Resources, Conservation & Recycling 204 (2024) 107497

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.elsevier.com/locate/resconrec

Application of question answering systems for intelligent agriculture production and sustainable management: A review

Tian Yang [a,c,d,e], Yupeng Mei [a,c,d,e], Ling Xu [a,c,d,e], Huihui Yu [a,b,c,*], Yingyi Chen [a,c,d,e]

[a] National Innovation Center for Digital Fishery, Beijing 100083, PR China
[b] School of Information Science & Technology, Beijing Forestry University, 35 Tsinghua East Road, Beijing 100083, PR China
[c] Key Laboratory of Smart Farming Technologies for Aquatic Animal and Livestock, Ministry of Agriculture and Rural Affairs, Beijing 100083, PR China
[d] Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing 100083, PR China
[e] College of Information and Electrical Engineering, China Agricultural University, 17 Tsinghua East Road, Beijing 100083, PR China

ARTICLE INFO

Keywords:
Question answering system
Intelligent agriculture
Knowledge graphs
Large language models

ABSTRACT

The increasing application of artificial intelligence in agriculture production and management has generated a large amount of data, leading to a demand for processing this data. This review focuses on the knowledge storage approaches in agricultural question answering systems, namely corpora, knowledge graphs, and large language models. These systems are built on massive amounts of data and aim to process and retrieve information effectively in the context of sustainable agriculture. Corpora refer to large collections of diverse documents that serve as foundational resources for training and fine-tuning question answering systems. Knowledge graphs capture structured and interconnected knowledge by representing entities, relationships, and attributes, enabling efficient organization and querying of information. Large language models, such as GPT-4, enhance the capacity of question answering systems to provide accurate and relevant responses. By exploring these three prominent knowledge storage approaches, this review analyses the methodology and impact of agricultural question answering systems, highlighting their applications in the production process. The findings provide important implications for future research in agriculture, and potential directions for further exploration.

1. Introduction

Extracting targeted decision-making information from unmanned farms (Wang et al., 2021b) has become a crucial technology in the advancement of intelligent agriculture and unmanned farm management. Question answering systems, as a key technology in this process, can play a pivotal role. Question answering system is a computer program designed to provide user with answers to their queries. The system utilizes natural language processing and machine learning algorithms to comprehend the context of a question and provide accurate and relevant answers. Compared to other information retrieval service such as search engine, question answering system provides a more precise and concise response to user queries which significantly reduces the time and effort required to derive appropriate information (Jin et al., 2023a; Zhu et al., 2021). Like in other fields, question answering systems in agriculture can also play a crucial role in advancing this sector. However, the development of question answering systems within the agricultural field is encountering some difficulties that limit its application. Some of these challenges are the lack of data standardization across various agricultural domains and the highly variable nature of production environments.

The question answering systems is a complex system based on various technologies such as document retrieval, knowledge graph and neural network. The knowledge storage architectures for question answering methods mainly includes three types: corpora, knowledge graphs, and large language models. Corpora based methods are used to retrieve relevant information from documents or datasets using keyword-based techniques (Izacard and Grave, 2021; Karpukhin et al., 2020; Wang et al., 2022c). Knowledge based methods rely on structured knowledge representations such as knowledge graphs or ontologies to answer the question (Gu et al., 2022; Jin et al., 2023a; Lan et al., 2021). Large language model based methods use deep learning models to enable the system to understand the context of the question and the related text corpus (Brown et al., 2020; OpenAI, 2023; Touvron et al., 2023). Deep learning technology is a type of artificial intelligence that utilizes large-scale neural networks to identify patterns in data. It has

* Corresponding authors at: National Innovation Center for Digital Fishery, Beijing 100083, PR China.
E-mail addresses: yuhh1990@bjfu.edu.cn (H. Yu), chenyingyi@cau.edu.cn (Y. Chen).

https://doi.org/10.1016/j.resconrec.2024.107497
Received 31 October 2023; Received in revised form 11 February 2024; Accepted 12 February 2024
Available online 22 February 2024
0921-3449/

many advantages over traditional machine learning techniques, such as higher accuracy, faster processing speed, and the ability to handle large amounts of complex data. In recent years, deep learning has been gradually applied in various fields of agricultural question and answering systems, and achieves impressive results. It has been proved to be beneficial in identifying diseases and pests, improving crop yields and reducing the amount of pesticides used (Kung et al., 2021; Lan et al., 2023; Rose Mary et al., 2021; Wang et al., 2021a). These advancements prove that deep learning is a valuable tool with a wide range of applications.

Compared with other fields in the technology industry (Chalkidis et al., 2021; Weiqiang et al., 2023; Pal et al., 2022; Yuan et al., 2022), question answering systems in the agricultural field have unique characteristics and hold significant potential. First of all, like all other vertical domains, agricultural question answering systems requires a deep understanding of specific domain knowledge (Lun and Hui, 2022), including various aspects of expertise such as crop cultivation, livestock management, soil science, and pest control. Moreover, agriculture is greatly influenced by seasonal and regional variations. Agricultural question answering systems need to provide specific answers based on factors such as climate, soil types, crop rotation, and local agricultural practices (Raj et al., 2022). Finally, there is a large amount of sensor data in agricultural scenarios. With the increasing adoption of precision agriculture technology, question answering systems in agriculture can integrate data from various sensors and IoT devices (Marinello et al., 2023). This integration enables the system to provide real-time insights and recommendations based on data collected from fields, weather stations, drones, and other agricultural monitoring tools.

Based on the above contents, the question answering systems have been applied to a wide variety of agriculture production and sustainable management. For agricultural-related knowledge and language, a question answering system in this field must be able to understand specific agricultural terms and concepts, handle various agricultural data sources, and provide appropriate answers based on the actual production conditions in agricultural scenarios. However, there is a lack of comprehensive surveys in the literature regarding agricultural question answering systems. In recent years, with the development of language models, various methods for agricultural question answering have emerged, highlighting the need for systematic research to integrate and unify these approaches.

Accordingly, we explore three prominent knowledge storage approaches: corpora, knowledge graphs, and large language models. The content is organized as follows. Section 2 systematically introduces the different knowledge storage architectures for question answering. In Section 3, the practical application of agricultural question answering systems in the production process in recent years is discussed. Finally, Section 4 highlights the different critical challenges posed by agricultural question answering systems.

2. Question answering system

Question answering systems are built on massive amounts of data and can generate coherent responses based on the input context data. However, the raw data is often diverse and can consist of books, articles, websites, and more. To unify and utilize these data more efficiently, different data pre-processing and organization schemes are proposed. The question answering system of agriculture mainly contains three types of models, corpora, knowledge graphs and large language models. Corpora refer to large collections of documents or texts that contain vast amounts of information. These documents are often diverse and can consist of books, articles, websites, and more. They act as the foundational resource for training and fine-tuning question answering systems. Knowledge Graphs, on the other hand, capture structured and interconnected knowledge by representing entities, relationships, and attributes. They provide a way to organize and query information effectively. Large Language Models, such as GPT-3, are essential components that enhance the capacity of question answering systems. Together, these components create a comprehensive architecture that empowers question answering systems to retrieve and provide accurate and relevant information to user queries.

2.1. Question answering system based on corpora in agriculture

The question answering system based on corpora is a classic question answering method generally involves three main parts: question processing, document processing and answer generation (Calijorne Soares and Parreiras, 2020) as shown in Fig. 1. The question processing stage involves understanding the input question and analysing its structure to identify its underlying meaning. Document processing stage is the key technology of this type of method, which involves locating pertinent information from a knowledge base or various external sources to yield all relevant information needed to answer the question. Finally, answer generation involves synthesizing the retrieved information into a coherent and accurate answer that speaks to the input question. These three parts essentially work in unison to develop an effective and comprehensive question answering method.

2.1.1. Information retrieval

Information retrieval for question answering method retrieves information from a large corpus of documents in response to a user query, and then sorts retrieval results by relevance. According to the representation of documents, information retrieval methods used in these methods can be divided into two main types, text-based and embedding-based information retrieval.

One hot encoding is one of the text-based methods which represents each word in a document query using a binary vector, where each bit corresponds to the presence or absence of a word. Global text-based methods such as the latent semantic indexing (LSA) algorithm (Deerwester et al., 1990) or latent Dirichlet allocation (LDA) algorithm (Campbell et al., 2001) take into account the co-occurrences of words, phrases and topics at the level of documents. Another commonly used method is the BM25 algorithm (Robertson and Zaragoza, 2009), which considers the relevance score of every word in a document, and the frequency of each word in both the query and the document. Term frequency-inverse document frequency (TF-IDF) algorithm (Ramos, 2003) counts the significance of a word increases proportionally with the number of times it appears in the corpus but is offset by the frequency of the word in the entire corpus.

Since text-based information retrieval methods search the corpus of documents for exact matches or relevant keywords, these methods may miss relevant information due to differences in phrasing or context. Embedding-based information retrieval methods can identify semantic relationships between words and concepts, which allows the system to find relevant information even if it does not contain the exact keywords used in the query. Word2vec (Mikolov et al., 2013) learns distributed vector representations of words and phrases in a large corpus, by training continuous bag of words (CBOW) and skip-gram (SG) to predict the context words or the target words of a given word or phrase in the corpus. GloVe (Global Vectors for Word Representation) (Pennington et al., 2014) is a distributional semantics model that has shown to outperform WordNet and other traditional methods for generating word embeddings by incorporating both local and global context information.

2.1.2. Machine reading comprehension

Some researchers (Liu et al., 2019) consider machine reading comprehension as a kind of specific question answering task, in which each question is given in the related context. When documents candidate from information retrieval method is viewed as the context, machine learning method can be a part of question answering system. In the process of machine reading comprehension, the pre-trained static word embedding model is a key step. The Embeddings from Language Models (ELMO) (Peters et al., 2018) method is one of the earliest types of

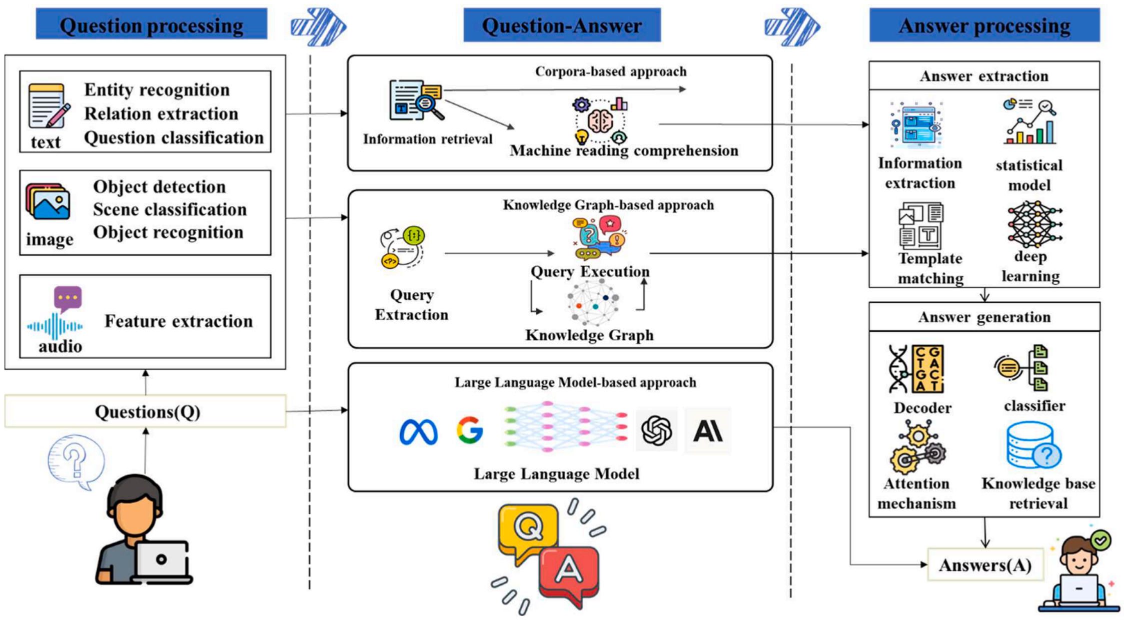

Fig. 1. Architecture of question answering method.

pre-trained word models. It is used to generate dynamic word vectors for pre-trained language model. Its core technique is to use a bi-directional LSTM model to train a dynamic word vector model by predicting the next word in the forward and reversing direction of the text sequence respectively. This type of model has solved the complex problem of polysemy. The Bidirectional Encoder Representations from Transformer (BERT) (Devlin et al., 2018) model is based on Transformer networks with self-attention, which increases the depth of the network and can extract the relationship features of words in sentences at the same time. Moreover, it can extract relationship features at different levels. Compared with traditional RNN models, it solves the problem of long-distance dependencies. Compared to Word2vec, it can obtain word meanings based on the context of the sentence, thus avoiding ambiguity.

2.2. Knowledge graphs in agriculture

Standardizing and formalizing the representation of agricultural data is important for applying question answering systems in agricultural area. Agricultural data can be complex and diverse, and lacks a standardized format, making it difficult to analyse and process using automated systems. By organizing and structuring agricultural data in a formalized way, employing techniques such as name entity recognition, literature analysis and knowledge graphs, the integration of agricultural data can be facilitated into automated question answering systems. As shown in Fig. 2, it shows the steps of constructing agricultural knowledge graphs before integrating knowledge graphs with question answering systems.

2.2.1. Construction of agricultural knowledge graph

In agricultural domain, name entity recognition and ontology serve crucial roles in applications that involve automated data processing such as precision agriculture and agricultural knowledge management. Name entity recognition (NER) seeks to identify and classify entities such as crop varieties, pests, and diseases from large volumes of text. These entities provide valuable information for decision-making processes and aid in the development of knowledge-based systems. Ontology, on the other hand, represents a formal model of the knowledge in a particular domain, including the relationships between entities. The ontology can

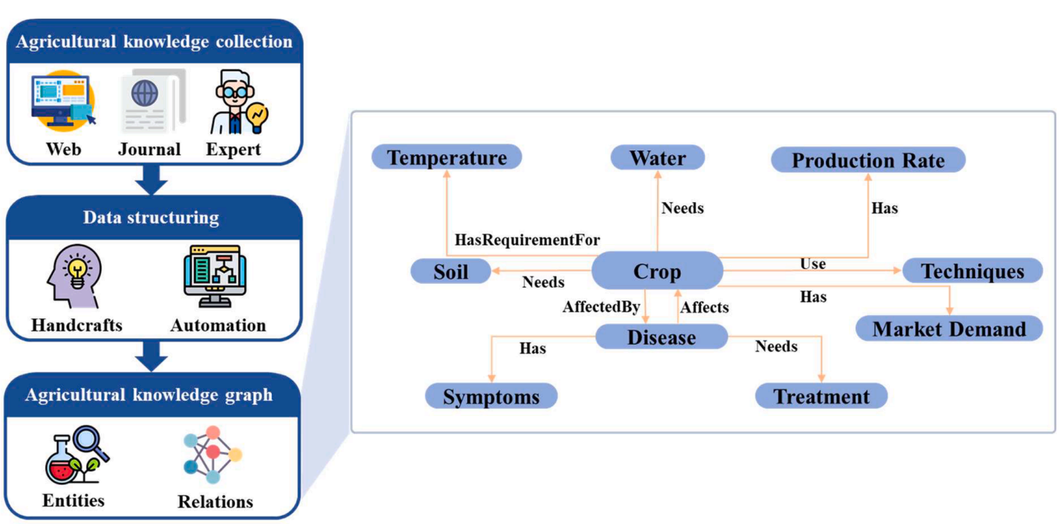

Fig. 2. Construction of agricultural knowledge graph.

facilitate information sharing and interoperability between different applications and systems within the agricultural domain.

Traditional name entity recognition methods (Biswas et al., 2019; Malik et al., 2015) are typically rule-based or statistical and rely on handcrafted features and dictionaries to recognize entities in text. While these methods have been successful in identifying entities in various domains, the challenge remains to achieve high accuracy and efficiency in recognizing entities in the agricultural domain, which involves dealing with complex and domain-specific terminologies.

To bridge the gap between agricultural product recognition and knowledge graph technologies, researchers have explored innovative approaches to enhance the performance of NER models (Liang et al., 2023; Liu et al., 2023; Wang et al., 2022a). One such effort was made by Zhang L and their team (Zhang et al., 2022), who developed a novel Chinese agricultural NER (CANER) model called KIWINER. This model addressed the limitations posed by common word segmentation tools and the feature extraction capabilities of the sequence encoding layer. By incorporating statistics-based new word detection and two novel modules, namely AttSoftlexicon (Criss-cross attention-based Softlexicon) and PCAT (Parallel connection criss-cross attention), the CANER model demonstrated significant improvements in recognizing named entities related to kiwifruit diseases and pests, particularly those with strong domain characteristics.

In addition to advancements in NER, ontology has also found its way into the agricultural domain to facilitate the retrieval of relevant information from machine-readable articles. Abad-Navarro F and their colleagues (Abad-Navarro et al., 2020) introduced a comprehensive pipeline for generating scientific literature knowledge graphs. This pipeline leverages the combined power of the Semantic Web and natural language processing techniques to enable computers to understand and process data, empowering the development of user applications for literature searches. The pipeline encompasses three key steps: RDF generation, which involves translating scientific publication files into RDF format, encompassing both metadata and contents; Annotation, where the article's content is annotated with ontology concepts, generating new RDF triples that describe these annotations; and Knowledge graph population, where the generated RDF files, along with the ontologies used for annotation, are populated into a knowledge graph. By employing these techniques, researchers can unlock the potential of knowledge graphs to organize and retrieve agricultural information efficiently.

The construction of an agricultural knowledge map requires a strong and robust infrastructure that is able to collate, store, and disseminate information efficiently. It is important that the infrastructure is designed to meet the specific needs of the agricultural community and support a range of data types and formats, including scientific literature, research reports, databases, and multimedia materials (Qin and Yao, 2021). To solve the problem of data sovereignty and control, format inconsistencies and different interpretations, Klose et al. (2019) present Wikinormia, a collaborative platform in which interested participants can describe and discuss their own new data formats, which provides an efficient system for the cooperative, flexible digitalization of agricultural workflows. Liu et al. (2020) proposes a general way to construct industry knowledge graph in finance, agriculture, medical treatment, e-commerce and other fields. It is especially difficult for crops whose names vary depending on the agriculture activity, edible parts or cultivation method in which various names are used.

2.2.2. Integrating knowledge graphs with question answering systems

Knowledge Based Question Answering (KBQA) is a method of answering questions posed by users in natural language by leveraging structured knowledge graphs and ontologies. Instead of relying on keyword matching and retrieval, KBQA systems utilize semantic parsing techniques to extract the user's intended meaning and map it to specific entities and relationships in the knowledge graph. By utilizing the structure and relationships of the knowledge graph, KBQA systems are able to provide more accurate and comprehensive answers to questions. This approach is increasingly important as the amount of structured and unstructured data continues to grow exponentially, and traditional search methods become less effective at finding relevant information.

FoodKG (Gharibi et al., 2020) is a novel software tool that employs an existing graph embedding technique trained on a controlled vocabulary called AGROVOC, which is published by the Food and Agriculture Organization of the United Nations. This tool can improve decision-making and knowledge discovery as well as to provide improved search results for data scientists in the Food, Energy, and Water domains. Based on AGROVOC, Panoutsopoulos and Brewster (2022) extracts agricultural concept mentions from text through the deployment of custom trained Named Entity Recognition models and the exploitation of Graph Neural Networks to recommend concept and relation additions towards predicting future AGROVOC states.

In large-scale domain dynamic knowledge graphs, there are several issues including incomplete acquisition of original data, low accuracy with knowledge extraction and knowledge fusion, as well as unnonuniform semantic relations between entities. Chen and Xing (2022) constructs dynamic knowledge graph based on "seven-step method" and apply this scheme in the field of agricultural informatization. Yang et al. (2022) proposed a method based on GCN (Graph Convolutional Networks) using the attention mechanism to constrain the information propagation in the process of graph convolution to effectively utilize the higher-order information of the graph, and finally obtained the classification of each data node by SVM.

Knowledge graphs can help to model and represent the relationship between meteorological data, such as temperature, rainfall, and soil moisture, and agricultural production, such as crop yield and quality. Chenglin et al. (2018) proposed a method of constructing China meteorology and agriculture knowledge graph based on semi-structured data, which is successfully applied to the automatic generation of crop meteorological reports. Zou et al. (2023) proposes a precise recommendation method for suitable planting areas of maize varieties based on a knowledge graph. The meteorology knowledge graph of maize ecological regions is constructed at county-scale and a RippleNet recommendation model is used to mine the potential spatial correlation of maize variety suitability in different meteorological environments.

In order to realize the accurate prediction of fruit tree diseases and pests in the text description, Guan et al. (2021) constructs a knowledge graph in the agricultural field, and encodes the knowledge in the agricultural field through the knowledge representation model, combines the description text provided by the user to obtain the representation vector of the fruit tree diseases and pests feature entity, and then passes the representation vector and the pest image representation vector through CNN-DNN-BiLSTM network recognizes fruit tree diseases and pests. The method can fully integrate agricultural knowledge graph and deep learning technology, and play a positive role in improving the diagnosis of fruit tree diseases and pests.

Knowledge graph technology has been playing an increasingly important role in the optimization of e-commerce platforms, especially in agriculture product e-commerce platforms. Knowledge graph technology can help e-commerce platforms better understand customer needs, preferences, and behaviour patterns, allowing for the optimization of marketing strategies and the development of new product offerings, ultimately leading to improved sales and revenue. Xie et al. (2022) proposes an agricultural products recommendation algorithm based on the combination of attention factor decomposer and knowledge graph. This method models the higher-order connectivity of the produce knowledge graph in an end-to-end manner under the space of the knowledge graph, recursively propagates embeddings from the neighbours of the nodes, and extracts the potential feature vectors of the produce by using the attention factor decomposer as the message aggregation of the neighbouring nodes.

Smart farming is a rapidly growing field that utilizes cutting-edge technologies to revolutionize traditional farming methods. One such technology is the knowledge graph, which is used to capture, organize, and communicate complex data and information related to the farming process. The knowledge graph can be integrated with big data analytics and cloud computing to generate valuable insights into farming operations, such as crop yields, soil conditions, weather patterns, and pest control. Additionally, the integration of the knowledge graph with the Internet of Things (IoT) technology provides real-time monitoring and control of farming equipment and infrastructure, allowing farmers to make informed decisions and optimize their farming practices. The application of the knowledge graph in smart farming systems has the potential to enhance crop productivity, reduce operational costs, and promote sustainable agriculture. Bhuyan et al. (2021) proposed a knowledge graph represented as a lattice to capture and perform reasoning on spatio-temporal agricultural data. Choudhary et al. (2020) analyze the crop data collected from an agriculture site in Rajasthan, India, that includes both Rabi and Kharif cropping patterns, and then link the collected data and the smart farm ontology to populate a knowledge graph. The generated knowledge graph is used to provide structural information and aggregate data by using SPARQL queries and is further used by machine learning models to predict the crop yield to benefit farmers and various stakeholders.

2.3. *Large language models in agriculture*

Large language models (LLMs) are developing fast and have revolutionized question answering in recent years (Siche and Siche, 2023; Silva et al., 2023). On the basis of pretrained language models, researchers have found that scaling models' size often leads to large performance improvement in downstream tasks (Zhao et al., 2023).

One of the most prominent LLM series is GPT (Generative Pre-trained Transformer). GPT-3 (Brown et al., 2020) extends the model size to 175 billion and achieves strong performance on many tasks, especially in task-agnostic, few-shot scene. Its ability to understand the semantics and nuances of natural language makes it a highly effective tool for question answering tasks. GPT-4 (OpenAI, 2023) is a large-scale, multimodal model which can accept image and text inputs and produce text outputs. LLaMa (Touvron et al., 2023a, 2023b) is another notable LLM series ranging from 7B to 70B parameters, which has attracted much attention from the research community due to its openness.

Existing LLMs mainly choose a mixture of public textual datasets as the pre-training corpus. After pre-training, LLMs gains the general ability to solve various downstream tasks. In addition to it, instruction tuning is required to adapt LLMs to specific goals (Wang et al., 2022d). Depending on zero shot and instrument tuning capabilities, LLMs have been applied in various domains, such as medical and legal fields. Tang et al. (2023) propose a new training paradigm that involves generating a vast quantity of high-quality synthetic data with labels utilizing ChatGPT and fine-tuning a local model for the downstream task, which results in significant improvements in the performance of named entity recognition and relation extraction. ChatAgri (Zhao et al., 2023a) conducts a preliminary comparative study on ChatGPT, PLMs-based fine-tuning methods, and PLMs-based prompt-tuning methods. A series of empirical results demonstrate that ChatGPT is an ideal solution for agricultural text classification. Fig. 3 shows the difference between ChatAgri and existing prompt learning paradigm using an agricultural sentiment analysis (Jin et al., 2024; Jin et al., 2023b; Parillas et al., 2022; Yadav et al., 2022) example.

In conclusion, utilizing LLMs for the agricultural domain can revolutionize the way farmers, researchers, and policymakers access and utilize agricultural information. By harnessing the power of these models, we can greatly enhance productivity, sustainability, and decision-making processes in agriculture, contributing to a more efficient and resilient food system.

3. Application of knowledge storage architecture in agricultural practices

In recent years, the agricultural industry has witnessed significant advancements in information sensing, acquisition and analytical processing technologies, enabling farmers and stakeholders to access vast amounts of data and information relevant to their practices. One crucial aspect of leveraging this wealth of knowledge lies in the design and implementation of an efficient knowledge storage architecture. Specifically, knowledge storage architecture can facilitate disease diagnosis and treatment, offer production advice, and enable comprehensive information inquiry. Table 1 presents the scenario, device, main method, performance and limitation of some agricultural question answering systems.

3.1. *Disease diagnosis and treatment*

Agricultural disease diagnosis and treatment is a critical aspect of farming and agriculture. Farmers must identify and treat diseases that affect their crops or livestock to ensure that their produce or animals are healthy and safe for consumption. Advanced technologies, such as question answering systems, can be used to improve the accuracy and speed of disease diagnosis and treatment. Farmers can input specific symptoms or characteristics of the affected crops or animals into the system, which will return potential diseases that match the description. The system can also recommend treatment methods and suggest

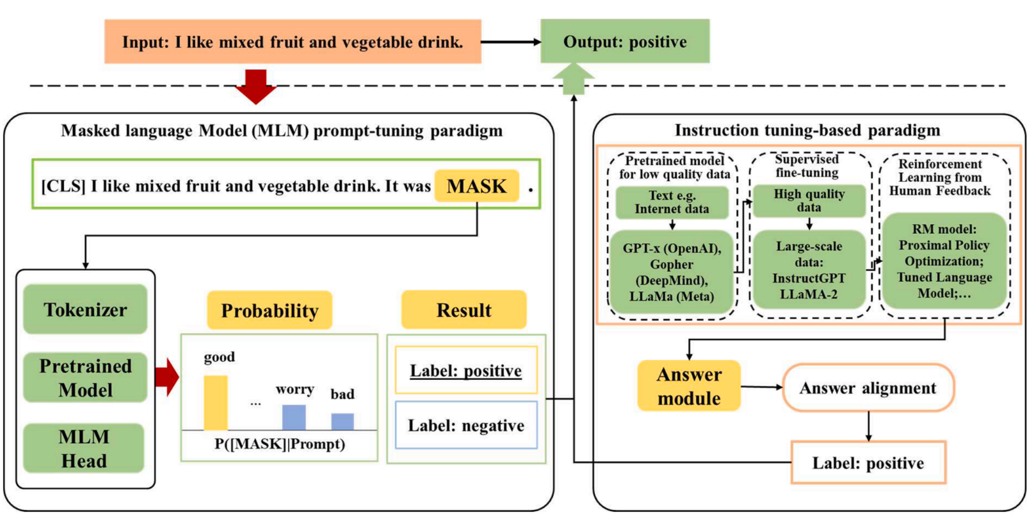

Fig. 3. The paradigm comparison of ChatAgri (Zhao et al., 2023a) and existing prompt learning paradigm using an agricultural sentiment analysis example.

Table 1
Summary of agricultural question answering systems.

Paper	Scenario	Device	Main Method	Performance	Limitation
(Sun et al., 2012)	Agriculture domain		Calculate the final similarity between the keywords using both similarity and correlation		Considering the correlation slows down the calculation speed
(Kawamura and Kawamura, 2014)	Plant Cultivation	Smartphone	Triplification of query sentences and graph pattern matching using SPARQL	Accuracy 66.7 %	Method only evaluated on specific questions
AGRI-QAS (Gaikwad et al., 2015)	Agriculture domain		QANUS(Ng and Kan, 2015) based on LUCENE	Accuracy 69 %	Cannot handle some question types such as asking a list
ADNAS (Devi and Dua, 2017)	Agriculture domain		SPARQL generated by triple formation using Stanford dependency tree	Recall 86.7 Precision 86.7	Ontology building part of the system is a time-consuming task
(Menaha et al., 2017)	Sports and agriculture		Remove stop words, search in google, and then use keyword matching algorithm		
(Sahni et al., 2018)	Cotton farming		Search over ontology graph aided by farmer's context as well as weather predictions made		New observations, specific local conditions and new experimentation requires ontology to be extended in time
(Kung et al., 2021)	Pig raising	CLI & GUI	bi-GRU, SNN, LSTM, SPARQL Query	Accuracy 87.79 %	Knowledge map cannot fully answer all the questions in the knowledge base
(Rose Mary et al., 2021)	Agriculture domain	Smartphone	RNN	Accuracy 97.83 %	
(Wang et al., 2022b)	Rice		LSTM	BLEU 35.3 %	requires a lot of supervision
(Lan et al., 2023)	Fruit tree disease		Unified visual question answering model combining the bilinear model and co-attentive learning	Accuracy 86.36 %	Inaccurate positioning of keywords and stochastic attention learning lead to incorrect prediction
(Huang et al., 2023)	Agriculture domain	Smartphone	BERT	F1 74.05 %	Cannot be directly transferred to a text classification task at other levels
(Rehman et al., 2023)	Agriculture domain		LSTM	F1 96.58 %	not capable of providing answers to queries that require real-time information
(Li et al., 2023)	Agricultural sensors		KBQA model consist of Bidirectional Entity Scoring Module, Question Embedding Module and Multi-hop Answer Selection Model	Accuracy 100 %	The neighbor entity from the topic entity can be reflected in the loss function rather than as a model input. It's necessary to increase the robustness of the model.
(Silva et al., 2023)	Agronomist certifications		Llama 2, GPT-3.5 and GPT-4 with Retrieval-Augmented Generation and Ensemble Refinement	Accuraccy 93 %	The models can be affected by potential limitations in their training data concerning.

preventative measures to avoid future outbreaks. By using question answering systems, farmers can quickly and effectively diagnose and treat agricultural diseases, leading to a healthier and more productive farm.

Suktarachan and Saint-Dizier (2009) emerged a rice disease question answering system from a need from the Thai Ministry of Agriculture. The main goal is to develop tools for e-Farming, in particular rice farming, so that farmers can easily get information on farming rice and rice diseases, for example via SMS. The Thai Ministry of Agriculture has large text databases on the way rice can be planted, on how to prepare and fertilize soils and on the numerous diseases rice may be subject to, the effects, the treatments, etc. Question answering is a particularly well-adapted approach to allow farmers to query in Thai (via SMS short messages) such databases. Yusof et al. (2018) develops an application that enables a user to recognize cocoa diseases afflict by the plant and provide user appropriate advice or treatments in shorter time period. The user will answer the questions based on cocoa plant condition or symptoms and the application generates the answer in form of disease and treatments. A rule-based and forward chaining inference engine has been used as part of the system development. With this application, it helps and allows the user to recognize cocoa diseases with useful treatments suggestion.

In recent years, computer vision technology has been increasingly applied to the field of agriculture for disease diagnosis in crop plants (Ouhami et al., 2021; Syed-Ab-Rahman et al., 2021). In addition to traditional text-based question answering methods, visual question answering (VQA) methods have been developed to incorporate image information into disease detection in agriculture. In Lan et al. (2023) study, a VQA model for fruit tree diseases based on multimodal feature fusion was designed. Fusing images and Q&A knowledge of disease management, the model obtains the decision-making answer by querying questions about fruit tree disease images to find relevant disease image regions. These VQA methods use deep learning techniques to process images of crop plants and detect the presence of diseases based on visual cues. This has proven to be a promising approach, especially for areas with limited human expertise to diagnose diseases in agriculture. Additionally, VQA methods provide a more intuitive and user-friendly way to diagnose diseases, enabling more farmers to easily access these diagnostic tools.

3.2. Production advice

With the help of natural language processing and machine learning techniques, the agricultural question answering system can use large databases of information to answer queries related to crop production and provide farmers with recommendations on how to optimize their yield. The system can provide advice on everything from soil quality and nutrient requirements to pest control and crop rotation. By integrating advanced data analytics and predictive modelling techniques, the system can also provide context-specific recommendations tailored to the unique needs of each farmer, taking into account factors such as weather patterns, local climate, and historical data. Ultimately, such a system could help farmers make more informed decisions and improve their overall production outcomes, leading to more sustainable and efficient agricultural practices.

Kung et al. (2021) designs and implements an intelligent knowledge question-and-answer system for pig farming. To identify intelligent questions and answers for raising livestock pigs, Siamese neural network (SNN) is adopted to obtain accurate answers in fetch the final answer. An SNN compares candidate answers to a livestock problem as two input values, previously trained by the bidirectional LSTM neural network. The final answer is determined from the cosine similarity value, which represents the highest relevance to the question. Finally, this study implements the intelligent pig-raising knowledge question answering system based on the proposed methodology. The evaluation results reveal that the proposed system is accurate and practicable.

Community question answering systems (CQAs) are emerging as a popular form of social media where people ask and answer questions to other people within a community. These platforms are designed to harness the collective knowledge and expertise of the community, enabling people to find answers to their queries quickly and accurately. In the question-and-answer (Q&A) communities of the "China Agricultural Technology Extension Information Platform", thousands of rice-related Chinese questions are newly added every day. The rapid detection of the same semantic question is the key to the success of a rice-related intelligent Q&A system. To allow the fast and automatic detection of the same semantic rice-related questions, Wang et al. (2021a) proposes a new method based on the Coattention-DenseGRU (Gated Recurrent Unit). According to the rice-related question characteristics, Word2vec was applied with the TF-IDF method to process and analyse the text data and compare it with the Word2vec, GloVe, and TF-IDF methods. Combined with the agricultural word segmentation dictionary, the applied Word2vec with the TF-IDF method effectively solved the problem of high dimension and sparse data in the rice-related text.

To improve the low accuracy of existing question answering models in providing real-time answers to users' questions in rice production, Wang et al. (2022b) propose a network architecture called Attention-ResLSTM-seq2seq for constructing a rice question-and-answer model. The model utilizes a 12-layer transformer-based GPT pre-training model to obtain text representations of rice question answering pairs. ResLSTM (Residual Long Short-Term Memory) is then employed to extract text features in both the encoder and decoder, with LSTM's output project matrix and output gate controlling spatial information flow. By reaching an optimal state, the network retains only the constant mapping value of the input vector, reducing network parameters and enhancing performance. Additionally, an attention mechanism is incorporated between the encoder and decoder to strengthen the weight of keyword feature information in the question. Experimental results demonstrate that the proposed Attention-ResLSTM-Seq2seq model achieves the highest scores in BLEU (35.3 %) and ROUGE (37.8 %) compared to six other generative question answering models related to rice.

Data collected by sensors in agricultural production can also be integrated into the question answering system. MAKBQA (Li et al., 2023) combines smart agriculture with information technology to enhance agricultural practices. The system collects crop growth and environmental data, such as soil moisture and temperature, for digital modelling. The sensor data the combines with existing agricultural knowledge from the internet to construct an agricultural knowledge base. The proposed question answering model achieves state-of-the-art performance on three publicly available datasets and demonstrates 100 % accuracy on the 3-hop test set of MetaQA.

3.3. Comprehensive information inquiry

In recent years, there has been a growing need for a comprehensive agricultural question answering system that can assist individuals and organizations in their quest for information on different farming practices and crops. The agricultural question answering system is a computer-based system that is designed to provide reliable and efficient responses to inquiries related to farming, regardless of the type of product cultivated. This system utilizes artificial intelligence and natural language processing algorithms to interpret questions asked in plain language and retrieve relevant information from large databases. With improved accessibility to this type of technology, farmers and individuals alike can access information from reliable sources thus facilitating informed decision making and ultimately contributing to increased efficiency and productivity in the agricultural sector.

There has been significant research in the field of question answering systems but there are few agriculture domain specific question answering system which returns actual answers by analysing unstructured data to the questions posed by the farmers. To address this limitation, Gaikwad et al. (2015) develops a system which gives answers to domain specific questions and evaluates them. The input for the system is a corpus of agriculture related documents (news articles, blogs etc. which are easily available on net) and a set of predefined question templates. For this purpose, we have worked upon an open source generic question answering framework, QANUS (Question Answering System by National University of Singapore) and modified it to make agriculture domain specific question answering system. QANUS adopts a pipelined approach to question answering, dividing question answering task into several sub-tasks including information base preparation, question processing, answer retrieval, and evaluation. In the Information Base Preparation (IBP) stage, an information source from which answers are to be derived can be set-up. The eventual information source is a LUCENE index of a corpus of documents given in XML format. Preprocessing of the documents that will make up the eventual information source is done here.

To effectively fulfil the comprehensive information inquiry needs of users, a question answering system requires a diverse range of knowledge management structures. Kawamura and Kawamura (2014) proposes a model of agricultural knowledge by Linked Open Data (LOD) with a view to establishing an open standard for agricultural data, allowing flexible schemas based on ontology alignment. Devi and Dua (2017) presents a question answering system on agriculture domain, the ADANS, to answer queries given in natural language. The system formulates an SPARQL from the queries given in natural language to query the resource description format data.

Through a comprehensive information inquiry question answering system, the farmers are able to get any interested agricultural information. Rose Mary et al. (2021) develops a chatbot to provide localized information such as current market prices for specific commodities in specific district and disease management. This method aids in the analysis of farmers' mindsets as well as the local structure of agricultural sector. While the technology provides a safe communication route for farmers, it also aids policymakers in comprehending their wants and concerns. Kisan Query Response System (KisanQRS) (Rehman et al., 2023) is also a question answering system designed for farmers' helpline centers, which utilizes deep learning techniques, integrating semantic and lexical similarities of farmers' queries and employing a rapid threshold-based clustering method The performance of KisanQRS is evaluated using a dataset of 34 million call logs from the Kisan Call Centre (KCC) operated by the Government of India. The query mapping module achieves a top F1-score of 96.58 % for a state, outperforming traditional techniques. The answer retrieval module achieves a competitive NDCG score of 96.20 % on 10,000 samples.

4. Challenges and future directions

4.1. Dataset collection and database construction

Agriculture dataset construction is a significant for question answer system of agriculture, but there are many limitations for the agriculture dataset construction. One of the major challenges in dataset collection for agricultural question answering systems is the availability of diverse and relevant data. In many cases, the data available may not fully represent the variety of crops, regions, and practices that exist in the agricultural industry. Additionally, the data may not be standardized or structured in a way that is suitable for use in a question answering system. To overcome these challenges, researchers need to work with agricultural experts and stakeholders to identify relevant data sources and to collect and process data in a way that is meaningful for the question answering system. This may involve the use of natural language processing techniques to extract structured data from unstructured sources such as text documents or social media.

Another challenge is the need for ongoing data updates and maintenance. Agricultural practices and technologies continuously evolve, and new questions may arise as a result. Additionally, the data used in question answering systems may become outdated or irrelevant over time. Therefore, it is important to establish a mechanism for ongoing data acquisition and maintenance. This may involve the establishment of partnerships with agricultural organizations and institutions to keep the dataset up-to-date, as well as the use of machine learning and other techniques to automatically identify and update data as needed. In addition, building a scalable and reliable data infrastructure is crucial for future research in the area of agriculture-based question answering system.

4.2. Technological challenges and breakthroughs of question answering system

4.2.1. Fit to different scenarios

One of the major challenges faced in developing question answering systems for agricultural scenarios is the highly variable nature of production environments. Even farms that serve a similar purpose can differ greatly in terms of topography, soil composition, climate, and other factors that can impact the crops grown and production processes used. This makes it difficult to generalize across farming scenarios and develop question answering systems that can accurately address the needs of all farmers. Without a deep understanding of each individual farmer's specific context, it is challenging to create effective question answering systems for the agricultural industry.

4.2.2. Visual question answering

Visual Question Answering, or VQA in short, is the task of answering natural language questions about visual content. It requires the combination of two different technologies: Computer Vision and Natural Language Processing. In agriculture, the use of VQA can have a significant impact on both the output and the quality of the products. VQA can help agricultural experts identify crop diseases, estimate crop yield, monitor fertility levels, and more. It can also aid farmers to make informed decisions about how to manage their farms with minimal environmental impact.

Despite the potential benefits of VQA in agriculture, there are still many challenges that need to be addressed. One of them is the lack of publicly available agricultural datasets. This limits the training of VQA models and benchmarking their performance. In addition, the images and questions that arise in the context of agriculture are more diverse and complex than those in other fields, requiring intelligent models that can handle the variability and context of agricultural questions. Therefore, developing effective VQA models in agriculture is still one of the future directions that needs to be actively pursued by researchers and industry experts.

4.2.3. Answer explainability

One of the significant challenges that agricultural question answering systems is the explainability of the system's answers. Explainability refers to the ability of the system to provide a clear and understandable explanation for how the system arrived at a particular answer. Although machine learning algorithms can provide accurate answers to agricultural questions, they lack a transparent explanation of their decision-making process. As a result, it becomes challenging for the end-users, agro-dealers, and farmers to trust the system's output, leading to low user adoption and acceptance. The use of visual and audio explanations can improve user acceptance and understanding. To improve the system's explainability, a combination of machine learning techniques, such as deep learning and NLP, can be integrated with human knowledge and expertise. This integration can help build a hybrid system capable of providing more accurate and transparent answers. Overall, enhancing explainability is crucial in agricultural question answering systems to increase user adoption and trust in the system's output.

4.2.4. Fairness and bias

Fairness and bias are some of the significant challenges in agricultural question answering systems. The issue of fairness arises when the system provides unequal responses or treatment to the users. In agriculture, the system should give equal responses to farmers regardless of their gender, race, ethnicity, or geographical location. Additionally, the system should ensure that the answers provided are fair and equitable to every user. Bias, on the other hand, is when the system provides different responses to different users based on their attributes. Therefore, in designing and developing agricultural question answering systems, fairness and bias need to be considered to provide equal and fair responses to all users.

4.2.5. Combined with intelligent chatbot

The development of intelligent chatbots, like ChatGPT (based on GPT-3.5 or GPT-4) and ERNIE Bot, has had a profound impact on agricultural question answering systems. The integration of prompt engineering and fine-tuning methodologies into chatbot development is pivotal for aligning chatbots with the agricultural domain. Harnessing the exceptional semantic understanding and multilingual capabilities of these chatbots holds promise for enhancing the versatility of agricultural question answering systems. The fusion of domain-specific knowledge and multilingual support, coupled with the potential integration of multimodal content, signifies a significant leap forward in agricultural question answering systems. Through these advancements, we can catalyse a more sustainable and efficient agricultural industry, empowering farmers with enhanced accessibility, real-time monitoring, and personalized recommendations.

4.3. Future applications in agriculture

4.3.1. Intelligent planting and breeding

The question answer system of agriculture has vast application prospects in the future development of smart agriculture. The question answering system optimizes efficiency and productivity by providing informed decision-making, timely issue resolution, and integration of real-time knowledge. It addresses specific production scenarios and environmental factors, reduces costs, and improves product quality. By leveraging sensor data and continuous knowledge updates, it ensures accuracy and self-updating for enhanced agricultural practices.

Efficient cultivation and breeding in agriculture are influenced by factors such as local climate, soil, and prevalent diseases. To make informed decisions, it is important to select varieties suitable for the region and environmental conditions and utilize environmental information in disease diagnosis, fertilization, and feeding processes. The question answering system can leverage multiple data sources, invoke existing relationship models, integrate local multidimensional information and specific production scenarios, and enhance product quality and yield.

In the growth process of crops and animal feeding, promptly addressing nutritional deficiencies, pest and disease problems, as well as ineffective fertilization, watering, and feeding, can reduce costs, improve productivity, and optimize product quality. The question answering system, with its natural language interface and optional voice interaction, can analyze the causes of events based on user-provided information, deliver preventive and control measures, and offer a fast and user-friendly experience.

Compared to traditional expert diagnosis, the question answering system can quickly utilize real-time updated knowledge sharing achievements. By standardizing and normalizing agricultural open data, different data sources and databases can achieve querying and exchange through alignment. Furthermore, the question answering system can integrate sensor data from smart agricultural production scenarios, provide feedback mechanisms to enhance system accuracy, optimize the richness and availability of knowledge, and achieve self-updating.

4.3.2. Intelligent logistics and transactions

Intelligent agricultural question answering system technology can provide a basis and reference for new agricultural product trading models in the future. With the rise of e-commerce as the mainstream method of shopping, intelligent logistics and transactions have rapidly developed, providing favourable conditions for the application of question answering systems in the transportation and sales of post-partum agricultural products.

As a popular form of online shopping, the increasing popularity of live streaming shopping services has created a platform for users to interact with anchors and ask questions about products, highlighting the importance of personalized recommendations and question answering services. However, on live streaming pages, there is a potential for consumer questions to be overlooked, resulting in consumers needing to independently search for answers or engage with other users. This process can be time-consuming and inefficient. To address this issue, intelligent question answering systems can provide consumers with quick and accurate service.

Question answering systems can play a crucial role in resolving the routine and repetitive issues faced by users throughout the logistics or delivery process. Additionally, by utilizing standardized marketing templates and filtering systems, question answering systems can contribute to precision marketing efforts and send targeted invitation notifications to potential customers. Moreover, artificial intelligence-powered question answering systems can automatically confirm delivery time and location with customers, thus saving valuable time for customer service personnel in appointment confirmation.

By integrating intelligent question answering systems into intelligent logistics and transactions, businesses can diminish their dependence on conventional communication methods. This allows for effective communication with customers while reducing unnecessary logistics activities, resource waste, and pollutant emissions. Additionally, it promotes efficient and personalized marketing, resulting in a reduction of production and sales of unnecessary product and service.

5. Conclusion

The increasing demand for effective information retrieval and support for decision-making in the field of sustainable agricultural development has led to the emergence of agricultural question answering systems. This study has focused on the knowledge storage architecture of question answer system and its impact on their overall effectiveness. By examining three prominent knowledge storage approaches, corpora, knowledge graphs, and large language models, we have provided insights into the advantages and limitations of each method. This analysis has allowed us to evaluate the effectiveness of these knowledge storage paradigms in practical applications within the agricultural production process.

The evaluation of agricultural question answering systems in recent years has provided valuable insights into their practicality. These systems have significantly improved the effectiveness of information retrieval and decision-making in various agricultural tasks, e.g., disease diagnosis, pests identification, and production prediction. However, it is important to note that the effectiveness of these systems heavily relies on the quality and coverage of the knowledge stored in the question answer systems. The knowledge storage approach in these systems, e.g., corpora, knowledge graphs, and large language models, plays a crucial role in determining the system's overall performance and utility.

The findings of this study emphasize the significance of knowledge storage architecture in agricultural question answering systems. The use of knowledge graphs and large language models have shown promising results in enhancing the performance and capability of these systems. However, further research is needed to address challenges such as knowledge representation, scalability, and system integration. Additionally, exploring hybrid approaches that combine the strengths of different knowledge storage paradigms may offer greater potential for advancing agricultural question answering systems.

CRediT authorship contribution statement

Tian Yang: Conceptualization, Writing – original draft, Visualization. **Yupeng Mei:** Writing – review & editing. **Ling Xu:** Writing – review & editing. **Huihui Yu:** Supervision. **Yingyi Chen:** Supervision.

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.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the National Natural Science Foundation of China "Intelligent identification method of underwater fish morphological characteristics based on binocular vision" (No. 62206021) and Beijing Digital Agriculture Innovation Consortium Project (BAIC10-2023).

References

Abad-Navarro, F., Bernabé-Diaz, J.A., García-Castro, A., Fernandez-Breis, J.T., 2020. Semantic publication of agricultural scientific literature using property graphs. Appl. Sci. 10 (3) https://doi.org/10.3390/APP10030861.

Bhuyan, B.P., Tomar, R., Gupta, M., Ramdane-Cherif, A., 2021. An ontological knowledge representation for smart agriculture. In: Proceedings of the IEEE International Conference on Big Data (Big Data), pp. 3400–3406. https://doi.org/10.1109/BIGDATA52589.2021.9672020.

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

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T.J., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. ArXiv, abs/2005.14165.

Calijorne Soares, M.A., Parreiras, F.S., 2020. A literature review on question answering techniques, paradigms and systems. J. King Saud Univ. Comput. Inf. Sci. 32 (6), 635–646. https://doi.org/10.1016/J.JKSUCI.2018.08.005.

Campbell, J.C., Hindle, A., Stroulia, E., 2001. Latent dirichlet allocation. J. Mach. Learn. Res. 139–159. https://doi.org/10.1016/B978-0-12-411519-4.00006-9.

Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D.M., Aletras, N., 2021. LexGLUE: a benchmark dataset for legal language understanding in English. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics, 1, pp. 4310–4330. https://doi.org/10.2139/SSRN.3936759.

Chen, Y., Xing, X., 2022. Constructing dynamic knowledge graph based on ontology modeling and Neo4j graph database. In: Proceedings of the 5th International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 522–525. https://doi.org/10.1109/ICAIBD55127.2022.9820199.

Chenglin, Q., Qing, S., Pengzhou, Z., Hui, Y., 2018. Cn-MAKG: China meteorology and agriculture knowledge graph construction based on semi-structured data. In: Proceedings of the IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), pp. 692–696. https://doi.org/10.1109/ICIS.2018.8466485.

Choudhary, N.K., Sree Laya Chukkapalli, S., Mittal, S., Gupta, M., Abdelsalam, M., Joshi, A., 2020. YieldPredict: a crop yield prediction framework for smart farms. In: Proceedings of the IEEE International Conference on Big Data (Big Data), pp. 2340–2349. https://doi.org/10.1109/BIGDATA50022.2020.9377832.

Deerwester, S., Dumais, S., Landauer, T., Furnas, G., Harshman, R., 1990. Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6.

Devi, M., Dua, M., 2017. ADANS: an agriculture domain question answering system using ontologies. In: Proceedings of the IEEE International Conference on Computing, Communication and Automation, ICCCA 2017, pp. 122–127. https://doi.org/10.1109/CCAA.2017.8229784, 2017-January.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2019, 1, pp. 4171–4186. https://arxiv.org/abs/1810.04805v2.

Gaikwad, S., Asodekar, R., Gadia, S., Attar, V.Z., 2015. AGRI-QAS question-answering system for agriculture domain. In: Proceedings of the International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015, pp. 1474–1478. https://doi.org/10.1109/ICACCI.2015.7275820.

Gharibi, M., Zachariah, A., Rao, P., 2020. FoodKG: a tool to enrich knowledge graphs using machine learning techniques. Front. Big Data 3. https://doi.org/10.3389/FDATA.2020.00012/PDF.

Gu, Y., Pahuja, V., Cheng, G., & Su, Y. (2022). Knowledge base question answering: a semantic parsing perspective. ArXiv.Org. 10.48550/ARXIV.2209.04994.

Guan, L., Zhang, J., Geng, C., 2021. Diagnosis of fruit tree diseases and pests based on agricultural knowledge graph. J. Phys. Conf. Ser. 1865 (4) https://doi.org/10.1088/1742-6596/1865/4/042052.

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. 13 (5), 2924. https://doi.org/10.3390/APP13052924, 2023, Vol. 13, Page 2924.

Izacard, G., Grave, E., 2021. Leveraging passage retrieval with generative models for open domain question answering. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics EACL 2021, pp. 874–880. https://doi.org/10.18653/V1/2021.EACL-MAIN.74.

Jin, W., Zhao, B., Yu, H., Tao, X., Yin, R., Liu, G., 2023a. Improving embedded knowledge graph multi-hop question answering by introducing relational chain reasoning. Data Min. Knowl. Discov. 37 (1), 255–288. https://doi.org/10.1007/S10618-022-00891-8/METRICS.

Jin, W., Zhao, B., Zhang, L., Liu, C., Yu, H., 2023b. Back to common sense: Oxford dictionary descriptive knowledge augmentation for aspect-based sentiment analysis. Inf. Process. Manag. 60 (3), 103260 https://doi.org/10.1016/J.IPM.2022.103260.

Jin, W., Zhao, B., Zhang, Y., Huang, J., Yu, H., 2024. WordTransABSA: enhancing Aspect-based Sentiment Analysis with masked language modeling for affective token prediction. Expert Syst. Appl. 238, 122289 https://doi.org/10.1016/J.ESWA.2023.122289.

Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., Yih, W.T., 2020. Dense passage retrieval for open-domain question answering. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, pp. 6769–6781. https://doi.org/10.18653/V1/2020.EMNLP-MAIN.550.

Kawamura, T., Kawamura, T., 2014. Question-answering for agricultural open data. In: Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8960, pp. 15–28. https://doi.org/10.1007/978-3-662-45947-8_2/COVER.

Klose, J., Schröder, M., Becker, S., Bernardi, A., & Ruckelshausen, A. (2019). Data preparation in agriculture through automated semantic annotation - basis for a wide range of smart services. ArXiv, abs/1911.06606.

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 (2), 906–922. https://doi.org/10.1002/agj2.20622.

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, 5287. https://doi.org/10.3389/FPLS.2022.1064399/BIBTEX.

Lan, Y., He, G., Jiang, J., Jiang, J., Zhao, W.X., Wen, J.R., 2021. A survey on complex knowledge base question answering: methods, challenges and solutions. In: Proceedings of the IJCAI International Joint Conference on Artificial Intelligence, pp. 4483–4491. https://doi.org/10.24963/IJCAI.2021/611.

Li, J., He, L., Wang, L., Wang, S., Ma, H., & Feng, K. (2023). MAKBQA: multi-hop knowledge base question answering system based on sensors and internet agricultural data. 33–40. 10.1109/SECON58729.2023.10287501.

Liang, J., Li, D., Lin, Y., Wu, S., Huang, Z., 2023. Named entity recognition of chinese crop diseases and pests based on RoBERTa-wwm with adversarial training. Agronomy 13 (3). https://doi.org/10.3390/AGRONOMY13030941.

Liu, C., Li, X., & Yu, Y. (2020). Research on construction technology of industry knowledge graph. 10.23977/CNCI2020079.

Liu, S., Zhang, X., Zhang, S., Wang, H., Zhang, W., 2019. Neural machine reading comprehension: methods and trends. Appl. Sci. 9 (18) https://doi.org/10.3390/app9183698.

Liu, Y., Wei, S., Huang, H., Lai, Q., Li, M., Guan, L., 2023. Naming entity recognition of citrus pests and diseases based on the BERT-BiLSTM-CRF model. Expert Syst. Appl. 234 https://doi.org/10.1016/J.ESWA.2023.121103.

Lun, Z., Hui, Z., 2022. Research on agricultural named entity recognition based on pre train BERT. Acad. J. Eng. Technol. Sci. 5 (4) https://doi.org/10.25236/AJETS.2022.050407.

Malik, N., Sharan, A., Hijam, D., 2015. Ontology development for agriculture domain. In: Proceedings of the 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 738–742.

Marinello, F., Zou, X., Liu, Z., Zhu, X., Zhang, W., Qian, Y., Li, Y., Karunathilake, E., Le, A.T., Heo, S., Chung, Y., Mansoor, S., 2023. The path to smart farming: innovations and opportunities in precision agriculture. Agriculture 13 (8). https://doi.org/10.3390/AGRICULTURE13081593.

Menaha, R., Surya, A.U., Nandhni, K., Ishwarya, M., 2017. Question answering system using web snippets. In: Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2017, pp. 387–390. https://doi.org/10.1109/I-SMAC.2017.8058377.

Mikolov, T., Chen, K., Corrado, G.S., Dean, J., 2013. Efficient estimation of word representations in vector space. In: Proceedings of the International Conference on Learning Representations.

Ng, J.P., & Kan, M.Y. (2015). *QANUS: an open-source question-answering platform*. https://arxiv.org/abs/1501.00311v1.

OpenAI. (2023). *GPT-4 technical report*. https://arxiv.org/abs/2303.08774v3.

Ouhami, M., Hafiane, A., Es-Saady, Y., El Hajji, M., Canals, R., 2021. Computer vision, IoT and data fusion for crop disease detection using machine learning: a survey and ongoing research. Remote Sens. 13 (13) https://doi.org/10.3390/RS13132486 (Basel).

Pal, A., Umapathi, L.K., Sankarasubbu, M., 2022. MedMCQA : a large-scale multi-subject multi-choice dataset for medical domain question answering. In: Proceedings of the ACM Conference on Health, Inference, and Learning. https://doi.org/10.48550/ARXIV.2203.14371.

Panoutsopoulos, H., Brewster, C., 2022. Data-driven update of AGROVOC using agricultural text corpora. In: Proceedings of the International Conference on Information and Communication Technologies for Sustainable Agri-Production and Environment.

Parillas, V.Q., Berina, J.M., Baesa, E.M., Raro, E.R., Palaoag, T., 2022. Sentiment analysis on hydroponic technology application for urban farming limitations. In: Proceedings of the International Conference on Emerging Technologies in Electronics, Computing and Communication (ICETECC). https://doi.org/10.1109/ICETECC56662.2022.10069517.

Pennington, J., Socher, R., Manning, C.D., 2014. GloVe: global vectors for word representation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, pp. 1532–1543. https://doi.org/10.3115/V1/D14-1162.

Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep contextualized word representations. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2018, 1, pp. 2227–2237. https://doi.org/10.18653/v1/n18-1202.

Qin, H., Yao, Y., 2021. Agriculture knowledge graph construction and application. J. Phys. Conf. Ser. (1), 1756. https://doi.org/10.1088/1742-6596/1756/1/012010.

Raj, S.D., Karthiban, M., Cse, M.E., 2022. Artificial intelligence in agriculture: a literature survey. Int. J. Creat. Res. Thoughts 10 (6), 2320–2882. www.ijcrt.org.

Ramos, J.E. (2003). *Using TF-IDF to determine word relevance in document queries*.

Rehman, M.Z.U., Raghuvanshi, D., Kumar, N., 2023. KisanQRS: a deep learning-based automated query-response system for agricultural decision-making. Comput. Electron. Agric. 213, 108180 https://doi.org/10.1016/J.COMPAG.2023.108180.

Robertson, S., Zaragoza, H., 2009. The probabilistic relevance framework: BM25 and beyond. Found. Trends Inf. Retr. 3 (4), 333–389. https://doi.org/10.1561/1500000019.

Rose Mary, C.A., Raji Sukumar, A., & Hemalatha, N. (2021). Text based smart answering system in agriculture using RNN. AgriRxiv, 2021. 10.31220/AGRIRXIV.2021.00071.

Sahni, S., Arora, N., Sen, S., Sarda, N.L., 2018. OntoAQ: ontology-based flexible querying system for farmers. Geospatial Infrastructure, Applications and Technologies: India Case Studies. Springer, pp. 201–215. https://doi.org/10.1007/978-981-13-2330-0_16/COVER.

Siche, R., Siche, N., 2023. El modelo de lenguaje basado en inteligencia artificial sensible - ChatGPT: análisis bibliométrico y posibles usos en la agricultura y pecuaria. Sci. Agropecu. 14 (1), 111–116. https://doi.org/10.17268/SCI.AGROPECU.2023.010.

Silva, B., Nunes, L., Estevão, R., Aski, V., & Chandra, R. (2023). *GPT-4 as an Agronomist Assistant? Answering Agriculture Exams Using Large Language Models*. https://arxiv.org/abs/2310.06225v2.

Suktarachan, M., Saint-Dizier, P., 2009. An application of lexical semantics annotation to question-answering in e-Farming. In: Proceedings of the 8th International Conference on Computational Semantics, IWCS, 2009, pp. 338–341. https://doi.org/10.3115/1693756.1693799.

Sun, B., Liang, Z., Zeng, Q., Zhao, H., & Ni, W. (2012). *Short text similarity computing method towards agriculture question and answering systems*. 10.2991/ICCIA.2012.61.

Syed-Ab-Rahman, S.F., Hesamian, M.H., Prasad, M., 2021. Citrus disease detection and classification using end-to-end anchor-based deep learning model. Appl. Intell. 52 (1), 927–938. https://doi.org/10.1007/S10489-021-02452-W.

Tang, R., Han, X., Jiang, X., & Hu, X. (2023). Does synthetic data generation of LLMs help clinical text mining? ArXiv. 10.48550/ARXIV.2303.04360.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: open and efficient foundation language models. ArXiv. 10.48550/ARXIV.2302.13971.

Touvron, H., Martin, L., Stone, K.R., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C.C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Scialom, T. (2023). Llama 2: open foundation and fine-tuned chat models. ArXiv. 10.48550/ARXIV.2307.09288.

Wang, C., Gao, J., Rao, H., Chen, A., He, J., Jiao, J., Zou, N., Gu, L., 2022a. Named entity recognition (NER) for Chinese agricultural diseases and pests based on discourse topic and attention mechanism. Evol. Intell. https://doi.org/10.1007/S12065-022-00727-W.

Wang, H., Wu, H., Zhu, H., Miao, Y., Wang, Q., Qiao, S., Zhao, H., Chen, C., Zhang, J., 2022b. A residual LSTM and Seq2Seq neural network based on GPT for Chinese rice-related question and answer system. Agriculture 12 (6), 813. https://doi.org/10.3390/AGRICULTURE12060813. Vol. 12, Page 813.

Wang, H., Zhu, H., Wu, H., Wang, X., Han, X., Xu, T., 2021a. A densely connected GRU neural network based on coattention mechanism for chinese rice-related question similarity matching. Agronomy 11 (7), 1307. https://doi.org/10.3390/AGRONOMY11071307. Vol. 11, Page 1307.

Wang, S., Xu, Y., Fang, Y., Liu, Y., Sun, S., Xu, R., Zhu, C., Zeng, M., 2022c. Training data is more valuable than you think: a simple and effective method by retrieving from training data. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics. https://doi.org/10.48550/ARXIV.2203.08773.

Wang, T., Xu, X., Wang, C., Li, Z., Li, D., 2021b. From smart farming towards unmanned farms: a new mode of agricultural production. Agriculture 11 (2), 1–26. https://doi.org/10.3390/AGRICULTURE11020145 (Switzerland).

Wang, Y., Kordi, Y., Mishra, S., Liu, A., Smith, N.A., Khashabi, D., & Hajishirzi, H. (2022). *Self-instruct: aligning language models with self-generated instructions*. 10.48550/ARXIV.2212.10560.

Weiqiang, J., Biao, Z., Yu, Z., Gege, S., Hang, Y., 2023. Fintech key-phrase: a new Chinese financial high-tech dataset accelerating expression-level information retrieval. ACM Trans. Asian Low-Resour. Lang. Inf. Process. https://doi.org/10.1145/3627989.

Xie, H., Yang, J., Huang, C., Wang, Z., Liu, Y., 2022. Recommendation algorithm for agricultural products based on attention factor decomposer and knowledge graph. In: Proceedings of the Asia Conference on Algorithms, Computing and Machine Learning (CACML), pp. 626–631. https://doi.org/10.1109/CACML55074.2022.00110.

Yadav, S., Kaushik, A., Sharma, M., Sharma, S., 2022. Disruptive technologies in smart farming: an expanded view with sentiment analysis. AgriEngineering 4 (2), 424–460. https://doi.org/10.3390/AGRIENGINEERING4020029.

Yang, T., Ran, F., Lu, M., & Yang, J. (2022). *Research on agricultural data mining model based on knowledge graph*. 27. 10.1117/12.2635381.

Yuan, Z., Mou, L., Wang, Q., Zhu, X.X., 2022. From easy to hard: learning language-guided curriculum for visual question answering on remote sensing data. IEEE Trans. Geosci. Remote Sens. 60 https://doi.org/10.1109/TGRS.2022.3173811.

Yusof, M.M., Rosli, N.F., Othman, M., Mohamed, R., Abdullah, M.H.A., 2018. M-DCocoa: M-agriculture expert system for diagnosing cocoa plant diseases. Adv. Intell. Syst. Comput. 700, 363–371. https://doi.org/10.1007/978-3-319-72550-5_35/COVER.

Zhang, L., Nie, X., Zhang, M., Gu, M., Geissen, V., Ritsema, C.J., Niu, D., Zhang, H., 2022. Lexicon and attention-based named entity recognition for kiwifruit diseases and pests: a Deep learning approach. Front. Plant Sci. 13 https://doi.org/10.3389/FPLS.2022.1053449/PDF.

Zhao, B., Jin, W., Ser, J., & Yang, G. (2023). ChatAgri: exploring potentials of ChatGPT on cross-linguistic agricultural text classification. ArXiv. 10.48550/ARXIV.2305.15024.

Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Wen, J.R. (2023). A survey of large language models. https://arxiv.org/abs/2303.18223v11.

Zhu, F., Lei, W., Wang, C., Zheng, J., Poria, S., & Chua, T.S. (2021). Retrieving and reading: a comprehensive survey on open-domain question answering. ArXiv.Org. http://arxiv.org/abs/2101.00774.

Zou, Y., Pan, S., Yang, F., Zhang, D., Han, Y., Zhao, X., Wang, K., Zhao, C., 2023. Precise recommendation method of suitable planting areas of maize varieties based on knowledge graph. Agriculture 13 (3), 526. https://doi.org/10.3390/AGRICULTURE13030526.