Computers and Electronics in Agriculture 235 (2025) 110396

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Review Article

The application progress and research trends of knowledge graphs and large language models in agriculture

Ruizi Gong [a], Xinxing Li [a,b,*]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, PR China
[b] Nanchang Institute of Technology, Nanchang 330044, PR China

ARTICLE INFO

Keywords:
Knowledge graphs
Large language models
Agricultural knowledge intelligent services

ABSTRACT

Enhancing agricultural productivity remains a crucial issue in agriculture, wherein agricultural knowledge intelligent services can improve the scientific and intelligent level of agricultural production through technologies such as knowledge coupling and inference decision-making. The application of knowledge graphs (KGs) in agriculture effectively supports the structured representation of agricultural data and helps manage agricultural data. On the other hand, the recent emergence of large language models (LLMs), with their strong language understanding and generation capabilities, provides new methods and ideas for knowledge services in agriculture. Though KGs and LLMs each have different strengths and limitations, their integration is believed to complement each other and has great potential to promote the development of agricultural knowledge intelligent services. In this paper, we review the current status of the application of KGs and LLMs in agriculture. We also discuss their complementary fusion as well as the prospect of their agricultural application, hoping to provide some references for the future development of agricultural knowledge intelligent services.

1. Introduction

In response to the growing global population and the increasing demand for food, the agricultural sector must continue to enhance productivity in a more efficient and intelligent manner (Ayoub Shaikh et al., 2022). Smart Agriculture, also known as Agriculture 4.0, refers to the optimization of agricultural inputs and outputs through various modern technologies, including sensors, image processing, machine learning, Internet of Things, artificial intelligence, and others (Alahe et al., 2024). These technologies help farmers enhance the efficiency of agricultural production (Gunaratnam et al., 2024). Furthermore, existing studies have demonstrated a growing inclination among farmers to adopt digital technologies to support decision-making processes related to agricultural production. Among these, the development of agricultural knowledge intelligent services has emerged as a prominent research focus in recent years (Zhao et al., 2022c). Agricultural knowledge intelligent services are capable of collecting, analyzing, and reasoning over extensive agricultural data by applying knowledge coupling, reasoning, decision-making, and other related technologies. This provides a scientifically sound and reliable theoretical foundation for agricultural practices (Zhao, 2023).

Knowledge graphs (KGs) offer a graphically representation of declarative knowledge, providing a structured representation of a large amount of knowledge in general or specific domains (Tamašauskaitė and Groth, 2023). They serve as valuable data sources for applications, such as intelligent question answering (Q&A), knowledge retrieval, and more. Currently, KG technology has demonstrated significant knowledge management benefits across a wide range of domains. In general domains, KG applications are nearing maturity, with notable examples such as Freebase (Bollacker et al., 2008), Yago (Suchanek et al., 2008), Wikidata (Vrandecic and Krtoetzsch, 2014). In specific domains, KG technology has made different progress and achievements in medical and other research areas. Similarly, in the field of agriculture, researchers have developed KGs to integrate agricultural expertise and data, offering substantial scientific and comprehensive support for the informatization and intelligence of the agricultural sector.

The advent of Large Language Models (LLMs) has initiated a paradigm shift in the field of natural language processing (NLP). LLMs have demonstrated remarkable capabilities in language comprehension and reasoning, enabling them to generate text that is not only satisfactory but also highly fluent. These models are now widely recognized and deployed globally as efficient tools for various language processing

* Corresponding author.
E-mail address: lxxcau@cau.edu.cn (X. Li).

https://doi.org/10.1016/j.compag.2025.110396
Received 17 December 2024; Received in revised form 10 March 2025; Accepted 8 April 2025
Available online 19 April 2025
0168-1699/© 2025 Elsevier B.V. All rights reserved, including those for text and data mining, AI training, and similar technologies.

tasks, with notable examples such as GPT4, the leading conversational multimodal LLM introduced by OpenAI. For the extension of agricultural knowledge services, LLMs can offer innovative approaches through their advanced text processing and generation capabilities, presenting significant potential for applications in the agricultural sector.

In contrast to other technologies currently employed in the agricultural sector, KG technology and LLM technology are more specifically oriented towards processing text-based data and delivering knowledge services within the agricultural domain. Furthermore, although KGs and LLMs offer distinct advantages in knowledge processing, each is also subject to certain limitations. For instance, KGs exhibit two prominent drawbacks: high construction costs and limited natural language understanding. Likewise, LLMs suffer from issues such as model hallucination and a lack of domain-specific knowledge. It has been argued that the characteristics of KGs and LLMs are mutually complementary (KG-LLM-Mcom), and their integration can address the respective shortcomings of each (Huang et al., 2024).

In this paper, we present a review of two techniques: KG and LLMs. The rationale behind this choice is based on the observation that both techniques are primarily focused on processing text-based data and delivering knowledge services. Moreover, a complementary potential between these two technologies has been identified. The objective of this paper is to explore the current state of applications and research trends of KGs and LLMs within the agricultural domain, both as individual technologies and in their integrated use. This exploration aims to provide insights for the future development and research of agricultural knowledge intelligent services.

To further clarify the exploratory focus of this review, the following outlines the strategies for the literature search. The primary sources of literature are drawn from the Web of Science database, supplemented by ArXiv platforms. This review focuses on the application of KGs and LLMs in the agricultural domain. In the section addressing the current state of KG and LLM applications in agriculture, the selection criterion prioritizes studies that explore the practical implementation of these technologies, as opposed to those exclusively focusing on a specific technology such as ontology construction and knowledge extraction technology. The selected literature highlights effective application related to KG or LLM, which can solve the problems in agriculture.

2. Agricultural knowledge graphs

2.1. Concepts and construction of knowledge graph

The concept of KGs originated from research on the Semantic Web (Shadbolt et al., 2006). After Google's initial implementation of this concept in 2012 to enhance its search engine performance (Singhal, 2012), the use of KGs gradually evolved and was later adopted across various domains and industries. Hogan et al. (2022) define a KG as "a graph of data intended to accumulate and convey knowledge of the real world, with nodes representing entities of interest and edges representing potentially diverse relationships between these entities." The representation captures semantic associations among the data, and further reasoning within the graph can be performed using specific techniques.

KG construction typically involves two primary methodologies: top-down and bottom-up (Hou and Niu, 2024), where "top" and "bottom" respectively represent the schema layer and data layer of KG. The data layer consists of different types of data, including structured, semi-structured, and unstructured data, which collectively form the factual foundation of the KG. Positioned above the data layer, the schema layer governs the expression paradigm and imposes constraints on these facts through restrictive conditions. Normally, the schema layer is constructed by domain experts, who define and validate entities, relationships and attributes to establish the ontology model that underpins the KG.

The top-down construction approach begins with the development of the ontology model upon which the KG will be based. Subsequently, the knowledge stored in the database is processed using this ontology model. In contrast, the bottom-up approach follows an opposite sequence: it first extracts entities and relationships directly from the data, then integrates those with high confidence into the knowledge base, and finally constructs the ontology model (Xu et al., 2016). A hybrid construction method combines both approaches (Zhou et al., 2022). In this method, the schema layer identifies preliminary entities for the data layer, while the data layer simultaneously updates the entities within the schema layer. This hybrid approach merges the speed of bottom-up construction with the accuracy of top-down construction, although it presents greater challenges in concurrently constructing the schema layer (Zhao et al., 2022a). Fig. 1 shows an overview of the KG construction process and the associated key technology.

2.2. Technologies for agricultural knowledge graphs

Although KG technology was proposed over a decade ago, its research and subsequent applications in the field of agriculture have been largely confined to the past five years. This phenomenon is closely linked to advancements in technology. Firstly, the progress of science and technology has led to a significant increase in the accumulation of agricultural research data. Furthermore, the evolution of KG technology is evident in the shift from manual to automated construction, which has enhanced the efficiency of agricultural KG construction and reduced associated costs, thereby accelerating the advancement of agricultural KG research.

The construction of agricultural KG includes key technologies such as ontology construction, knowledge extraction, knowledge fusion, and knowledge reasoning. A brief overview of these four critical technologies is provided below.

2.2.1. Ontology construction

An ontology is an explicit, formal specification of concepts and their relationships within a shared conceptualization (Gruber, 1993), (Wilson et al., 2023). A well-designed ontology provides a clear structure and a standardized semantic representation of the KG (Goldstein et al., 2021). Commonly used ontology construction tools include Protégé (Gennari et al., 2003), which offers a flexible and highly extensible development environment, and Ontolingua (Farquhar et al., 1997), which facilitates ontology maintenance and team collaboration.

Ontology construction methods can be categorized into three types: manual, semi-automatic, and automatic construction. The manual construction of agricultural ontologies is usually carried out by experts within the agricultural domain. While manual construction offers advantages such as authority, professionalism, and accuracy, it also associated with drawbacks, including high construction costs and slow construction speed.

The semi-automatic construction method, which builds ontologies based on existing agricultural thesauri and ontologies, is the most widely used approach for constructing agricultural ontologies and has achieved a considerable degree of maturity. However, mainstream methods are often hindered by issues such as missing evaluation sessions and cumbersome procedural steps. The Protégé-based construction method is currently the most mature and widely adopted semi-automatic approach (Li et al., 2023).

The automatic construction method represents an emerging approach for automatically extracting concepts from knowledge data in the agricultural domain using technologies such as machine learning (Wang et al., 2020) and deep learning (Al-Aswadi et al., 2020). This method is characterized by low construction costs and high construction speed. However, the quality of constructed ontology model still has room for improvement, which positions it as a promising direction for future research in agricultural ontology construction.

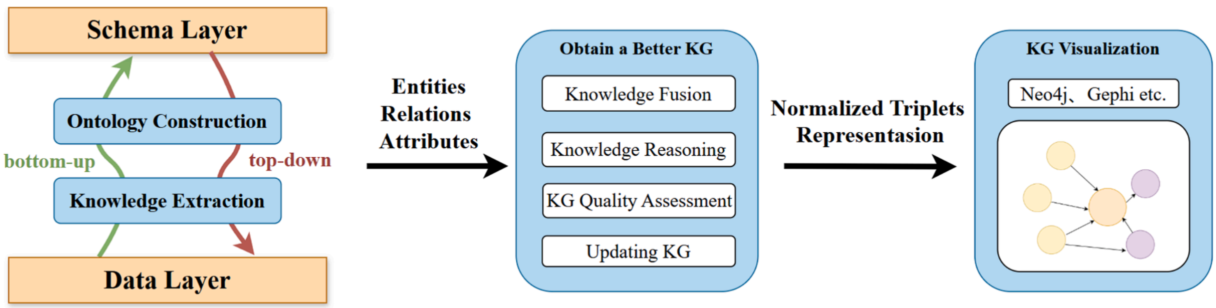

Fig. 1. Schematic of KG construction process and key technologies.

2.2.2. Knowledge extraction

Knowledge extraction is a critical component in the construction of KGs, as the effectiveness of this process directly impacts the overall quality of the KG. Agricultural data, which is inherently dispersed, voluminous, and multimodal, requires precise and efficient extraction techniques to ensure the accuracy and validity of the resulting knowledge representations.

The goal of knowledge extraction techniques is to derive the three fundamental elements of a knowledge unit, including entity, relation, and attribute from structured, semi-structured, or unstructured data. For structured data, there are already standardized languages like R2RML that can convert SPARQL to SQL and complete the knowledge extraction process (Rodríguez-Muro and Rezk, 2015). However, for unstructured data, the knowledge extraction process becomes more laborious. As technology advances, the knowledge extraction process carried out from the original rule-based (Chatterjee and Kaushik, 2017) to leveraging machine learning (Wang and Wang, 2014) and deep learning (Nismi Mol and Santosh Kumar, 2024). This gradual evolution helps reduce labor costs and improve the efficiency of extraction.

Entity extraction, also known as named entity recognition, involves the identification and extraction of various categories of entities within the unstructured agricultural knowledge texts. High-quality extraction of agricultural entities can provide more effective data support for subsequent tasks, such as information retrieval in downstream applications (Zhao et al., 2022b). Relation extraction, on the other hand, focuses on acquiring and analyzing the semantics of the relationships between entities (Dong et al., 2022). In addition to the solutions for separate extraction of entities and relations, methods for joint extraction of entities and relations have also been explored, often leveraging deep learning techniques (Chen et al., 2024a), (Qiao et al., 2022).

2.2.3. Knowledge fusion

Data sources within the data layer are often heterogeneous, and the inclusion of diverse data sources enriches the KG. To address challenges such as ontology heterogeneity and varying data quality across multiple sources, it is necessary to use knowledge fusion technology to integrate and unify the knowledge from different data sources, enabling the high-quality expansion of KGs derived from multiple data sources.

Knowledge fusion consists of two parts: ontology matching and entity alignment. Ontology matching mainly focuses on the fusion of the schema layer within a KG, aiming to identify the correspondences between ontologies across multiple sources of databases within the same agricultural sub-domain (Annane et al., 2018), thus enabling the enrichment and expansion of domain-specific ontologies. Entity alignment techniques, on the other hand, focus on the fusion of the data layer in KGs, aligning and merging objects that refer to the same real-world entity across different data sources (Zeng et al., 2022). These techniques ensure more accurate alignment of entities within the fused and extended KG, improving the precision of knowledge integration.

2.2.4. Knowledge reasoning

Knowledge reasoning refers to the process of inferring new knowledge from existing knowledge (Chen et al., 2020). It plays a crucial role in enriching and expanding the KG and is commonly applied to support downstream applications such as decision-making and recommendation systems. Knowledge reasoning techniques mainly include logic rule-based methods, representation learning-based approaches, and neural network-based methods.

Currently, representation learning-based and neural network-based methods are more widely employed due to their advanced technologies and superior modeling capabilities. The convolutional neural network model KG_CNN, integrates data from the cow disease KG into the convolutional neural network architecture, yielding better classification performance compared to traditional convolutional neural networks models (Gao et al., 2021). The GT-LSTM framework proposed by Cheng et al. (2024) implements the use of time series and spatial sequence information to predict winter wheat yields at the county level, where the inclusion of the graph neural network layer improves the accuracy of the prediction.

2.3. Current status of knowledge graph application in agriculture

KG is commonly utilized as a knowledge base in various application, providing robust support for addressing specific research problems. It is characterized by its structured knowledge, ease of operation, and high accuracy. Additionally, downstream applications, including information retrieval, decision support, intelligent Q&A, can be developed based on the KG, leveraging technologies such as knowledge reasoning and deep learning.

After reviewing recent research on agricultural KGs, we categorize the prominent application scenarios of KGs in agriculture into five main areas: agricultural research support, precision agriculture decision-making, disease control, agricultural policy recommendations, and food safety support. The following sections will provide an overview of these categories to illustrate the current KG applications in agriculture.

2.3.1. Agricultural research support

Information retrieval within the KG has the potential to advance research in agricultural sciences, particularly in areas as breeding, variety improvement, and other research activities that rely on identifying specific information within large-scale datasets.

For instance, KGs can assist researchers in uncovering potential associations within complex information. Zhang et al. (2024) developed a KG designated as PKGMG to study pig gut microbiota, from which they retrieved information and identified the localization of key microbes influencing feed efficiency. This provided robust support for the scientific investigation of the relationship between pig gut microbiota and feed efficiency. Another study created a gene-to-trait KG of plant photoreceptors to assess associations between plant traits and plant photoreceptors, which will help plant biologists select potential gene targets to improve beneficial traits in plants and support biological breeding (Mawkhiew et al., 2021).

KGs can also offer researchers rapid, accurate and visualized information retrieval services. A KG of milk yield traits and associated genes in dairy cows, developed by Hu et al. (2023), enhances the efficiency of literature searches and facilitates the rapid retrieval of information related to target genes or traits.

There are also instances where KGs are created to assist researchers in designing agricultural robots. TKG is a specialized KG that provide knowledge about agricultural cultivation, the electromechanical system, automatic control, etc. The researchers employed word co-occurrence network technology combined with expert annotation to develop an efficacy thesaurus. By obtaining design technology suggestions derived from the TKG, they ultimately designed a citrus picking robot, fully utilizing the information contained within the KG (Jin et al., 2021). In contrast to traditional design methods, which rely heavily on the designer's expertise or intuition, this approach offers a broader range of information and perspectives throughout the design process.

2.3.2. Precision agriculture decision-making

Precision agriculture has the potential to substantially enhance production efficiency and resource utilization by incorporating advanced information technologies and data analytics, thereby offering more precise decision support for agricultural production. In this context, KGs can effectively integrate and extract valuable insights from various domains, such as fertilizer application and production planning, to provide farmers with actionable recommendation and solutions. This will help optimize agricultural practices, boost crop yields and reduce resource waste.

Based on the constructed C3POKG, Darnala et al. (2023) developed two KG-based decision support systems for farmers: Elzeard and La Pépinière. In these systems, farmers input information about their farms, which then enables the systems to deliver relevant insights regarding crop planning and production processes, thereby assisting farmers in making decisions.

KGs can provide decision support for precision fertilization. Ge et al. (2024) proposed a rice fertilization recommendation system that utilized the PairRE model for knowledge extraction and developed a recommendation model based on the case-based reasoning technique. This system is capable of recommending a tailored fertilizer plan for farmers, aligned with the rice lifecycle. Another proposed method for accurate rice fertilization decision-making involves the use of an agricultural spatio-temporal multi-modal KG, which can be combined with crop growth monitoring data to identify potential nutrient deficiencies in the crop. Furthermore, the KG can then determine the most suitable agricultural fertilization model for the plot based on the current plot condition (Xu et al., 2023).

Considering external factors, such as meteorological disasters, is crucial for maintaining the efficient and sustainable characteristics of precision agriculture in the face of climate change and extreme weather events. (Wu et al., 2024) proposed a method for constructing an agrometeorological disaster KG to analyze these disasters. By utilizing graph retrieval technology, this method can automatically generate an analysis report on agrometeorological disasters, thereby effectively assisting agricultural workers in mitigating their impacts.

2.3.3. Disease control

Disease control plays a crucial role in mitigating issues such as yield loss and economic damage caused by diseases. Constructing a disease KG for a specific crop or livestock species can integrate dispersed information, thereby enhancing the efficiency of information acquisition for producers in addressing disease-related challenges.

One application of the disease KG is the retrieval of information and Q&A based on the knowledge it contains. This type of application offers precise and efficient knowledge support in the domain of specific crop diseases by constructing the KG as a highly specialized and targeted knowledge base. Wang et al. (2024a) developed a digital diagnostic system for tomato leaf pests and diseases based on a KG, enabling the retrieval and browsing of pest and disease-related knowledge, thereby offering theoretical support for the disease control. Huang et al. (2023), on the other hand, employed a rule-based approach in conjunction with Cypher query statements to achieve intelligent Q&A on tea pest knowledge. The resulting Q&A system, built upon the tea pest KG, is capable of accurately interpreting queries and delivering precise response. Yan et al. (2025) developed the CropDP-KG, a KG focused on crop and disease in China, encompassing detailed information on species and types of pests and diseases affecting crops in the region. Additionally, they created an integrated knowledge service system that facilitates knowledge retrieval and intelligent Q&A, offering users a comprehensive and extensive resource on pest and disease-related information.

Another type of application integrates KG techniques with image processing, sensor technologies, and other advanced methods to directly predict disease species. Zhang et al. (2020) constructed a KG that stores environmental information related to the occurrence of wheat stripe rust and proposed a prediction model based on a combination of the KG and bidirectional long short-term memory (Bi-LSTM) networks. This model leverage weather and environmental data to predict disease outbreaks and provides technical support for the management of wheat stripe rust. The proposed method was validated using a wheat stripe rust disease dataset. The prediction accuracy for wheat stripe rust was 93.21 %, representing a 4.5 percentage point improvement over the accuracy of the Bi-LSTM-based method alone. In the deep learning and KG based cassava disease prediction method proposed by Chhetri et al. (2023), the decision engine integrates the output of the visual model, which process images, with the output of the semantic model, which incorporates sensor observations, to determine the final prediction results. The semantic model then retrieves disease information from the ontology based on these results and generates natural language explanations, making it well-suited for real-world agricultural scenarios.

KGs also play a role in the field of animal disease control. While research in this area is relatively less extensive compared to crop pest control, they offer a unique advantage in integrating and applying veterinary knowledge. A study introduces an intelligent diagnostic approach for dairy cow diseases based on a KG that stores veterinary knowledge and medical records (Gao et al., 2021). This method enhances the efficiency and accuracy of clinical diagnosis for dairy cow diseases. Experimental results demonstrate the model's effectiveness in diagnosing dairy cow disease, with F1 scores of 85.87 % for the KG_CNN model based solely on the KG and 86.77 % for the KGTL_CNN model that integrates the KG with transfer learning. These scores represent improvement of 6.58 % and 7.7 %, respectively, over a conventional CNN model.

2.3.4. Agricultural policy recommendations

KG can facilitate the systematic integration of the complex interconnections between agricultural production, ecological resources, and policy elements, assisting agricultural decision-makers to uncover deep patterns that are challenging to capture through traditional analytical methods.

UrbanAgriKG focuses on extracting entities and relationships related to urban agriculture and capturing the complex interconnections within the urban agricultural system through knowledge-based reasoning techniques (Bhuyan et al., 2024), such as TransE, TransH, and TransR. These techniques can provide critical insights for various types of decision-making in the context of urban agriculture, including guiding agricultural investment decisions for urban policy makers, evaluating the status of food security and sustainability in urban areas, and analyzing the environmental impacts of urban agricultural production.

The KG integrates agricultural data from Bangladesh, referred to as BDAKG, can link other datasets for joint analysis (Nath et al., 2024). From this, researcher gain valuable insights into agricultural production, sustainability, and related factors in Bangladesh. For example, BDAKG suggests that the government of Bangladesh reduce rice production and increase the cultivation of wheat, soybeans to mitigate carbon emissions. Additionally, it recommends the enhancement of forestry development in the northern region of the country to maintain ecological balance.

These studies validate the decision-support capabilities of KG technology in optimizing agricultural production structures and balancing eco-efficiency, providing novel technical tools for agricultural policymakers.

2.3.5. Food safety support

Food safety continues to be a critical global issue that demands sustained attention (Stuart et al., 2024). KGs, as effective tools for information integration and analysis, demonstrate potential in applications such as case retrieval, intelligent Q&A, and food traceability within the realm of food safety.

Zhang et al. (2021b) constructed a KG of food safety cases and enabled efficient case retrieval using the BERT language representation model and calculation of case attribute similarity. Relevant food safety cases can be retrieved by searching keywords such as 'melamine' and 'poisoning'.

Integrating LLM technology can enhance traditional Q&A system, significantly lowering the barrier to using KGs for information acquisition. An innovative intelligent perceptual Q&A system for food KGs, based on LLMs, combines the expertise stored in the KG with the natural language generation capabilities of LLMs (Song et al., 2024). When a question posed matches relevant information in the KG, the identified structured information will be input into the LLM through a predefined prompt template, leading to the final output. If the information cannot be found in the KG, the LLM will directly generate the answer. This approach can provide users with fluent, professional, and accurate nutritional advice or food compliance counseling, among other services.

Tracking the quality of agricultural products is a crucial component in ensuring food safety. An effective traceability system can identify and address potential quality issues in time, thereby safeguarding food safety and ensuring regulatory compliance. Based on a KG that integrates knowledge related to the quality of agricultural products, Jing and Li (2024) developed an agricultural product quality traceability system. This system employs information mining through graph algorithms and deep graph traversal techniques to detect potential quality issues. The system was tested on carrot supply chain, and the detection accuracy for batch failure was 75 %.

Table 1 summarizes the applications of agricultural KGs discussed above. Currently, the integration of KG technology with knowledge reasoning, deep learning and other technologies has been progressively implemented in the agricultural sector, thereby enabling a range of agricultural knowledge services, such as information retrieval, decision-making reasoning, and intelligent Q&A, particularly in subfields such as precision fertilization and crop pest disease control.

Based on the directions of the improvement summarized in Table 1, future research trends in application of agricultural KGs are expected to focus on the following aspects. For data layer, future research will focus on expanding the coverage of trusted data and exploring method to effectively integrate multimodal data from images, sensor data, and other types of information to enhance the real-time applicability and practicality of decision-making. For technical improvement, the development of agricultural KGs will evolve towards greater intelligence and automation, driven by advancements in artificial intelligence

Table 1
Summary of agricultural KG applications.

Category	Reference	Key Technologies	Application Effect	Improvement Directions
Agricultural Research Support	(Zhang et al., 2024)	ontology mapping, quality assessment	contributing to the study of the relationship between pig gut microbes and feed efficiency	KG expansion and research directions enrichment
	(Mawkhiew et al., 2021)	graph visualization	assisting plant biologists in identifying potential gene targets to enhance agronomically beneficial traits in plants	KG expansion
	(Hu et al., 2023)	OpenIE, Neo4j	a rapid information retrieval platform for milk yield traits and associated genes in dairy cows	reducing labor costs and KG expansion
	(Jin et al., 2021)	TF-IDF, WCONet	successfully designed a citrus picking robot	multi-source KG fusion
Precision Agriculture Decision-making	(Darnala et al., 2023)	SAMOD	help farmers make decisions based on farm information provided by farmers	improving the accuracy of knowledge fusion
	(Ge et al., 2024)	PairRE, CBR	recommending targeted fertilization plans for farmers based on the life cycle of rice	KG expansion and knowledge reasoning enhancement
	(Xu et al., 2023)	SWRL rules	supporting farmers in making precision fertilization decisions	iterative optimization of spatio-temporal, multi-modal KGs
	(Wu et al., 2024)	graph retrieval technology	assisting agriculturists in defense against agrometeorological disasters	method area migration and actual meteorological data analysis
Disease Control	(Wang et al., 2024a)	ALBERT-BiLSTM-CRF	providing farmers with accurate pest and disease control advice	expanding the domain scope and optimizing the knowledge extraction model
	(Huang et al., 2023)	Neo4j, Cypher	realizing the effect of a basic Q&A on tea pests and diseases knowledge	enhance semantic comprehension and contextualization ability
	(Yan et al., 2025)	named entity recognition	constructed a pest and disease knowledge service platform with knowledge retrieval, overview and Q&A.	advancing the intelligent level of Q&A
	(Zhang et al., 2020)	Bi-LSTM	wheat stripe rust can be predicted based on the living environment of wheat	reduce the computing and storage burdens
	(Chhetri et al., 2023)	deep learning	cassava diseases can be predicted using image and sensor data	expanding domain knowledge and enhance model performance
	(Gao et al., 2021)	deep learning, transfer learning	proposed a method to diagnose diseases in dairy cows through KG-driven deep learning and transfer learning	model improvement with more complex architecture and better medical knowledge
Agricultural Policy Recommendations	(Bhuyan et al., 2024)	TransE/H/R	helping to accelerate the decision-making process in urban agriculture	enhancing data diversity
	(Nath et al., 2024)	online analytical processing	provide reliable recommendations for sustainable agricultural development in Bangladesh	expanding domain knowledge
Food Safety Support	(Zhang et al., 2021b)	BERT	handling food safety case search tasks	feature extension and sample expansion
	(Song et al., 2024)	LLM	constructed a food safety intelligent perceptual Q&A System	KG expansion
	(Jing and Li, 2024)	knowledge embedding, representation learning	constructed a KG-based agricultural product quality traceability system	accuracy and efficiency improvement

technology. For instance, the application of LLM technologies, as will be discussed later, have the potential to enable tasks such as knowledge extraction, thereby reducing the labor costs associated with constructing large-scale KGs. In addition, by introducing LLM for natural language understanding and interpretation, the system can lower the user threshold, and better provide agricultural knowledge intelligent services to agricultural workers with varying levels of literacy.

3. Agricultural large language models

3.1. Introduction to large language models

LLMs are artificial intelligence systems trained on billion-scale data volumes from books, articles and other internet-based content (Thirunavukarasu et al., 2023). The emergence of LLMs represents a significant advance in the field of NLP. LLMs are capable of processing enormous datasets, demonstrating robust generalization capabilities. This is achieved through the use of self-supervised learning, eliminating the need for labor-intensive manual labeling. Additionally, LLMs excel at managing long-term dependencies via the self-attention mechanism, resulting in a high degree of coherence and flexibility in text generation (Guo et al., 2024). As the quantity of training data and model sizes continue to grow, LLMs have become invaluable tools in advancing human–computer interaction, knowledge acquisition, content generation, offering vast application potential.

In 2017, transformer architecture, the inaugural fully attention-based sequence-to-sequence model was introduced (Vaswani et al., 2017), thereby laying a strong foundation for the rapid advancement of pre-trained language models. A year later, in 2018, GPT developed by OpenAI (Radford et al., 2018.), and BERT developed by Google (Devlin et al., 2019), made groundbreaking appearances one after another. These models demonstrated the immense potential of the pre-trained approach and catalyzed the widespread adoption of the pretraining-followed-by-fine-tuning paradigm for tackling NLP tasks. Subsequently, LLMs have undergone continuous updates and iterations, marked by a progressive increase in the number of parameters, leading to steady improvement in model performance. From the original GPT, with 330 million parameters, to today's multimodal LLM, GPT-4, capable of processing and generating multimodal information such as images, speech, and video, in addition to text. LLMs have rapidly evolved in just five years to become widely recognized and broadly utilized artificial intelligence systems. However, challenges related to data bias, privacy, and security, among others, remain. Further research and improvements are essential for the continued development of LLMs.

3.2. Key technologies for large language models

3.2.1. Transformer architecture

In the field of NLP, traditional recurrent neural networks and convolutional neural networks encounter several challenges when processing sequential data, primarily manifested in their insufficient ability to capture long-distance dependency and their low parallel processing efficiency. The introduction of the transformer architecture represents a significant breakthrough in NLP, addressing these limitations and enabling more efficient sequence modeling.

The transformer architecture was introduced by Google in the paper *Attention is All You Need* (Vaswani et al., 2017). As indicated by the title, the core innovation of the transformer architecture lies in its use of the self-attention mechanism to process both input and output sequences.

The attention mechanism assigns different weights to each element in the sequence, computes a weighted sum, and produces the output. The specific formula is shown below. In this formula, Q, K, V, and d_k respectively represent the query, key, value, and the dimension of the key. The self-attention mechanism, a specialized form of the attention mechanism, derives its queries, keys, and values from the same input sequence, enabling the architecture to effectively capture long-distance dependency.

$$Attention(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

The transformer architecture employs an encoder-decoder structure that enables the model to encode the input sequence and decode the output sequence, facilitating parallel computation. In contrast to traditional neural network architectures, where computations are performed sequentially, this design greatly improves the speed of model training.

3.2.2. Fine-tuning

Fine-tuning is a critical technique for adapting LLMs to specific downstream tasks. This process typically follows the unsupervised pre-training phase, during which the model is trained on a vast amounts of text data. Fine-tuning is achieved by utilizing manually annotated data to make targeted adjustments to the pre-trained model, enabling its output to meet the requirements and specific format of downstream tasks.

The full parameter fine-tuning (FFT) approach involves updating all the parameters of a pre-trained language model. This is achieved by optimizing each weight in the model using supervised data tailored to the corresponding downstream task. This comprehensive update typically leads to outstanding fine-tuning performance and is commonly used as a control group in fine-tuning technique research (Sun et al., 2023), (Razuvayevskaya et al., 2024).

Although FFT outperforms other fine-tuning methods in terms of effectiveness, the growing scale of pre-trained language models has highlighted a major drawback: excessive computational resource consumption. As a result, researchers have sought alternative solutions. Parameter-efficient fine-tuning (PEFT) has emerged as a prominent topic in fine-tuning research area. In this approach, only a small subset of the model's parameters is updated, while the majority remain unchanged, leading to a substantial reduction in computational resource requirements (Ding et al., 2022). Some PEFT methods have demonstrated the ability to achieve performance comparable to FFT methods in training LLMs. Ding et al. (2023) evaluated and compared the performance of FFT with four PEFT methods: prompt-tuning, prefix-tuning, LoRA, and adapter. The experiment highlights the potential of PEFT, with each approach exhibiting optimal performance in distinct downstream tasks.

3.3. Current status of large language model applications in agriculture

With the rapid development of LLM technology, its application research in the agricultural sector has also made initial strides. Currently, the focus of LLM applications in agriculture is primarily on the development of agricultural-specific LLM and the use of LLMs as tools for text-based information processing or generation. These applications include tasks such as data preprocessing, specific text generation, and information extraction. However, the adoption of LLMs in agriculture remains in its early stages. Many challenges, including issues like model hallucination and information security, continue to hinder the widespread implementation of agricultural LLMs in areas such as agricultural extension and related services (Tzachor et al., 2023).

3.3.1. Agricultural-specific large language model

Two primary approaches for constructing agricultural LLMs are retraining the base model with a large volume of agricultural data for pre-training, or fine-tuning the base model with a smaller amount of agricultural data, as illustrated in Fig. 2.

Retraining the base model with a large corpus of agricultural data enables the model to acquire extensive expertise in the agricultural subdomain, enhancing the specialization and domain adaptability of the LLM. Yang et al. (2024) developed PLLaMa, a LLM for plant science, using LLaMa-2 7B and LLaMa-2 13B as the foundational models. They

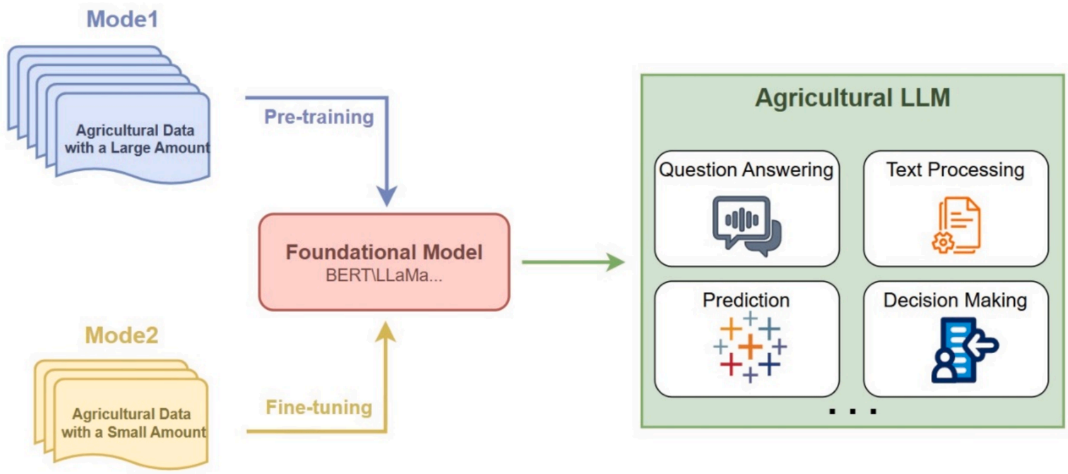

Fig. 2. Agricultural large language model construction in two modes.

employed 1.5 million academic articles in plant science as the extended pre-training dataset, which greatly improved PLLaMa's knowledge in the field of plant science. Following this, the instruction-based-fine-tuning process and expert feedback phase enabled PLLaMa to deliver accurate and professional responses to domain-specific queries. PLLaMa achieved 60 % accuracy on ten held-out plant science quizzes and received high recognition from global experts on zero-shot questions.

Unlike PLLaMa, the use of non-text-based data as pre-training data presents another future application for agricultural LLMs beyond text-based knowledge, with potential applications in tasks such as biological breeding and spatio-temporal data analysis. Mendoza-Revilla et al. (2024) developed AgroNT, a LLM focused on edible plant genomes, following the BERT modeling paradigm. The pre-training process utilized a dataset comprising 10.5 million genome sequences from 48 species. Experimental results demonstrated excellent performance in predicting regulatory features, RNA processing, and gene expression.

The drawbacks of this type of training method include high training overhead, computational resources and hardware costs. Therefore, in addition to using large amounts of agricultural sub-domain data to extend the pre-training dataset for constructing an agriculture-specific LLM as described above, another emerging research trend involves enhancing or evaluating the agricultural capabilities of existing LLMs through techniques such as fine-tuning and retrieval-augmented generation (RAG).

In order to evaluate the ability of LLM in answering agriculture-related questions, Silva et al. (2023) assessed several mainstream LLMs, including LLaMa2, GPT-3.5, and GPT-4, based on their ability to answer agricultural queries. The study also investigated the potential of RAG and ensemble refinement (ER) techniques to enhance the agricultural performance of LLMs. The experimental results indicate that GPT-4 achieves optimal performance, reaching 93 % accuracy in the agronomist certification test, surpassing the 88 % accuracy of earlier generalized models. Furthermore, RAG and ER are shown to enhance the performance of GPT-4 on region-specific agricultural problems.

Different models exhibit varying performance on distinct tasks, and selecting the most suitable model based on its performance is essential before deploying a LLM for application development. To develop an intelligent Q&A system for fruit and vegetable agrotechnical knowledge, Wang et al. (2023b) fine-tuned and optimized four LLMs, Baichuan2-13B-Chat, ChatGLM2-6B, Llama-2-13B-Chat, and ChatGPT, using LoRa, prompt-tuning, and RAG. Researchers evaluated these models in terms of knowledge entity recognition and Q&A abilities. The results demonstrate that ChatGLM outperforms the other three models, achieving an average accuracy of 89.2 % after fine-tuning and optimization. Furthermore, model hallucination was reduced by 43.5 % compared to pre-optimization performance. Consequently, ChatGLM was selected as the foundational pre-trained model for building the intelligent Q&A system, which is capable of generating high-quality answers to users' queries and providing sources of information.

The selection of fine-tuning methods also impacts model performance. Kpodo et al. (2024) developed AgRoBERTa, a specialized LLM for the agricultural extension domain, by constructing a Q&A dataset, AgXQA, and fine-tuning a RoBERTa model for agricultural extension. In this process, three types of fine-tuning methods were evaluated: simple fine-tuning, traditional fine-tuning, and PEFT. The experiments demonstrate that the pre-trained bottleneck adapter in PEFT achieves the best exact matching score of 55.15 % and F1 score of 78.89 % on the test set. By fine-tuning the model for agricultural extension domain-specific tasks, AgRoBERTa significantly improves performance on agriculture-related problems, achieving a 94.37 % agreement rate with expert assessments, surpassing GPT-3.5′s 92.62 % agreement rate.

3.3.2. Agricultural text processing

Leveraging advanced NLP processing capabilities, LLMs can effectively process complex texts and perform specific tasks, such as text categorization and knowledge extraction, through the use of Q&A, prompt engineering, and others.

Agricultural text categorization facilitates the automatic management and indexing of large amounts of agriculture-related unstructured data, thereby improving the efficiency and accuracy of information acquisition in the agricultural domain. The ChatAgri framework is designed to address the challenges in agricultural text categorization through ChatGPT (Zhao et al., 2023). ChatAgri enables ChatGPT to perform effective agricultural text categorization through three key phases: question construction, Q&A inference, and specification and alignment of answers. Several baseline methods (e.g., SVM, Random Forest, BERT-based fine-tuning) were established and compared, with experimental results demonstrating that ChatAgri achieved superior accuracy and F1scores in the agricultural text categorization task. Additionally, the experiments investigated ChatAgri's performance with the use of advanced prompt strategies, where the Chain-of-Thought prompting strategy significantly improve task performance across all datasets. The study demonstrates that ChatGPT excels in tasks requiring fine-grained semantic understanding, cross-linguistic comprehension, and generation of multiple randomized responses.

LLM-based Q&A system automate the extraction of knowledge from text to structured data. FINDER is a system that combines LLM and embedding-based retrieval techniques to extract entities, attributes, and descriptive terms from unstructured agricultural text and convert them into structured data (Peng et al., 2023). The system performs information extraction through a four-phase multi-round Q&A process. FINDER does not require the users to possess expertise in NLP, thereby lowering the barrier for information extraction. The evaluation results demonstrate that the system achieves an accuracy of 88 % and a recall of 84 % in the final phase. The main avenues for improvement and the challenges of this research include enhancing the accuracy of information extraction by adjusting the prompts and pipeline structure, as well as addressing the inconsistency in LLM performance with regard to attribute identification and value matching.

3.3.3. Agricultural text generation

The fluent text generation capabilities of the LLM can be integrated with techniques such as prompt engineering to produce outputs that are better aligned with the diverse needs of agricultural research. Properly applied prompt engineering can also reduce the output uncertainty of the LLM to some extent, thereby improving its reliability in the agricultural domain (Wang et al., 2024b).

For instance, one study integrates multimodal AI technology with LLM for the detection and prevention of potato diseases (Zhu et al., 2025). Users can upload image or text descriptions of potato leaves to the online potato disease platform. Following multimodal AI processing, the model delivers tailored disease prevention and control recommendations based on recognition results, using predefined templates for specific disease. In comparison to traditional disease detection applications, the incorporation of the LLM offers farmers supplementary knowledge services and enhances the overall intelligence and efficiency of the agricultural intelligent system.

LLMs can generate answers that meet the expected criteria through prompt engineering. AgriVLM is an agricultural intelligence analysis framework built upon a visual LLM, designed to identify crop disease species and crop growth stages (Yu and Lin, 2024). In this framework, ChatGLM serves as the LLM, with carefully designed prompts enabling the model to introduce more relevant information when presenting recognition results, thereby generating outputs that better align with human expectations.

We present a summary of the applications of LLMs in agriculture in Table 2. The table outlines the datasets used for model training or data processing, key technologies related to LLMs in the research, and the application effects. Fine-tuning and prompt engineering are frequently-used techniques that enable LLMs better align with the specific needs of the agricultural domain. In terms of practical applications, by leveraging the robust text processing and comprehension capabilities of LLMs, these models are increasingly employed as NLP tools in agricultural research, assisting researchers in accomplishing text processing tasks and reducing human costs.

Table 2
Summary of LLM applications in agriculture.

Category	Reference	Data Usage	Key Technologies	Application Effect
Agricultural-specific large language model	(Yang et al, 2024)	plant science-related article	pretraining, instruction-based fine-tuning	achieved 60 % accuracy on ten held-out plant science quizzes
	(Mendoza-Revilla et al., 2024)	edible plant reference genomes contained in the Ensembl Plants database	pretraining, PEFT using the IA3 technique	efficiently perform the tasks of predicting regulatory features, RNA processing, and gene expression
	(Silva et al., 2023)	agriculture exams and benchmark datasets from Brazil, India, and the USA	RAG, ER	achieved 93 % accuracy in the agronomist certification test
	(Wang et al., 2023b)	fruit and vegetable agrotechnical knowledge	LoRa, prompt fine-tuning, RAG	the average accuracy of the fine-tuned and optimized model reached 89.2 %, and hallucination was reduced by 43.5 %
	(Kpodo et al., 2024)	Q&A data on agricultural extension services	simple fine-tuning, traditional fine-tuning, PEFT	achieved a 94.37 % agreement rate with expert assessments
Agricultural text processing	(Zhao et al., 2023)	Amazon-Food-Comments, PestObserver-France, Natural-Hazards-Twitter, Natural-Hazards-Type, Agri-News-Chinese	prompt engineering	accomplish agricultural text categorization tasks efficiently and accurately
	(Peng et al., 2023)	unstructured pest data	embedding-based retrieval	agricultural information extraction was achieved with 88 % accuracy and 84 % recall
Agricultural text generation	(Zhu et al., 2025)	textual description of potato diseases	prompt engineering	generate prevention suggestions for specific diseases
	(Yu and Lin, 2024)	agricultural conversation data	prompt engineering	generate descriptions and explanations of disease types and crop growth stages

3.4. The prospects of large language models in agricultural applications

The application of LLMs in agriculture demonstrates significant flexibility and is anticipated to be applied in a broader range of scenarios in the future. Agricultural-specific LLMs have the potential to function as virtual agricultural experts, offering knowledge support for various subfields within agriculture. These models can also be integrated into intelligent systems, such as agricultural robots to facilitate Q&A interactions. Through an intelligent Q&A system, LLMs can provide targeted assistance to agricultural workers in areas such as crop management, policy guidance, and more. With powerful text processing and generation capabilities of LLMs, they also serve as effective tools for various agricultural research applications, particularly when combined with other artificial intelligence technologies. For instance, in the detection of crop pest diseases or the identification of livestock diseases, the LLMs can generate professional textual explanations and recommend countermeasures using techniques such as prompt engineering. This further enhances the overall intelligence of the system on the basis of identifying the types of diseases. The advent of LLMs offers the potential to refine artificial intelligence systems, providing farmers with more accurate and convenient agricultural services.

However, LLMs face numerous obstacles and challenges in their implementation of LLMs in agricultural settings due to issues such as model hallucination and output uncertainty. Kuska et al., (2024) suggest that the influence and trust in LLMs for agriculture can be strengthened by implementing monitoring systems to assess the quality of the output information, developing global guidelines, and introducing other supportive measures. In the future, challenges related to regional applicability, trust in data sources, and the verifiability of facts still need to be addressed within the agricultural LLM domain. Therefore, to enhance the implementation of LLMs in agricultural contexts, future efforts must focus not only on advancing technical capabilities but also on addressing issues related to data trust and policy support.

4. Mutual complementarity between knowledge graphs and large language models

4.1. Complementary characteristics of knowledge graph and large language model

KG and LLM are both technologies closely related to the field of NLP. However, they each possess distinctive advantages and limitations. While LLMs have garnered public attention for their extensive knowledge base, powerful text processing capabilities, and generation abilities, they are also subject to several inherent shortcomings stemming from their construction mechanisms. These limitations include non-interpretability (Danilevsky et al., 2020), output uncertainty, model hallucination (Liu et al., 2024), and a lack of domain-specific expertise. On the other hand, KGs offer advantages such as structured knowledge representation, specialization in specific domains, and interpretability (Zhang et al., 2021a). However, KGs also face challenges, including difficulties in large-scale construction and completion, as well as limited language comprehension abilities. The comparison between LLMs and KGs reveals that their respective characteristics exhibit complementary properties, as illustrated in Fig. 3. Consequently, the integrated application of these two technologies is expected to mitigate their individual shortcomings and overcome the limitations of relying on a single technology.

It is evident that integrating these two technologies presents a significant challenge, as discrepancies in data formats and model architectures contribute to the time and space complexity of the integration process (Wang et al., 2023a). Pan et al. (2024) proposed a roadmap for fusing LLMs and KGs, which includes two parallel frameworks for KG-enhanced LLMs and LLM-enhanced KGs, as well as a unified framework for synergizing KGs and LLMs. Representative studies within these three frameworks are introduced below.

4.1.1. Knowledge graph-enhanced large language models

The issue of LLM hallucination can be mitigated by leveraging the high accuracy of knowledge within the KG. Potential solutions include utilizing knowledge from the KG to construct the corpus for pre-training, as well as developing a knowledge-enhanced pre-trained language model.

TEKGEN is a model developed by Agarwal et al. (2020) to textualize KGs into a corpus called KELM. KELM can be directly used as natural language text in the pre-training process of augmented language models. This approach enables the language model to be indirectly trained with knowledge from the KG, thereby ensuring the accuracy of the knowledge source and alleviating the issue of hallucination.

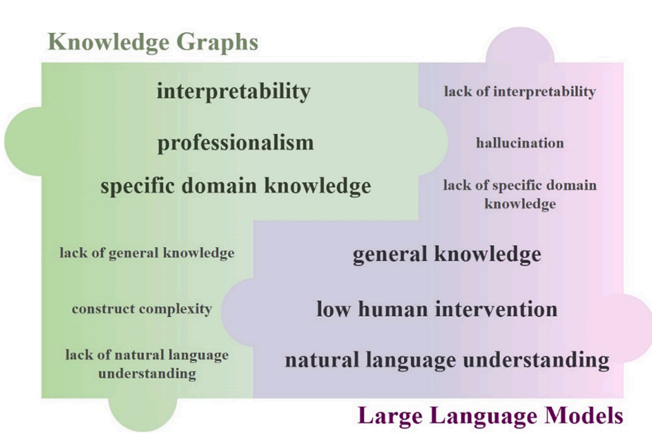

Fig. 3. Complementary characteristics of KGs and LLMs.

An alternative methodology involves the construction of knowledge-enhanced pre-trained language models. ERNIE is an augmented language model designed to accomplish innovative pre-training objectives, such as selecting appropriate entities from the KG for alignment. By introducing knowledge-driven tasks, ERNIE outperforms BERT in tasks like entity classification and relationship classification (Zhang et al., 2019). ERNIE 3.0 further expands the scope of language modeling. In contrast to earlier versions, ERNIE 3.0 is designed with a pre-training framework that integrates both an auto-regressive network and an auto-encoding network. It is trained on a corpus totaling 4 TB, comprising a plain text corpus and a large-scale KG. In this framework, the triplets in the KG are represented as a series of tokens directly connected to the input text (Sun et al., 2021). In another example, K-BERT, which is compatible with BERT parameters, transforms sentences into a knowledge-rich sentence tree by injecting structured knowledge triplets from the KG into sentences. Additionally, the soft position and visible matrix are adjusted to prevent the knowledge from deviating from the original meaning, thereby preserving the intended meaning of the sentence. (Liu et al., 2019).

In addition to enhancing the knowledge capacity of LLMs by integrating KGs into the LLM in various forms, KGs can also be leveraged to improve the retrieval and generation capabilities of LLMs. RAG as a widely adopted technique, combines pre-trained parametric and non-parametric memory to more effectively incorporate external knowledge. Upon receiving an input text, RAG retrieves the KG relevant to the text through Maximum Inner Product Search, thereby obtaining pertinent facts. These facts are then used to enhance text generation through sequence-to-sequence based LLM, leading to higher-quality text generation (Lewis et al., 2021). KGLM, as another example, can select and incorporate contextually relevant facts from the KG to generate text with descriptions that tightly aligned with the exact facts, outperforming generalized LLMs in factual text generation (Logan IV et al., 2019).

In response to the frequently-criticized lack of interpretability in LLMs, the graph-structural characteristics of knowledge within KGs offer a potential solution for interpreting the LLM outputs. XplainLLM introduces a framework for elucidating the inference behavior of LLMs. By utilizing this framework without requiring additional training, it enhances the interpretability of LLMs, mitigates hallucinations, and improves model performance (Chen et al., 2024b). KagNet, a novel knowledge-aware graph network module, is designed to model subgraphs, providing a reliable foundation for the inference process within LLMs (Lin et al., 2019).

4.1.2. Large language model-enhanced knowledge graphs

LLMs are instrumental in accomplishing tasks related to knowledge extraction and knowledge completion in the construction of KGs. The ChatIE framework leverages the capabilities of ChatGPT to break down the zero-shot information extraction task into multiple rounds of simple quizzes, thereby achieving strong knowledge extraction results (Wei et al., 2024). Similarly, ChatExtract also utilizes the functionality of ChatGPT, but it differentiates itself by focusing on design of follow-up prompts that enable high-quality knowledge extraction using LLMs (Polak and Morgan, 2024). Knowledge completion in KG construction involves filling in missing components within structured knowledge triplets. The KICGP framework, which integrates LLMs with a KGC retriever, employs an in-context learning strategy to guide the LLM and address the long-tail problem in knowledge-completion tasks (Wei et al., 2023).

The reasoning process in KGs is typically conducted within the graph itself, but the advent of LLMs has expanded the capabilities of traditional reasoning methods. Huang (2024) evaluates the GPT-4's ability to perform conversationally reasoning over KGs and introduces LLM-ARK, a framework that employs a full-textual environment prompt to gather state information for each reasoning step. This approach enables LLMs to engage in continuous learning throughout the reasoning task, demonstrating exceptional performance in reasoning tasks.

The remarkable effectiveness of LLMs in Q&A has been widely recognized. As a result, there is a growing interest in leveraging LLMs to enhance the performance of Q&A applications based on KGs. QA-GNN encodes the Q&A context using pre-trained language model and integrates it with a KG to form a joint graph. The model then updates their representations interactively using graph neural networks, demonstrating significant improvements in Q&A tasks (Yasunaga et al., 2022). ChatKBQA focuses on generating logical forms by utilizing a fine-tuned LLM and unsupervised retrieval techniques to enhance retrieval efficiency and accuracy. This approach shows great promise for tackling explanatory and knowledge-based Q&A tasks (Luo et al., 2024).

4.1.3. Synergize knowledge graphs and large language models

Distinguishing from the aforementioned two types of frameworks, research on collaborative KGs and LLMs primarily focuses on the development of synergized frameworks that aim to integrate their respective advantages. As a unified model for knowledge embedding and pre-trained language representation, KEPLER encodes entity descriptions as entity embeddings and jointly trains knowledge embeddings and masked language modeling goals on the same pre-trained language representation model. It demonstrates efficacy in both NPL and knowledge-embedding tasks (Wang et al., 2021). JointLK integrates task textual information and an external KG through a joint reasoning module based on a bidirectional attention mechanism. Irrelevant graph nodes are removed through a dynamic pruning module, which has been shown to effectively leverage KGs and language models for joint reasoning, despite its continued reliance on the LLM for output (Sun et al., 2022).

4.2. Current applications using the mutual complementarity between knowledge graphs and large language models in agriculture

Agricultural KG, which generally functions as a specialized database for agricultural sub-domains, offers a wealth of structured information and plays a critical role in capturing the complex relationships and knowledge involved in agricultural production. However, the dispersed and modality-diverse nature of agricultural information makes the construction of such KGs costly. In contrast, LLMs possess advanced NLP and generation capabilities, allowing for flexible user interaction and serving as valuable tools for various types of agricultural research. While LLMs are capable of addressing a broader range of problems, the issue of modeling illusion remains a challenge in practical agricultural applications. Therefore, future research should leverage mutual complementarity between KG and LLM (KG-LLM-Mcom) to mitigate the limitations of both KG and LLM in agricultural applications. In terms of technology, KG-LLM-Mcom has made some progress in research. However, applications of this concept in agriculture remain relatively limited. The existing studies primarily focus on utilizing LLMs to enhance the efficiency of agricultural KG construction and the joint use of LLMs and KGs for collaborative decision-making and reasoning.

Language models can improve the efficiency and quality of knowledge extraction during the construction of KGs. Wang and Zhao (2024) utilized the advantages of LLMs in language understanding to accomplish the tasks of entity recognition and relation extraction in the construction of AGKG, an agricultural engineering technology KG. Initially, key entities are accurately identified from agricultural literature using the BERT word embedding model and BIO coding scheme, combining BiGRU neural network and a CRF layer. Subsequently, complex relationships are effectively extracted using generative language models that leverage BERT features, entity labels, and are input into the BiGRU network. The loss function is optimized using the Adam optimization algorithm, which enhances both the automation and accuracy of KG construction.

To enhance both the accuracy and efficiency of the elaeagnus angustifolia diseases detection system's decision support functions, Zhao et al. (2024) developed a module that integrates both a KG encoder and a LLM encoder within the system. The KG encoder processes structured data related to elaeagnus angustifolia diseases, such as disease types and preventive measures, while the LLM encoder handles natural language data to extract textual information related to elaeagnus angustifolia diseases. Additionally, the dynamic attention mechanism, applied atop the two encoders, enables the system to dynamically adjust the flow of information between the KG and LLM. This mechanism allows for increased emphasis on symptoms and preventive measures when generating diagnostic reports. The final joint inference layer of the system consolidates information from both the KG and LLM to further optimize the decision output.

Dong et al. (2024) proposed a coupled optimal scheduling model for power supply, irrigation, and storage in agricultural parks, which fully leverages the advantages of KGs and LLMs. First, the agricultural water use KG was constructed through LLM. The LLM further integrated the structured knowledge from the KG with its reasoning capabilities, allowing it to combine real-time meteorological data and crop growth stages to dynamically adjust water demand predictions. The LLM also improved the analysis and prediction accuracy of photovoltaic power time series data by being fine-tuned with weighting parameters. The KG provided rich semantic associations to support the LLM in performing more accurate reasoning and scenario generation using meteorological APIs and soil moisture sensor data. These efforts offered data support for the operation and scheduling of the two-layer optimization model, optimizing energy consumption and system operation costs in the agricultural park.

Utilizing multimodal KGs and multimodal LLMs can provide higher quality and smarter services in the agriculture sector. However, the construction of multimodal KGs or LLMs is more complex, and the KG-LLM-Mcom technology may facilitate the processing of multimodal information. Lv et al. (2024) innovatively proposed a construction method for multimodal KGs for vegetables, leveraging the capabilities of pre-trained language models. The study successfully matched text entities and images related to vegetables within self-constructed datasets for corn and cabbage by preforming fine-grained comparisons between image regions and text entity vocabularies, alongside fine-tuning the pre-trained model. This approach achieved a recall of 76.7 %. This approach offers a innovative methodology for the development of multimodal KGs.

Table 3 provides a summary of current research on the application of KG-LLM-Mcom in agriculture, outlining the respective roles of KG and LLM, their joint application methods, and the resulting effects. The application of KG-LLM-Mcom in agriculture is still limited, with studies focusing on leveraging LLMs to enhance the efficiency of agricultural KG construction, as well as the combined use of LLMs and KGs for decision-making and reasoning.

4.3. The research gaps and way-forward

There is a limited number of studies applying the concept of KG-LLM-Mcom in the agricultural domain, and research gaps remain in both the technical and application aspects. In terms of technique, further in-depth research is required to address the issues of hallucination and non-interpretability in LLMs by leveraging the structured knowledge and domain-specific expertise in KGs. Additionally, improving the knowledge reasoning and intelligent Q&A capabilities of agricultural KGs through the natural language understanding abilities of LLMs presents the way-forward. These issues need to be explored and addressed within the agricultural context by designing appropriate framework that effectively integrates KGs and LLMs, thereby providing professional, accurate, and interpretable knowledge services with humanized features.

In the future, with the continued development of technology, the joint application of the KG-LLM-Mcom framework will be applied to broader agricultural scenarios. More complex and multimodal data in

Table 3
Summary of agricultural KG-LLM-Mcom applications.

Reference	Role of KG	Role of LLM	Method of Joint Application	Application Effect
(Wang and Zhao, 2024)	structured agricultural knowledge storage	entity identification and relation extraction	using the BERT word embedding model and the BIO coding scheme, combining BiGRU neural network and CRF layer	extract knowledge from agricultural literature efficiently and accurately
(Zhao et al., 2024)	processing structured data as one of the encoders	processing natural language information as one of the encoders	introducing dynamic attention mechanisms and joint reasoning layers	decision support for elaeagnus angustifolia diseases control
(Dong et al., 2024)	provide structured knowledge of agricultural water needs in different scenarios	short-term photovoltaic power prediction, structured data extraction assistance, meteorological and soil data analysis	LLM receives and reasons about the environmental and crop information, KG determines the corresponding water use scenarios for agriculture and irrigation.	generate scenarios of power supply-irrigation production simulation processes in agricultural parks to support subsequent irrigation strategies decision-making
(Lv et al., 2024)	provide the architecture of the knowledge graph	matching vegetable images with text entities	LLM aid efficient generation of a multimodal KG	matching of vegetable images to corresponding entities achieves 76.7 % recall

agriculture will be effectively utilized. For instance, the KG integrates information such as agricultural policies, food production laws and regulations, while the LLM can serve as a humanized policy Q&A assistant. By integrating with the development of intelligent interactive Q&A platform, knowledge services can be embedded in agricultural smart devices or smartphone terminals, facilitating the broader dissemination of agricultural knowledge.

In addition, by integrating other artificial intelligence technologies, such as Internet of Things and sensors, the idea of KG-LLM-Mcom will enhance the automatic analysis of collected field data and environmental monitoring information. This will provide substantial support for precision agriculture, disease control, and other related areas, thereby advancing the completeness of the smart agricultural system.

Overall, by combining the strengths of KGs and LLMs, this approach can elevate the comprehensive level of agricultural services. It is anticipated that, in the future, research applying the KG-LLM-Mcom concept to agriculture will increase, driving further development of agricultural knowledge intelligent services.

5. Conclusion

KGs and LLMs play a crucial role in advancing agricultural knowledge intelligent services, offering distinct advantages in the processing of agricultural knowledge. These technologies provide theoretical support and guidance to agricultural producers, enabling more informed and efficient decision-making within the agricultural sector.

This review begins with an introduction to KG and LLM technologies, followed by an overview of the current status of their applications in the agricultural domain. KGs are adept at processing and storing textual knowledge in a structured manner within the agricultural domain, providing reliable and precise agricultural knowledge support to producers. Future research should focus on advancing dataset expansion, processing cross-lingual and cross-domain data, and aligning multimodal data to enhance the effectiveness of agricultural knowledge in more complex agricultural scenarios.

LLMs are NLP tools that have evolved at an exceptional pace in recent years. The pre-training and fine-tuning paradigm endow LLMs with the capacity to adapt to various downstream tasks while processing large-scale datasets. Currently, research on agricultural LLMs is focused on developing specialized LLMs and evaluating LLMs' agricultural capabilities. Future research will primarily aim to enhance the credibility and interpretability of these models, as well as to explore their practical applications in agricultural contexts.

The review also examines the current status of KG-LLM-Mcom technology and its applications in the agricultural domain. The idea of KG-LLM-Mcom aims to integrate the advantages of KG and LLM, thereby mitigating their respective limitations. While the use of KG-LLM-Mcom technology in agriculture is still in its early stages, its potential and inherent value should not be overlooked. This review aims to introduce the concept of KG-LLM-Mcom and provide new perspectives and references for the development of agricultural knowledge intelligent services, ultimately enhancing the dissemination of precise and professional agricultural knowledge to agricultural workers.

CRediT authorship contribution statement

Ruizi Gong: Writing – review & editing, Writing – original draft, Visualization, Investigation. **Xinxing Li:** Supervision, Resources, Project administration, Funding acquisition.

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.

Acknowledgements

The authors would like to acknowledge the financial support by The 2115 Talent Development Program of China Agricultural University.

Data availability

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

References

Agarwal, O., Ge, H., Shakeri, S. and Al-Rfou, R., 2020. Knowledge graph based synthetic corpus generation for knowledge-enhanced language model pre-training. arXiv preprint arXiv:2010.12688.
Alahe, M.A., Wei, L., Chang, Y., Gummi, S.R., Kemeshi, J., Yang, X., Won, K., Sher, M., 2024. Cyber security in smart agriculture: Threat types, current status, and future trends. Computers and Electronics in Agriculture 226, 109401. https://doi.org/10.1016/j.compag.2024.109401.
Al-Aswadi, F.N., Chan, H.Y., Gan, K.H., 2020. Automatic ontology construction from text: a review from shallow to deep learning trend. Artif Intell Rev 53, 3901–3928. https://doi.org/10.1007/s10462-019-09782-9.
Annane, A., Bellahsene, Z., Azouaou, F., Jonquet, C., 2018. Building an effective and efficient background knowledge resource to enhance ontology matching. J. Web Semant. 51, 51–68. https://doi.org/10.1016/j.websem.2018.04.001.
Ayoub Shaikh, T., Rasool, T., Rasheed Lone, F., 2022. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. Computers and Electronics in Agriculture 198, 107119. https://doi.org/10.1016/j.compag.2022.107119.
Bhuyan, B.P., Tomar, R., Singh, T.P., Ramdane-Cherif, A., 2024. UrbanAgriKG: A knowledge graph on urban agriculture and its embeddings. Open Agriculture 9, 20220265. https://doi.org/10.1515/opag-2022-0265.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J., 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, SIGMOD '08. Association for Computing Machinery, New York, NY, USA, pp. 1247–1250. https://doi.org/10.1145/1376616.1376746.

Chatterjee, N., Kaushik, N., 2017. RENT: Regular Expression and NLP-Based Term Extraction Scheme for Agricultural Domain, in: Satapathy, S.C., Bhateja, V., Joshi, A. (Eds.), PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DATA ENGINEERING AND COMMUNICATION TECHNOLOGY, ICDECT 2016, VOL 1, Advances in Intelligent Systems and Computing. Presented at the 1st International Conference on Data Engineering and Communication Technology (ICDECT), Springer-Verlag Singapore Pte Ltd, Singapore, pp. 511–522. https://doi.org/10.1007/978-981-10-1675-2_51.

Chen, X., Chen, T., Zhao, J., Wang, Yaojun, 2024. AgriBERT: A Joint Entity Relation Extraction Model Based on Agricultural Text, in: Cao, C., Chen, H., Zhao, L., Arshad, J., Asyhari, T., Wang, Yonghao (Eds.), Knowledge Science, Engineering and Management. Springer Nature, Singapore, pp. 254–266. https://doi.org/10.1007/978-981-97-5495-3_19.

Chen, X., Jia, S., Xiang, Y., 2020. A review: Knowledge reasoning over knowledge graph. Expert Systems with Applications 141, 112948. https://doi.org/10.1016/j.eswa.2019.112948.

Chen, Z., Chen, J., Singh, A., Sra, M., 2024. XplainLLM: A Knowledge-Augmented Dataset for Reliable Grounded Explanations in LLMs, arXiv:2311.08614.

Cheng, E., Wang, F., Peng, D., Zhang, B., Zhao, B., Zhang, W., Hu, J., Lou, Z., Yang, S., Zhang, H., Lv, Y., 2024. A GT-LSTM Spatio-Temporal Approach for Winter Wheat Yield Prediction: From the Field Scale to County Scale. IEEE Trans. Geosci. Remote Sensing 62, 1–18. https://doi.org/10.1109/TGRS.2024.3418046.

Chhetri, T.R., Hohenegger, A., Fensel, A., Kasali, M.A., Adekunle, A.A., 2023. Towards improving prediction accuracy and user-level explainability using deep learning and knowledge graphs: A study on cassava disease. Expert Systems with Applications 233, 120955. https://doi.org/10.1016/j.eswa.2023.120955.

Gao, M., Wang, H., Shen, W., Su, Z., Liu, H., Yin, Y., Zhang, Y., Zhang, Y., 2021. Disease Diagnosis of Dairy Cow by Deep Learning Based on Knowledge Graph and Transfer Learning. Int J Bioautomation 25, 87–100. https://doi.org/10.7546/ijba.2021.25.1.000812.

Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., Sen, P., 2020. A Survey of the State of Explainable AI for Natural Language Processing, arXiv:2010.00711.

Darnala, B., Amardeilh, F., Roussey, C., Todorov, K., Jonquet, C., 2023. C3PO: a crop planning and production process ontology and knowledge graph. Front. Artif. Intell. 6, 1187090. https://doi.org/10.3389/frai.2023.1187090.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805.

Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C.-M., Chen, W., Yi, J., Zhao, W., Wang, X., Liu, Z., Zheng, H.-T., Chen, J., Liu, Y., Tang, J., Li, J., Sun, M., 2023. Parameter-efficient fine-tuning of large-scale pre-trained language models. Nat Mach Intell 5, 220–235. https://doi.org/10.1038/s42256-023-00626-4.

Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C.-M., Chen, W., Yi, J., Zhao, W., Wang, X., Liu, Z., Zheng, H.-T., Chen, J., Liu, Y., Tang, J., Li, J., Sun, M., 2022. Delta Tuning: A Comprehensive Study of Parameter Efficient Methods for Pre-trained Language Models, arXiv:2203.06904.

Dong, J., Shu, H., Liu, R., Long, W., Zhang, G., 2024. Large Language Models Empowering Scenario Generation and Dual-layer Optimization for Coupled Operations of Power Supply, Irrigation, and Water Storage in Multiple Agricultural Parks. High Voltage Engineering 50, 2906–2917. https://doi.org/10.13336/j.1003-6520.hve.20240871.

Dong, Z., Wang, Y., Ma, C., Li, Z., 2022. Food Safety Relation Extraction Model Based on Adversarial Training and Capsule Network. Science Technology and Engineering 22, 10162–10168.

Farquhar, A., Fikes, R., Rice, J., 1997. The Ontolingua Server: a tool for collaborative ontology construction. International Journal of Human-Computer Studies 46, 707–727. https://doi.org/10.1006/ijhc.1996.0121.

Ge, W., Zhou, J., Zheng, P., Yuan, L., Rottok, L.T., 2024. A recommendation model of rice fertilization using knowledge graph and case-based reasoning. Computers and Electronics in Agriculture 219, 108751. https://doi.org/10.1016/j.compag.2024.108751.

Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubézy, M., Eriksson, H., Noy, N.F., Tu, S.W., 2003. The evolution of Protégé: an environment for knowledge-based systems development. International Journal of Human-Computer Studies 58, 89–123. https://doi.org/10.1016/S1071-5819(02)00127-1.

Goldstein, A., Fink, L., Ravid, G., 2021. A Framework for Evaluating Agricultural Ontologies. Sustainability 13, 6387. https://doi.org/10.3390/su13116387.

Gruber, T.R., 1993. A translation approach to portable ontology specifications. Knowledge Acquisition 5, 199–220. https://doi.org/10.1006/knac.1993.1008.

Gunaratnam, A., Thayananthan, T., Thangathurai, K., Abhiram, B., 2024. Computer vision in livestock management and production, in: Tarafdar, A., Pandey, A., Gaur, G.K., Singh, M., Pandey, H.O. (Eds.), Engineering Applications in Livestock Production. Academic Press, pp. 93–128. https://doi.org/10.1016/B978-0-323-98385-3.00002-5.

Guo, W., Yang, Y., Wu, H., Zhu, H., Miu, Y., Gu, J., 2024. Big Models in Agriculture: Key Technologies, Application and Future Directions. Smart Agriculture 6, 1–13. https://doi.org/10.12133/j.smartag.SA202403015.

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G.D., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., Ngomo, A.-C.-N., Polleres, A., Rashid, S.M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., Zimmermann, A., 2022. Knowledge Graphs. ACM Comput. Surv. 54, 1–37. https://doi.org/10.1145/3447772.

Hou, C., Niu, P., 2024. Review of Research Status and Prospects of Agricultural Knowledge Graphs. Transactions of the Chinese Society for Agricultural Machinery 55, 1–17. https://doi.org/10.6041/j.issn.1000-1298.2024.06.001.

Hu, H., Gao, J., Liu, Z., 2023. Research and Construction of Genetic Knowledge Graph of Milk Yield Traits in Dairy Cows. Computer Engineering and Applications 59, 299–305. https://doi.org/10.3778/j.issn.1002-8331.2108-0005.

Huang, B., Wu, S., Wang, W., Yang, Y., Liu, J., Zhang, Z., Chen, N., Yang, H., 2024. KG-LLM-MCom: A Survey on Integration of Knowledge Graph and Large Language Model. Journal of Wuhan University (Natural Science Edition) 70, 397–412. https://doi.org/10.14188/j.1671-8836.2024.0040.

Huang, Q., Tao, Y., Ding, S., Liu, Y., Marinello, F., 2023. An Intelligent Q&A Module for Tea Diseases and Pests Based on Automatic Knowledge Graph Construction. In: 2023 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor). Presented at the 2023 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor). IEEE, Pisa, Italy, pp. 66–70. https://doi.org/10.1109/MetroAgriFor58484.2023.10424245.

Huang, Y., 2024. Evaluating and Enhancing Large Language Models for Conversational Reasoning on Knowledge Graphs, arXiv:2312.11282.

Jin, Y., Liu, J., Wang, X., Li, P., Wang, J., 2021. Technology Recommendations for an Innovative Agricultural Robot Design Based on Technology Knowledge Graphs. Processes 9, 1905. https://doi.org/10.3390/pr9111905.

Jing, R., Li, P., 2024. Knowledge graph for integration and quality traceability of agricultural product information. Front. Sustain. Food Syst. 8, 1389945. https://doi.org/10.3389/fsufs.2024.1389945.

Kpodo, J., Kordjamshidi, P., Nejadhashemi, A.P., 2024. AgXQA: A benchmark for advanced Agricultural Extension question answering. Computers and Electronics in Agriculture 225, 109349. https://doi.org/10.1016/j.compag.2024.109349.

Kuska, M.T., Wahabzada, M., Paulus, S., 2024. AI for crop production – Where can large language models (LLMs) provide substantial value? Computers and Electronics in Agriculture 221, 108924. https://doi.org/10.1016/j.compag.2024.108924.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., Kiela, D., 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, arXiv:2005.11401.

Li, A., Xu, Y., Chi, Y., 2023. Review of Ontology Construction and Applications. Information Studies: Theory & Application 46, 189–195. https://doi.org/10.16353/j.cnki.1000-7490.2023.11.024.

Lin, B.Y., Chen, X., Chen, J., Ren, X., 2019. KagNet: Knowledge-Aware Graph Networks for Commonsense Reasoning, arXiv:1909.02151.

Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., Wang, P., 2019. K-BERT: Enabling Language Representation with Knowledge Graph, arXiv:1909.07606.

Liu, Y., Yao, Y., Ton, J.-F., Zhang, X., Guo, R., Cheng, H., Klochkov, Y., Taufiq, M.F., Li, H., 2024. Trustworthy LLMs: a Survey and Guideline for Evaluating Large Language Models' Alignment, arXiv:2308.05374.

Logan IV, R.L., Liu, N.F., Peters, M.E., Gardner, M. and Singh, S., 2019. Barack's wife hillary: Using knowledge-graphs for fact-aware language modeling. arXiv preprint arXiv:1906.07241.

Luo, H., E, H., Tang, Z., Peng, S., Guo, Y., Zhang, W., Ma, C., Dong, G., Song, M., Lin, W., Zhu, Y., Luu, A.T., 2024. ChatKBQA: A Generate-then-Retrieve Framework for Knowledge Base Question Answering with Fine-tuned Large Language Models, in: Findings of the Association for Computational Linguistics ACL 2024. Presented at the Findings of the Association for Computational Linguistics ACL 2024, Association for Computational Linguistics, Bangkok, Thailand and virtual meeting, pp. 2039–2056. https://doi.org/10.18653/v1/2024.findings-acl.122.

Lv, B., Wu, H., Chen, W., Chen, C., Miao, Y., Zhao, C., 2024. VEG-MMKG: Multimodal knowledge graph construction for vegetables based on pre-trained model extraction. Computers and Electronics in Agriculture 226, 109398. https://doi.org/10.1016/j.compag.2024.109398.

Mawkhiew, H.B., Sahoo, L., Kharshiing, E.V., 2021. Gene-to-trait knowledge graphs show association of plant photoreceptors with physiological and developmental processes that can confer agronomic benefits. Genet Resour Crop Evol 68, 2727–2735. https://doi.org/10.1007/s10722-021-01214-4.

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

Nismi Mol, E.A., Santosh Kumar, M.B., 2024. End-to-end framework for agricultural entity extraction – A hybrid model with transformer. Computers and Electronics in Agriculture 225, 109309. https://doi.org/10.1016/j.compag.2024.109309.

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. https://doi.org/10.1109/TKDE.2024.3352100.

Peng, R., Liu, K., Yang, P., Yuan, Z., Li, S., 2023. Embedding-based Retrieval with LLM for Effective Agriculture Information Extracting from Unstructured Data, arXiv:2308.03107.

Polak, M.P., Morgan, D., 2024. Extracting Accurate Materials Data from Research Papers with Conversational Language Models and Prompt Engineering. Nat Commun 15, 1569. https://doi.org/10.1038/s41467-024-45914-8.

Nath, R.P.D., Das, T.R., Das, T.C., Raihan, S.M., 2024. Knowledge Graph Generation and Enabling Multidimensional Analytics on Bangladesh Agricultural Data. IEEE Access 12, 87512–87531. https://doi.org/10.1109/ACCESS.2024.3416388.

Qiao, B., Zou, Z., Huang, Y., Fang, K., Zhu, X., Chen, Y., 2022. A joint model for entity and relation extraction based on BERT. Neural Comput. Appl. 34, 3471–3481. https://doi.org/10.1007/s00521-021-05815-z.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., 2018. Improving Language Understanding by Generative Pre-Training.

Razuvayevskaya, O., Wu, B., Leite, J.A., Heppell, F., Srba, I., Scarton, C., Bontcheva, K., Song, X., 2024. Comparison between parameter-efficient techniques and full fine-tuning: A case study on multilingual news article classification. PLoS ONE 19, e0301738. https://doi.org/10.1371/journal.pone.0301738.
Rodríguez-Muro, M., Rezk, M., 2015. Efficient SPARQL-to-SQL with R2RML mappings. Journal of Web Semantics 33, 141–169. https://doi.org/10.1016/j.websem.2015.03.001.
Shadbolt, N., Berners-Lee, T., Hall, W., 2006. The Semantic Web Revisited. IEEE Intell. Syst. 21, 96–101. https://doi.org/10.1109/MIS.2006.62.
Silva, B., Nunes, L., Estevão, R., Aski, V., Chandra, R., 2023. GPT-4 as an Agronomist Assistant? Answering Agriculture Exams Using Large Language Models, arXiv:2310.06225.
Singhal, 2012. Official Google Blog: Introducing the Knowledge Graph: things, not strings.
Song, X., Jiang, Q., Zhang, W., Wu, X., 2024. Food Knowledge Graph-based Intelligent Perception Question Answering. Journal of Chinese Institute of Food Science and Technology 24, 1–12. https://doi.org/10.16429/j.1009-7848.2024.12.001.
Stuart, L., Hobbins, M., Niebuhr, E., Ruane, A.C., Pulwarty, R., Hoell, A., Thiaw, W., Rosenzweig, C., Muñoz-Arriola, F., Jahn, M., Farrar, M., 2024. Enhancing Global Food Security: Opportunities for the American Meteorological Society. Bulletin of the American Meteorological Society 105, E760–E777. https://doi.org/10.1175/BAMS-D-22-0106.1.
Suchanek, F.M., Kasneci, G., Weikum, G., 2008. YAGO: A Large Ontology from Wikipedia and WordNet. Web Semant. 6, 203–217. https://doi.org/10.1016/j.websem.2008.06.001.
Sun, X., Ji, Y., Ma, B., Li, X., 2023. A Comparative Study between Full-Parameter and LoRA-based Fine-Tuning on Chinese Instruction Data for Instruction Following Large Language Model, arXiv:2304.08109.
Sun, Y., Shi, Q., Qi, L., Zhang, Y., 2022. JointLK: Joint Reasoning with Language Models and Knowledge Graphs for Commonsense Question Answering, in: Carpuat, M., de Marneffe, M.-C., Meza Ruiz, I.V. (Eds.), Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Presented at the NAACL-HLT 2022, Association for Computational Linguistics, Seattle, United States, pp. 5049–5060. https://doi.org/10.18653/v1/2022.naacl-main.372.
Sun, Y., Wang, S., Feng, S., Ding, S., Pang, C., Shang, J., Liu, J., Chen, X., Zhao, Y., Lu, Y., Liu, Weixin, Wu, Z., Gong, W., Liang, J., Shang, Z., Sun, P., Liu, Wei, Ouyang, X., Yu, D., Tian, H., Wu, H., Wang, H., 2021. ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation, arXiv:2107.02137.
Tamašauskaitė, G., Groth, P., 2023. Defining a Knowledge Graph Development Process Through a Systematic Review. ACM Trans. Softw. Eng. Methodol. 32, 1–40. https://doi.org/10.1145/3522586.
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 29, 1930–1940. https://doi.org/10.1038/s41591-023-02448-8.
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 4, 941–948. https://doi.org/10.1038/s43016-023-00867-x.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30 (NIPS 2017), Advances in Neural Information Processing Systems. Presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS), Neural Information Processing Systems (nips), La Jolla.
Vrandecic, D., Krtoetzsch, M., 2014. Wikidata: a free collaborative knowledgebase. Communications of the Acm 57, 78–85. https://doi.org/10.1145/2629489.
Wang, C., Wang, F., 2014. Study on Recognition of Chinese Agricultural Named Entity with Conditional Random Fields. Journal of Hebei Agricultural University 37, 132–135. https://doi.org/10.13320/j.cnki.jauh.2014.0027.
Wang, H., Zhao, R., 2024. Knowledge graph of agricultural engineering technology based on large language model. Displays 85, 102820. https://doi.org/10.1016/j.displa.2024.102820.
Wang, K., Miao, Y., Wang, X., Li, Y., Li, F., Song, H., 2024a. Research on the construction of a knowledge graph for tomato leaf pests and diseases based on the named entity recognition model. Front. Plant Sci. 15, 1482275. https://doi.org/10.3389/fpls.2024.1482275.
Wang, L., Chen, X., Deng, X., Wen, H., You, M., Liu, W., Li, Q., Li, J., 2024b. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. npj Digit. Med. 7, 1–9. https://doi.org/10.1038/s41746-024-01029-4.
Wang, M., Yin, T., Yan, H., Hu, J., Chen, J., 2023a. Knowledge Graphs and Large Language Models technology development and application. Cyber Security and Data Governance 42, 126–131. https://doi.org/10.19358/j.issn.2097-1788.2023.S1.022.
Wang, T., Gu, H., Wu, Z., Gao, J., 2020. Multi-source knowledge integration based on machine learning algorithms for domain ontology. Neural Comput & Applic 32, 235–245. https://doi.org/10.1007/s00521-018-3806-5.
Wang, T., Wang, N., Cui, Y., Liu, J., 2023b. Agricultural technology knowledge intelligent question-answering system based on large language model. Smart Agriculture 5, 105–116. https://doi.org/10.12133/j.smartag.SA202311005.
Wang, X., Gao, T., Zhu, Z., Zhang, Z., Liu, Z., Li, J., Tang, J., 2021. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. Transactions of the Association for Computational Linguistics 9, 176–194. https://doi.org/10.1162/tacl_a_00360.
Wei, X., Cui, X., Cheng, N., Wang, X., Zhang, X., Huang, S., Xie, P., Xu, J., Chen, Y., Zhang, M., Jiang, Y., Han, W., 2024. ChatIE: Zero-Shot Information Extraction via Chatting with ChatGPT, arXiv:2302.10205.
Wei, Y., Huang, Q., Kwok, J.T., Zhang, Y., 2023. KICGPT: Large Language Model with Knowledge in Context for Knowledge Graph Completion. Findings of the Association for Computational Linguistics: EMNLP 2023, 8667–8683. https://doi.org/10.18653/v1/2023.findings-emnlp.580.
Wilson, R.S.I., Goonetillake, J.S., Indika, W.A., Ginige, A., 2023. A conceptual model for ontology quality assessment: A systematic review. SW 14, 1051–1097. https://doi.org/10.3233/SW-233393.
Wu, D., Liu, X., Zai, S., Zhang, L., Feng, X., 2024. Analysis of Agrometeorological Hazard Based on Knowledge Graph. Agriculture 14, 1130. https://doi.org/10.3390/agriculture14071130.
Xu, D., Lu, W., Xu, R., Zhang, H., Wang, Y., You, L., Feng, Z., 2023. A method of deciding precision fertilization of rice based on spatio-temporal multi-modal knowledge graph of agriculture. Journal of Huazhong Agricultural University 42, 281–292. https://doi.org/10.13300/j.cnki.hnlkxb.2023.03.032.
Xu, Z., Sheng, Y., He, L., Wang, Y., 2016. Review on Knowledge Graph Techniques. Journal of University of Electronic Science and Technology of China 45, 589–606. https://doi.org/10.3969/j.issn.1001-0548.2016.04.012.
Yan, R., An, P., Meng, X., Li, Y., Li, D., Xu, F., Dang, D., 2025. A knowledge graph for crop diseases and pests in China. Sci Data 12, 222. https://doi.org/10.1038/s41597-025-04492-0.
Yang, X., Gao, J., Xue, W., Alexandersson, E., 2024. PLLaMa: An Open-source Large Language Model for Plant Science, arXiv:2401.01600.
Yasunaga, M., Ren, H., Bosselut, A., Liang, P., Leskovec, J., 2022. QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering, arXiv:2104.06378.
Yu, P., Lin, B., 2024. A Framework for Agricultural Intelligent Analysis Based on a Visual Language Large Model. Applied Sciences 14, 8350. https://doi.org/10.3390/app14188350.
Zeng, W., Zhao, X., Li, X., Tang, J., Wang, W., 2022. On entity alignment at scale. The VLDB Journal 31, 1009–1033. https://doi.org/10.1007/s00778-021-00703-3.
Zhang, J., Chen, B., Zhang, L., Ke, X., Ding, H., 2021a. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. AI Open 2, 14–35. https://doi.org/10.1016/j.aiopen.2021.03.001.
Zhang, J., Jiang, Q., Du, Z., Geng, Y., Hu, Y., Tong, Q., Song, Y., Zhang, H.-Y., Yan, X., Feng, Z., 2024. Knowledge graph-derived feed efficiency analysis via pig gut microbiota. Sci Rep 14, 13939. https://doi.org/10.1038/s41598-024-64835-6.
Zhang, S., Wang, Z., Wang, Z., 2020. Prediction of wheat stripe rust disease by combining knowledge graph and bidirectional long short term memory network. Transactions of the Chinese Society of Agricultural Engineering 36, 172–178. https://doi.org/10.11975/j.issn.1002-6819.2020.12.021.
Zhang, X., Li, Z., Sun, Y., 2021b. Food Case Retrieval Method Based on Knowledge Graph and BERT. Computer Applications and Software 38, 137–146. https://doi.org/10.3969/j.issn.1000-386x.2021.07.024.
Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., Liu, Q., 2019. ERNIE: Enhanced Language Representation with Informative Entities, arXiv:1905.07129.
Zhao, B., Jin, W., Del Ser, J., Yang, G., 2023. ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification. Neurocomputing 557, 126708. https://doi.org/10.1016/j.neucom.2023.126708.
Zhao, C., 2023. Agricultural Knowledge Intelligent Service Technology: A Review. Smart Agriculture 5, 126–148. https://doi.org/10.12133/j.smartag.SA202306002.
Zhao, J., Sun, S., Guo, J., Zhong, Y., Wang, H., Wang, M., Qin, Y., 2022a. RESEARCH ON THE CONSTRUCTION OF KNOWLEDGE GRAPH AND RECOMMENDATION TECHNOLOGY OF AGRICULTURAL ONLINE LEARNING RESOURCES. Computer Applications and Software 39, 69–75. https://doi.org/10.3969/j.issn.1000-386x.2022.08.010.
Zhao, P., Wang, W., Liu, H., Han, M., 2022b. Recognition of the Agricultural Named Entities with Multifeature Fusion Based on ALBERT. IEEE Access 10, 98936–98943. https://doi.org/10.1109/ACCESS.2022.3206017.
Zhao, R., Yang, C., Zheng, J., Li, J., Wang, J., 2022c. Agricultural Intelligent Knowledge Service: Overview and Future Perspectives. Smart Agriculture 4, 105–125. https://doi.org/10.12133/j.smartag.SA202207009.
Zhao, X., Chen, B., Ji, M., Wang, X., Yan, Y., Zhang, J., Liu, S., Ye, M., Lv, C., 2024. Implementation of Large Language Models and Agricultural Knowledge Graphs for Efficient Plant Disease Detection. Agriculture 14, 1359. https://doi.org/10.3390/agriculture14081359.
Zhou, W., Yu, J., Wang, R., Deng, H., Fan, C., 2022. Construction and Application of Knowledge Map of Case Library on Construction Machinery Equipment. Machinery 60, 1–9+18.
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. Computers and Electronics in Agriculture 229, 109824. https://doi.org/10.1016/j.compag.2024.109824.