodo, J., Kordjamshidi, P., Nejadhashemi, A.P., 2024. AgXQA: a benchmark for advanced agricultural extension question answering. Comput. Electron. Agric. 225, 109349. doi:10.1016/j.compag.2024.109349.

Kraisnikovic, C., Harb, R., Plass, M., Al Zoughbi, W., Holzinger, A., Mueller, H., 2025. Fine-tuning language model embeddings to reveal domain knowledge: an explainable artificial intelligence perspective on medical decision making. Eng. Appl. Artif. Intell. 139, 109561. doi:10.1016/j.engappai.2024.109561.

Kung, H.-Y., Yu, R.-W., Chen, C.-H., Tsai, C.-W., Lin, C.-Y., 2021. Intelligent pig-raising knowledge question-answering system based on neural network schemes. Agron. J. 113, 906–922. doi:10.1002/agj2.20622.

Lahat, D., Adali, T., Jutten, C., 2015. Multimodal data fusion: an overview of methods, challenges, and prospects. Proc. IEEE 103, 1449–1477. doi:10.1109/JPROC.2015.2460697.

Lai, T.M., Zhai, C., Ji, H., 2023. KEBLM: knowledge-enhanced biomedical language models. J. Biomed. Inform. 143, 104392. doi:10.1016/j.jbi.2023.104392.

Lan, Y., Guo, Y., Chen, Q., Lin, S., Chen, Y., Deng, X., 2023. Visual question answering model for fruit tree disease decision-making based on multimodal deep learning. Front. Plant Sci. 13, 1064399. doi:10.3389/fpls.2022.1064399.

Lester, B., Al-Rfou, R., Constant, N., 2021. The Power of Scale for Parameter-Efficient Prompt Tuning.

Li, X.L., Liang, P., 2021. Prefix-Tuning: Optimizing Continuous Prompts for Generation.

Li, Z., Cao, Z., Li, P., Zhong, Y., Li, S., 2023a. Multi-hop question generation with knowledge graph-enhanced language model. Appl. Sci.-Basel 13, 5765. doi:10.3390/app13095765.

Li, J., Li, D., Savarese, S., Hoi, S., 2023b. BLIP-2: Bootstrapping Language-Image Pre-Training with Frozen Image Encoders and Large Language Models.

Li, B., Zhang, Y., Chen, L., Wang, J., Pu, F., Yang, J., Li, C., Liu, Z., 2023c. MIMIC-IT: Multi-Modal in-Context Instruction Tuning.

Li, Y., Han, S., Ji, S., 2024a. VB-LoRA: Extreme Parameter Efficient Fine-Tuning with Vector Banks.

Liang, Z., Fang, Z., Huang, H., Wang, Z., Hong, Y., Liu, K., Shang, P., 2023. Engineering exploration and research on large language models. Proceedings of the 2023 5th International Conference on Internet of Things, Automation and Artificial Intelligence. Presented at the IoTAAI 2023: 2023 5th International Conference on Internet of Things, Automation and Artificial Intelligence, ACM, Nanchang China, pp. 174–179 doi:10.1145/3653081.3653111.

Liu, V., Chilton, L.B., 2022. Design guidelines for prompt engineering text-to-image generative models. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–23.

Liu, X., Ji, K., Fu, Y., Tam, W.L., Du, Z., Yang, Z., Tang, J., 2022a. P-tuning: prompt tuning can be comparable to fine-tuning across scales and tasks. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022): (Short Papers), Vol 2. Presented at the 60th Annual meeting of the association-for-computational-Linguistics (ACL), Assoc computational Linguistics-ACL, Stroudsburg, pp. 61–68.

Liu, H., Tam, D., Muqeeth, M., Mohta, J., Huang, T., Bansal, M., Raffel, C.A., 2022b. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. Adv. Neural Inf. Proces. Syst. 35, 1950–1965.

Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., Tang, J., 2023. GPT understands, too. AI Open doi:10.1016/j.aiopen.2023.08.012.

Lu, Y., Lu, X., Zheng, L., Sun, M., Chen, S., Chen, B., Wang, T., Yang, J., Lv, C., 2024. Application of multimodal transformer model in intelligent agricultural disease detection and question-answering systems. Plants-Basel 13, 972. doi:10.3390/plants13070972.

Luo, H., Sun, Q., Xu, C., Zhao, P., Lou, J., Tao, C., Geng, X., Lin, Q., Chen, S., Zhang, D., 2023. Wizardmath: empowering mathematical reasoning for large language models via reinforced evol-instruct. arXiv preprint arXiv:2308.09583.

Maksimovich, K.Yu., Kalichkin, V.K., Fedorov, D.S., Aleschenko, V.V., 2024. Application of machine learning methods for crop rotation selection in organic farming system. E3S Web Conf. 486, 01028. doi:10.1051/e3sconf/202448601028.

McCoy, R.T., Yao, S., Friedman, D., Hardy, M., Griffiths, T.L., 2023. Embers of Autoregression: Understanding Large Language Models Through the Problem They are Trained to Solve. doi:10.48550/arXiv.2309.13638.

Mendoza-Revilla, J., Trop, E., Gonzalez, L., Roller, M., Dalla-Torre, H., de Almeida, B.P., Richard, G., Caton, J., Carranza, N.L., Skwark, M., Laterre, A., Beguir, K., Pierrot, T., Lopez, M., 2024. A foundational large language model for edible plant genomes. Commun. Biol. 7, 835. doi:10.1038/s42003-024-06465-2.

Miao, J., Thongprayoon, C., Suppadungsuk, S., Krisanapan, P., Radhakrishnan, Y., Cheungpasitporn, W., 2024. Chain of thought utilization in large language models and application in nephrology. Medicina 60, 148. doi:10.3390/medicina60010148.

Moiseev, F., Dong, Z., Alfonseca, E., Jaggi, M., 2022. SKILL: Structured Knowledge Infusion for Large Language Models.

Mtega, W.P., Ngoepe, M., 2019. A framework for strengthening agricultural knowledge systems for improved accessibility of agricultural knowledge in Morogoro region of Tanzania. J. Librariansh. Inf. Sci. 51, 629–642. doi:10.1177/0961000617742456.

Mu, Y., Zhang, Q., Hu, M., Wang, W., Ding, M., Jin, J., Wang, B., Dai, J., Qiao, Y., Luo, P., 2023. EmbodiedGPT: Vision-Language Pre-Training Via Embodied Chain of Thought.

Neethirajan, S., 2024. Decoding the Language of Chickens - An Innovative NLP Approach to Enhance Poultry Welfare. doi:10.1101/2024.04.29.591707.

Nidamanuri, R.R., 2024. Deep learning-based prediction of plant height and crown area of vegetable crops using LiDAR point cloud. Sci. Rep. 14, 14903. doi:10.1038/s41598-024-65322-8.

Nismi Mol, E.A., Santosh Kumar, M.B., 2024. End-to-end framework for agricultural entity extraction – a hybrid model with transformer. Comput. Electron. Agric. 225, 109309. doi:10.1016/j.compag.2024.109309.

OpenAI, 2023. GPT-4 Technical Report arXiv preprint arXiv:2303.08774.

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., Wu, X., 2024. Unifying large language models and knowledge graphs: a roadmap. IEEE Trans. Knowl. Data Eng. 36, 3580–3599. doi:10.1109/TKDE.2024.3352100.

Ponti, E.M., Sordoni, A., Bengio, Y., Reddy, S., 2022. Combining Modular Skills in Multitask Learning.

Qian, Y., Chen, X., Wang, Y., Zhao, J., Ouyang, D., Dong, S., Huang, L., 2023. Agricultural text named entity recognition based on the BiLSTM-CRF model. Fifth International Conference on Computer Information Science and Artificial Intelligence (CISAI 2022). SPIE, pp. 525–530 doi:10.1117/12.2667761.

Qian, S., Liu, B., Sun, C., Xu, Z., Ma, L., Wang, B., 2024. CroMIC-QA: the cross-modal information complementation based question answering. IEEE Trans. Multimedia 26, 8348–8359. doi:10.1109/TMM.2023.3326616.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I., 2021. Learning Transferable Visual Models from Natural Language Supervision.

Rahmani, K., Raza, M., Gulwani, S., Le, V., Morris, D., Radhakrishna, A., Soares, G., Tiwari, A., 2021. Multi-modal program inference: a marriage of pre-trained language models and component-based synthesis. Proc. ACM Program. Lang. 5, 1–29. doi:10.1145/3485535.

Rathnayake, H., Sumanapala, J., Rukshani, R., Ranathunga, S., 2022. Adapter-based fine-tuning of pre-trained multilingual language models for code-mixed and code-switched text classification. Knowl. Inf. Syst. 64, 1937–1966. doi:10.1007/s10115-022-01698-1.

Sarsa, S., Denny, P., Hellas, A., Leinonen, J., 2022. Automatic generation of programming exercises and code explanations using large language models. Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 1. Presented at the ICER 2022: ACM Conference on International Computing Education Research. ACM, Lugano and Virtual Event Switzerland, pp. 27–43 doi:10.1145/3501385.3543957.

Sitokonstantinou, V., Porras, E.D.S., Bautista, J.C., Piles, M., Athanasiadis, I., Kerner, H., Martini, G., Sweet, L., Tsoumas, I., Zscheischler, J., Camps-Valls, G., 2024. Causal Machine Learning for Sustainable Agroecosystems. doi:10.48550/arXiv.2408.13155.

Sui, Y., He, Y., Liu, N., He, X., Wang, K., Hooi, B., 2024. FiDeLiS: Faithful Reasoning in Large Language Model for Knowledge Graph Question Answering. doi:10.48550/arXiv.2405.13873.

Sun, A., Jiang, M., Ma, Y., 2006. A maximum entropy model based answer extraction for Chinese question answering. Fuzzy Systems and Knowledge Discovery: Third International Conference, FSKD 2006, Xi'an, China, September 24-28, 2006. Proceedings 3. Springer, pp. 1239–1248.

Suzuki, J., Sasaki, Y., Maeda, E., 2002. SVM answer selection for open-domain question answering. COLING 2002: The 19th International Conference on Computational Linguistics doi:10.3115/1072228.1072347.

Tang, R., Yang, J., Tang, J., Aridas, N.K., Talip, M.S.A., 2024. Design of agricultural question answering information extraction method based on improved BILSTM algorithm. Sci. Rep. 14, 1–12. doi:10.1038/s41598-024-70534-z.

Thirunavukarasu, A.J., Ting, D.S.J., Elangovan, K., Gutierrez, L., Tan, T.F., Ting, D.S.W., 2023. Large language models in medicine. Nat. Med. 1–11.

Thorat, S.A., Jadhav, V., 2020. A review on implementation issues of rule-based chatbot systems. Proceedings of the International Conference on Innovative Computing & Communications (ICICC) doi:10.2139/ssrn.3567047.

Tzachor, A., Devare, M., Richards, C., Pypers, P., Ghosh, A., Koo, J., Johal, S., King, B., 2023. Large language models and agricultural extension services. Nat. Food 1–8.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Lukasz, Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Proces. Syst. 30.

Veena, G., Kanjirangat, V., 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.

Vizniuk, A., Diachenko, G., Laktionov, I., Siwocha, A., Xiao, M., Smoląg, J., 2025. A comprehensive survey of retrieval-augmented large language models for decision making in agriculture: unsolved problems and research opportunities. J. Artific. Intellig. Soft Comp. Res. 15, 115–146. doi:10.2478/jaiscr-2025-0007.

Wang, H., Zhao, R., 2024. Knowledge graph of agricultural engineering technology based on large language model. Displays 85, 102820. doi:10.1016/j.displa.2024.102820.

Wang, J., Liu, Z., Zhao, L., Wu, Z., Ma, C., Yu, S., Dai, H., Yang, Q., Liu, Y., Zhang, S., 2023a. Review of large vision models and visual prompt engineering. Meta-Radiology doi:10.1016/j.metrad.2023.100047 100047.

Wang, J., Shi, E., Yu, S., Wu, Z., Ma, C., Dai, H., Yang, Q., Kang, Y., Wu, J., Hu, H., Yue, C., Zhang, H., Liu, Y., Pan, Y., Liu, Z., Sun, L., Li, X., Ge, B., Jiang, X., Zhu, D., Yuan, Y., Shen, D., Liu, T., Zhang, S., 2024. Prompt Engineering for Healthcare: Methodologies and Applications. doi:10.48550/arXiv.2304.14670.

Wrzecińska, M., Czerniawska-Piątkowska, E., Kowalewska, I., Kowalczyk, A., Mylostyvyi, R., Stefaniak, W., 2023. Agriculture in the face of new digitization technologies. UBSRAS 3, 9–17. doi:10.56407/bs.agrarian/3.2023.09.

Xia, Y., Sun, N., Wang, H., Yuan, X., Gu, L., Wang, C., Gao, Q., 2020. Research on knowledge question answering system for agriculture disease and pests based on knowledge graph. J. Nonlinear Convex Anal. 21, 1487–1496.

Yang, T., Mei, Y., Xu, L., Yu, H., Chen, Y., 2024a. Application of question answering systems for intelligent agriculture production and sustainable management: a review. Resour. Conserv. Recycl. 204, 107497. doi:10.1016/j.resconrec.2024.107497.

Yang, S., Yuan, Z., Li, S., Peng, R., Liu, K., Yang, P., 2024b. GPT-4 as Evaluator: Evaluating Large Language Models on Pest Management in Agriculture. doi:10.48550/arXiv.2403.11858.

Yang, Z., 2019. XLNet: generalized autoregressive pretraining for language understanding. Advances in neural information processing systems 32.

Yang, S., Liu, Z., Mayer, W., Ding, N., Wang, Y., Huang, Y., Wu, P., Li, W., Li, L., Zhang, H.-Y., Feng, Z., 2025. ShizishanGPT: An agricultural large language model integrating tools and resources. In: Barhamgi, M., Wang, H., Wang, X. (Eds.), Web Information Systems Engineering – WISE 2024, Lecture Notes in Computer Science. Springer Nature Singapore, Singapore, pp. 284–298 doi:10.1007/978-981-96-0573-6_21.

Yao, X., Hao, X., Liu, R., Li, L., Guo, X., 2024. AgCNER, the first large-scale Chinese named entity recognition dataset for agricultural diseases and pests. Sci. Data 11, 769. doi:10.1038/s41597-024-03578-5.

Yasunaga, M., Ren, H., Bosselut, A., Liang, P., Leskovec, J., 2021. QA-GNN: reasoning with language models and knowledge graphs for question answering. arXiv preprint arXiv:2104.06378.

Yeh, S.-Y., Hsieh, Y.-G., Gao, Z., Yang, B.B., Oh, G., Gong, Y., 2023. Navigating text-to-image customization: From lycoris fine-tuning to model evaluation. The Twelfth International Conference on Learning Representations.

Yi, W., Zhang, L., Kuzmin, S., Gerasimov, I., Liu, M., 2025. Agricultural large language model for standardized production of distinctive agricultural products. Comput. Electron. Agric. 234, 110218. doi:10.1016/j.compag.2025.110218.

Yuan, Y., Zhan, Y., Xiong, Z., 2023. Parameter-efficient transfer learning for remote sensing image-text retrieval. IEEE Trans. Geosci. Remote Sens. 61, 5619014. doi:10.1109/TGRS.2023.3308969.

Zaib, M., Zhang, W.E., Sheng, Q.Z., Mahmood, A., Zhang, Y., 2022. Conversational question answering: a survey. Knowl. Inf. Syst. 64, 3151–3195. doi:10.1007/s10115-022-01744-y.

Zha, Z., Wang, J., Li, Z., Zhu, X., Song, W., Xiao, Y., 2023. M2ConceptBase: a fine-grained aligned multi-modal conceptual Knowledge Base. CoRR. doi:10.48550/ARXIV.2312.10417.

Zhan, Y., Yuan, Y., Xiong, Z., 2023. Mono3dvg: 3d visual grounding in monocular images. in: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 6988–6996.

Zhan, Y., Xiong, Z., Yuan, Y., 2025. SkyEyeGPT: unifying remote sensing vision-language tasks via instruction tuning with large language model. ISPRS-J. Photogramm. Remote Sens. 221, 64–77. doi:10.1016/j.isprsjprs.2025.01.020.

Zhang, Y., Dai, H., Kozareva, Z., Smola, A., Song, L., 2018. Variational reasoning for question answering with knowledge graph. Proceedings of the AAAI Conference on Artificial Intelligence.

Zhang, J., Guo, M., Geng, Y., Li, M., Zhang, Y., Geng, N., 2021. Chinese named entity recognition for apple diseases and pests based on character augmentation. Comput. Electron. Agric. 190, 106464. doi:10.1016/j.compag.2021.106464.

Zhang, Y., Ji, Q., Xu, X., Cheng, Z., Xiao, G., 2023. Knowledge graph relation path network for multi-hop intelligent question answering. Acta Electonica Sinica 1–8.

Zhang, R., Han, J., Liu, C., Gao, P., Zhou, A., Hu, X., Yan, S., Lu, P., Li, H., Qiao, Y., 2023. LLaMA-Adapter: Efficient Fine-Tuning of Language Models with Zero-Init Attention.

Zhao, B., Tu, H., Wei, C., Mei, J., Xie, C., 2023. Tuning LayerNorm in Attention: Towards Efficient Multi-Modal LLM Finetuning.

Zheng, W., Yu, J.X., Zou, L., Cheng, H., 2018. Question answering over knowledge graphs: question understanding via template decomposition. Proc. VLDB Endowm. 11, 1373–1386. doi:10.14778/3236187.3236192.
Zheng, W., Zheng, K., Gao, L., Zhangzhong, L., Lan, R., Xu, L., Yu, J., 2024. GRU–transformer: a novel hybrid model for predicting soil moisture content in root zones. Agronomy 14, 432. doi:10.3390/agronomy14030432.
Zhou, K., Yang, J., Loy, C.C., Liu, Z., 2022. Learning to prompt for vision-language models. Int. J. Comput. Vis. 130, 2337–2348. doi:10.1007/s11263-022-01653-1.
Zhou, P., Min, W., Fu, C., Jin, Y., Huang, M., Li, X., Mei, S., Jiang, S., 2024. FoodSky: A Food-oriented Large Language Model that Passes the Chef and Dietetic Examination. doi: 10.48550/arXiv.2406.10261.
Zhu, H., Shi, W., Guo, X., Lyu, S., Yang, R., Han, Z., 2025. Potato disease detection and prevention using multimodal AI and large language model. Comput. Electron. Agric. 229, 109824. doi:10.1016/j.compag.2024.109824.