Review article

Foundation models in smart agriculture: Basics, opportunities, and challenges

Jiajia Li [a], Mingle Xu [b], Lirong Xiang [c], Dong Chen [a], Weichao Zhuang [d], Xunyuan Yin [e,*], Zhaojian Li [f,*]

[a] *Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA*
[b] *Department of Electronic Engineering, Core Research Institute of Intelligent Robots, Jeonbuk National University, South Korea*
[c] *Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, NC, USA*
[d] *School of Mechanical Engineering, Southeast University, Nanjing, China*
[e] *School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, Singapore*
[f] *Department of Mechanical Engineering, Michigan State University, East Lansing, MI, USA*

ARTICLE INFO

ABSTRACT

The past decade has witnessed the rapid development and adoption of machine and deep learning (ML & DL) methodologies in agricultural systems, showcased by great successes in applications such as smart crop management, smart plant breeding, smart livestock farming, precision aquaculture farming, and agricultural robotics. However, these conventional ML/DL models have certain limitations: they heavily rely on large, costly-to-acquire labeled datasets for training, require specialized expertise for development and maintenance, and are mostly tailored for specific tasks, thus lacking generalizability. Recently, large pre-trained models, also known as foundation models (FMs), have demonstrated remarkable successes in language, vision, and decision-making tasks across various domains. These models are trained on a vast amount of data from multiple domains and modalities. Once trained, they can accomplish versatile tasks with just minor fine-tuning and minimal task-specific labeled data. Despite their proven effectiveness and huge potential, there has been little exploration of applying FMs to agriculture artificial intelligence (AI). Therefore, this study aims to explore the potential of FMs in the field of smart agriculture. In particular, conceptual tools and technical background are presented to facilitate the understanding of the problem space and uncover new research directions in this field. To this end, recent FMs in the general computer science (CS) domain are reviewed, and the models are categorized into four categories: language FMs, vision FMs, multimodal FMs, and reinforcement learning FMs. Subsequently, the process of developing agriculture FMs (AFMs) is outlined and their potential applications in smart agriculture are discussed. In addition, the unique challenges and risks associated with developing AFMs are discussed, including model training, validation, and deployment. Through this study, the advancement of AI in agriculture is explored by introducing AFMs as a promising paradigm that can significantly mitigate the reliance on extensive labeled datasets and enhance the efficiency, effectiveness, and generalization of agricultural AI systems. To facilitate further research, a well-classified and actively updated list of papers on AFMs is organized and accessible at https://github.com/JiajiaLi04/Agriculture-Foundation-Models.

1. Introduction

Modern agricultural systems have been undergoing a revolution through the tight integration of recent advances in information and communication technologies (ICT). This has led to a prosperous field known as smart farming (also referred to as smart agriculture) (Gondchawar et al., 2016; Wolfert et al., 2017; Walter et al., 2017), showing promise in enhancing agricultural outputs, increasing farming efficiency, and improving the quality of the final product. These technologies feature multi-disciplinary advancements involving unmanned aerial/ground vehicles (UAVs/UGVs), image processing, machine learning, big data, cloud computing, and wireless sensor networks (WSNs) (Walter et al., 2017; Wolfert et al., 2017; Moysiadis et al., 2021), enabling farmers to make well-informed decisions regarding planting,

---

tending, and harvesting to maximize productivity and profits. However, efficient extraction of relevant information from diverse data sources, especially imaging data, presents a significant challenge. Traditional data mining approaches often struggle to uncover meaningful insights from such complex data (Wolfert et al., 2017). For instance, in agriculture, these techniques may face difficulties in extracting relevant features from datasets comprising diverse variables such as soil composition, weather patterns, crop health metrics, and geographical information (Wolfert et al., 2017).

Deep learning (DL), on the other hand, has demonstrated remarkable capabilities in processing complex and high-dimensional data due to its ability to learn hierarchical representations of data (LeCun et al., 2015). Unlike traditional data mining techniques, which often rely on handcrafted feature extraction methods, DL models can automatically learn relevant features from raw data through multiple layers of abstraction. DL methods excel in feature extraction, pattern classification, and learning high-quality image representations, showing promising performance across various agricultural domains (Kamilaris and Prenafeta-Boldú, 2018). These applications include weed control (Hasan et al., 2021; Chen et al., 2022; Li et al., 2024; Chen et al., 2024), plant disease detection (Mohanty et al., 2016; Xu et al., 2021), plant counting (Pathak et al., 2022; Li et al., 2023c), plant phenotyping (Li et al., 2020; Xiang and Wang, 2023), postharvest quality assessment (Mendigoria et al., 2021; Zhou et al., 2022), robotic fruit harvesting (Chu et al., 2021; Zhang et al., 2022a; Chu et al., 2023), among others. Despite the progress, these approaches heavily rely on supervised training, which hinges on large-scale, task-specific, and high-quality labeled datasets (Sun et al., 2017). However, collecting and annotating such datasets is extremely time-consuming, resource-intensive, and expensive, thereby, creating a high bar for resource-limited applications (Lu and Young, 2020; Xu et al., 2023a). This challenge is particularly pronounced in specific applications such as plant disease detection, weed recognition, and fruit defect detection, where constraints related to biological materials, imaging conditions, and the need for precise annotations (Lu and Young, 2020). Moreover, the collected datasets are often limited in their generalizability to even similar agricultural domains or applications, necessitating repetition in the data collection and model development process (Ghazi et al., 2017; Thenmozhi and Reddy, 2019). This repetitive process not only adds to the overall time and monetary costs but also hinders the generalizability, efficiency, and scalability of DL frameworks in agricultural applications.

To mitigate the aforementioned issues, many approaches have been proposed, including transfer learning (Weiss et al., 2016; Zhuang et al., 2020), few-shot learning (Wang et al., 2020; Yang et al., 2022a), label-efficient learning (Zhou, 2018; Li et al., 2023a), among others. Specifically, transfer learning has emerged as a widely adopted approach in agricultural applications and beyond (Espejo-Garcia et al., 2020; Chen et al., 2022; Abbas et al., 2021; Dang et al., 2022), which leverages DL models pre-trained on large-scale image datasets, such as ImageNet (Deng et al., 2009), Microsoft COCO (Lin et al., 2014), and PlantCLEF2022 (Goëau et al., 2022). The pre-trained models are then fine-tuned for specific tasks. For instance, 35 state-of-the-art DL models are evaluated for weed classification using transfer learning approach (Chen et al., 2022). Few-shot learning, on the other hand, focuses on training models to quickly adopt new tasks or domains using only a limited number of labeled examples, by leveraging prior knowledge or meta-learning approaches. In contrast, label-efficient learning (Li et al., 2023a) employs weak supervision or no supervision techniques to mitigate the effects of laborious and time-consuming labeling challenges. However, these methods are often pre-trained on data from a single domain and modality, which restricts their adaptability to other domains and applications (Raffel et al., 2020). The above observations highlight the importance of incorporating diverse data sources to enhance their generalization capabilities across different domains and applications.

Large pre-trained models are also referred to as FMs (Bommasani et al., 2021; Moor et al., 2023; Mai et al., 2023) designed to produce a wide and general variety of outputs, undergo extensive training on vast and diverse datasets. FMs are enabled to tackle numerous downstream tasks, such as speech recognition and natural language understanding (Radford et al., 2019; Brown et al., 2020; OpenAI, 2022; Ouyang et al., 2022; Touvron et al., 2023), computer vision (Rombach et al., 2022; Ramesh et al., 2022; Kirillov et al., 2023; Kang et al., 2023), and decision making (Reed et al., 2022; Team et al., 2023b). FMs are generally trained on large-scale datasets from diverse domains and modalities through self-supervised learning (Liu et al., 2021), and are capable of handling a wide range of applications from various domains with minimum fine-tuning and no/little task-specific labeled data (Bommasani et al., 2021; Moor et al., 2023; Mai et al., 2023). For example, the Generative Pre-trained Transformer 3 language model (GPT-3, Brown et al., 2020) demonstrates a breakthrough in in-context learning, enabling the model to perform novel tasks without explicit training, by leveraging text explanations (known as "prompts") containing just a few examples. Subsequently, ChatGPT (OpenAI, 2022), a derivative of GPT-3, is developed and publicized by OpenAI for chat-based interactions and has brought about a transformative evolution in natural language processing, enabling unprecedentedly engaging conversational experiences. In the computer vision domain, FMs such as the segment anything model (SAM, Kirillov et al., 2023), trained on over 1 billion masks on 11M licensed and privacy-respecting images, can solve a range of downstream segmentation problems with new image distributions and tasks with zero-shot generalization where a model is trained to recognize classes or categories that it has never seen during training (Yang et al., 2022a). DeepMind Ada (Team et al., 2023b), equipped with a customized Transformer architecture (Vaswani et al., 2017) using distillation with a teacher-student approach (Schmitt et al., 2018), brings FMs to reinforcement learning (RL) that generalizes to new tasks in a few-shot setting. Despite the aforementioned advances, FMs' adoption in agriculture has received scarce attention.

As such, this study aims to explore the potential of developing and applying FMs for agricultural applications. The main contributions and the technical advancements of this paper are summarized as follows.

- Recent FMs (Section 2) in the general computer science (CS) domain are systematically reviewed and categorized into four key categories: language FMs, vision FMs, multimodal FMs, and reinforcement learning FMs.
- The process of developing AFMs (Fig. 1) is outlined and their potential applications (Section 3) in smart agriculture are discussed, shedding light on their diverse capabilities.
- This study also identifies and discusses the unique challenges and potential pitfalls associated with developing AFMs (Section 4), including model training, validation, and deployment.

This comprehensive examination of AFMs can serve as a valuable resource for newcomers to the field, as well as experienced researchers seeking to innovate within the agricultural space.

The nomenclature used in this article is presented in Table 1.

2. Foundation models (FMs)

In this section, different kinds of FMs, their distinct characteristics, and their applicability to various domains or tasks are explored.

2.1. Categories of FMs

In this subsection, a comprehensive review of LLMs & FMs within the computer science (CS) domain is presented, and the models are categorized into four distinct types based on the primary data modality and learning paradigm they address: language FMs (LFMs), vision FMs (VFMs), multimodal FMs (MFMs), and reinforcement learning FMs (RLFMs) (see Fig. 2). Readers who are familiar with these concepts or more interested in agricultural applications may skip this section and refer to Fig. 2 for a concise summary.
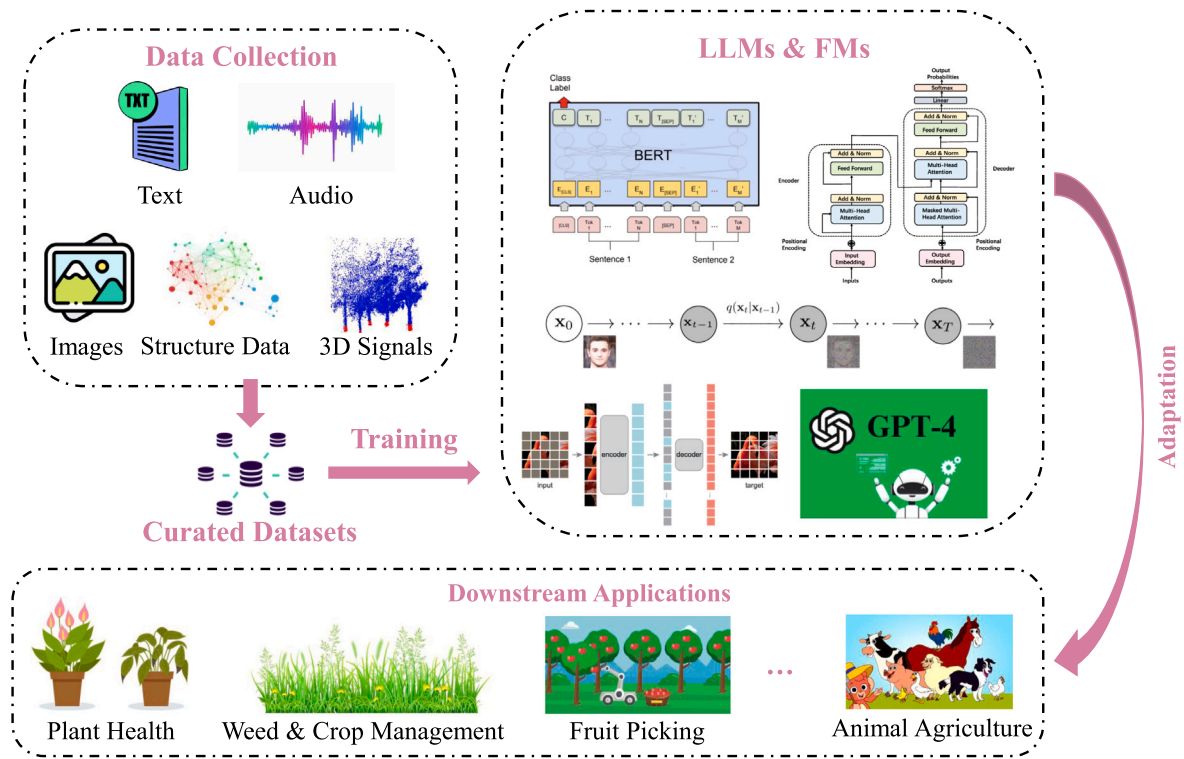
Fig. 1. An overview of building FMs for agricultural applications. The pipeline depicts the process of (multi-modal) data collection and cleaning, training or fine-tuning of LLMs and FMs, and downstream adaptation. The terms "BERT" and "GPT-4" refer to Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018) and Generative Pre-trained Transformer 4 (GPT-4) (Achiam et al., 2023), respectively.

Table 1
Nomenclature.

Nomenclature		InstructGPT	Instruction generative pretrained transformer (Ouyang et al., 2022)
ML	Machine learning	SAM	Segment anything model (Kirillov et al., 2023)
DL	Deep learning	SAA+	Segment any anomaly + (Cao et al., 2023b)
AI	Artificial intelligence	SEEM	Segment everything everywhere model (Zou et al., 2023)
GAN	Generative adversarial network	DMs	Diffusion models
VQA	Visual question answering	CLIP	Contrastive language-image pre-training (Radford et al., 2021)
FMs	Foundation models	ViT	Vision transformer
LFMs	Language foundation models	DALL·E2	Hierarchical text-conditional image generation with CLIP latent (Ramesh et al., 2022)
VFMs	Vision foundation models	GLIDE	Guided language to image diffusion for generation and editing (Nichol et al., 2021)
MFMs	Multimodal foundation models	GigaGAN	Large-scale GAN for text-to-image synthesis (Kang et al., 2023)
RLFMs	Reinforcement learning foundation models	BLIP	Bootstrapping language-image pre-training (Li et al., 2022)
BERT	Bidirectional encoder representations from transformers (Devlin et al., 2018)	VLP	Vision-language pre-training
PaLM	Pathways language model (Wei et al., 2022)	KOSMOS-1	A multimodal large language model (Huang et al., 2023)
LLAMA	Large language model meta AI (Touvron et al., 2023)	MLLM	Multimodal large language model
GLaM	Generalist language Model (Du et al., 2022)	DRL	Deep reinforcement learning
GPT-2	Generative pretrained transformer 2 (Radford et al., 2019)	DQN	Deep Q-network (Mnih et al., 2015)
GPT-3	Generative pretrained transformer 3 (Brown et al., 2020)	PPO	Proximal policy optimization (Schulman et al., 2017)
GPT-4	Generative pretrained transformer 4 (OpenAI, 2023)	Gato	A aeneralist agent (Reed et al., 2022)
Chat-GPT	Chat generative pretrained transformer (OpenAI, 2022)	AdA	Adaptive agent (Team et al., 2023b)

2.1.1. Language foundation models (LFMs)

Over the past few years, the field of computational natural language has seen a complete transformation (Devlin et al., 2018; Brown et al., 2020) due to the rise of LLMs. Essentially, language modeling (Nadkarni et al., 2011) involves predicting the subsequent element in a given sequence of tokens, relying on a self-supervised objective that requires no manual labeling beyond a raw text corpus.

The $\underline{B}$idirectional $\underline{E}$ncoder $\underline{R}$epresentations from $\underline{T}$ransformers (BERT, (Devlin et al., 2018)) is one of the most prominent models, with two versions: $\text{BERT}_{\text{BASE}}$ and $\text{BERT}_{\text{LARGE}}$, containing 110M and 340M parameters respectively. By simply incorporating an additional output layer for fine-tuning, the pre-trained BERT model can be adapted to generate cutting-edge models for a variety of tasks, such as question answering and language inference. Remarkably, this

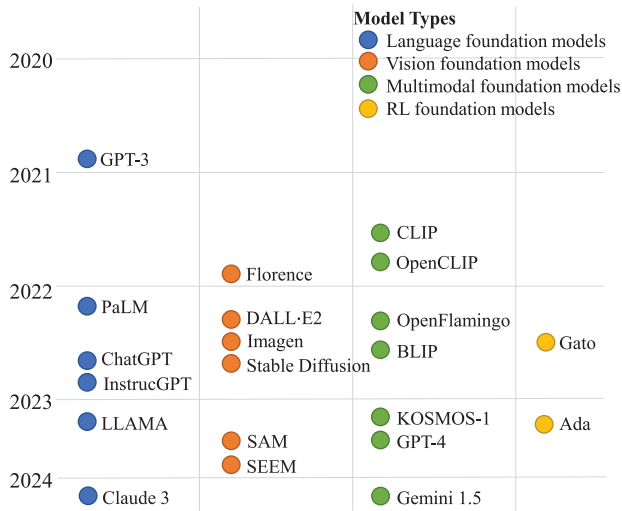## Foundation Models Timeline

Fig. 2. Timeline of FMs.

can be achieved without requiring substantial modifications to the task-specific architecture, even surpassing the performance of many dedicated task-specific architectures. Subsequently, Google proposes the Pathways Language Model (PaLM, Wei et al., 2022) and Large Language Model Meta AI (LLAMA, Touvron et al., 2023), further advancing the field of natural language processing. PaLM, with 540B parameters, is designed with a dense decoder-only Transformer model trained with the Pathways system (Barham et al., 2022) to enable efficient training. PaLM surpasses the few-shot performance of prior large models, such as GLaM (Du et al., 2022), GPT-3 (Brown et al., 2020), and LLAMA (Touvron et al., 2023), on numerous very difficult tasks, such as nature language inference (Bowman et al., 2015), in-context reading comprehension (Jenkins et al., 2003), and question answering (Kwiatkowski et al., 2019).

Participating in the development of large language models, OpenAI has proposed noteworthy contributions such as Generative Pretrained Transformer 2 (GPT-2, Radford et al., 2019), 3 (GPT-3, Brown et al., 2020), and 4 (GPT-4, OpenAI, 2023), having grown progressively more powerful, with 1.5 billion, 175 billion, and a staggering number of parameters, respectively, and setting remarkable benchmarks in language processing. ChatGPT (OpenAI, 2022), a specialized version of these models, has been extensively fine-tuned for human-like conversation. By adjusting the final output layer, the pre-trained GPT models can be tailored to deliver exceptional performance across a wide range of tasks, such as text generation, translation, and dialogue systems, outperforming many task-specific models without requiring extensive modifications to their foundational architecture. Recently, OpenAI has shown that InstructGPT (Ouyang et al., 2022) with only 1.3B parameters are much better at following users' instructions than GPT-3, by using an existing technique called reinforcement learning from human feedback (RLHF, Christiano et al., 2017). More recently, the development of Claude 3[1] represents the cutting-edge in AI systems, encompassing undergraduate-level expert knowledge (MMLU), graduate-level expert reasoning (GPQA), basic mathematics (GSM8K), and beyond. It demonstrates nearly human-like levels of comprehension and fluency across intricate tasks, marking the forefront of general intelligence (Anthropic, 2024).

With the continuous growth in data availability and computing power, language modeling has become a trustworthy method for crafting ever more powerful models.

2.1.2. Vision foundation models (VFMs)

Computer vision, a field dedicated to understanding and analyzing visual data, has traditionally relied on task-specific models tailored to specific objectives. These models, such as ResNet (He et al., 2016) for image classification, Faster RCNN (Ren et al., 2015) for object detection, Detectron (Girshick et al., 2018) for instance segmentation, and DCGAN (Radford et al., 2015) for image generation, have made significant contributions to advancing visual perception tasks. However, it is important to note that these task-specific models are designed to excel in their respective applications and may not be directly applicable to other computer vision tasks (Bommasani et al., 2021).

With the emergence of vision FMs, there is a paradigm shift towards more versatile and general-purpose models that can tackle multiple vision tasks simultaneously. These FMs leverage deep learning techniques, and diverse, and large-scale training data to offer a unified framework for various computer vision tasks, enabling researchers to explore new frontiers in visual understanding and analysis (Bommasani et al., 2021). For example, Florence (Yuan et al., 2021), a newly introduced computer vision FM, offers an extensive range of capabilities for visual understanding. By leveraging universal visual-language representations from vast image-text datasets, Florence seamlessly adapts to various computer vision tasks including classification, retrieval, object detection, visual question answering (VQA), image captioning, video retrieval, and action recognition. Notably, Florence achieves outstanding performance in transfer learning scenarios, excelling in fine-tuning, linear probing, few-shot transfer, and zero-shot transfer for novel images and objects. Recently, the Segment Anything Model (SAM, Kirillov et al., 2023) is introduced by Meta AI for image segmentation. With its flexible and efficient architecture, SAM achieved impressive zero-shot performance on numerous segmentation tasks (Zhang et al., 2023f; Ma and Wang, 2023; Yang et al., 2023a; Tang et al., 2023). To promote research in FMs for computer vision, the SA-1B dataset, the largest of its kind, providing over 1 billion masks from 11 million images, is open-sourced.[2] Inspired by SAM, the Segment Any Anomaly + (SAA+, Cao et al., 2023b) is proposed for zero-shot anomaly segmentation with hybrid prompt regularization to improve the adaptability of modern FMs. Furthermore, in Tang et al. (2023), the authors show that SAM shows promising performance in camouflaged object detection (COD) tasks, in which objects are seamlessly integrated into their surroundings. In Yang et al. (2023a), SAM demonstrates superior performance in the poultry industry, specifically in the context of segmenting cage-free hens, outperforming SegFormer (Xie et al., 2021) and SETR (Liu et al., 2022) in both whole and part-based chicken segmentation. The Segmenting Everything Everywhere Model (SEEM, Zou et al., 2023) represents a significant advancement in the field of image segmentation. It sets a new benchmark for simultaneously segmenting everything everywhere in an image, including semantic, instance, and panoptic segmentation (Kirillov et al., 2019). Notably, SEEM supports a wide range of prompts, including visual, textual, and referring region prompts, which can be combined in various ways to enable versatile and interactive referring segmentation. This model represents a breakthrough that pushes the boundaries of what is achievable in comprehensive image segmentation tasks.

Vision FMs have also demonstrated outstanding performance in the field of image generation, showcasing their ability to produce high-quality, diverse, and realistic images that push the boundaries of computer-generated visuals. The current attempts in this area can be categorized into two main branches: generative adversarial networks (GANs, Goodfellow et al., 2020) and diffusion probability models (diffusion models (DMs Yang et al., 2022c). Stable Diffusion (Rombach et al., 2022), a latent text-to-image DM, is introduced to generate detailed images conditioned on text descriptions. The model leverages a frozen CLIP ViT-L/14 text encoder (Radford et al., 2021) to incorporate

---

text prompts, combining an 860M UNet and a 123M text encoder. The Stable Diffusion model achieves state-of-the-art scores in various tasks, such as image inpainting and class-conditional image synthesis. Around the same time, OpenAI released DALL·E2 (Ramesh et al., 2022), another model for generating realistic images based on text inputs. DALL·E2 consists of a prior stage, responsible for generating a CLIP (Radford et al., 2021) image embedding from a given text caption, and a decoder stage, which generates an image conditioned on the generated image embedding. This architecture enables language-guided image manipulations in a zero-shot manner. Moreover, Saharia et al. (2022) introduces Imagen, which exhibits superior sample quality and image-text alignment compared to previous works such as VQ-GAN+CLIP (Crowson et al., 2022), Stable Diffusion (Rombach et al., 2022), GLIDE (Nichol et al., 2021) and DALL·E2 (Ramesh et al., 2022) with an efficient U-Net diffusion network architecture and powerful large transformer language models.

While the previous works discussed above have primarily focused on diffusion models (DMs), GigaGAN (Kang et al., 2023) represents a significant advancement in the realm of image generation by adopting a generative adversarial network (GAN) approach. GigaGAN offers significantly faster inference time compared to DM-based approaches (Rombach et al., 2022; Ramesh et al., 2022; Saharia et al., 2022) while supporting various applications like latent interpolation, style mixing, and vector arithmetic operations. Building upon the StyleGAN2 framework (Karras et al., 2020), GigaGAN enhances training stability and model capacity by retaining a bank of filters and employing a combination of self-attention (image-only) and cross-attention (image-text) with the convolutional layers for improved performance. Additionally, the authors introduce multi-scale training and multi-stage approaches to enhance image-text alignment and generate low-frequency details. GigaGAN, with 1.0B parameters, achieves a zero-shot Fréchet Inception Distance (FID) of 9.09 on the COCO2014 dataset, outperforming Imagen (3.0B) and DALL·E2 (5.5B) in terms of image quality evaluation.

Looking ahead, VFMs are expected to continue evolving and pushing the boundaries of what is possible in computer vision. As research progresses, it can be anticipated that such models will be further advanced with improved capabilities in terms of image understanding, scene understanding, and content generation. Furthermore, the fusion of vision models with other modalities, such as language and audio, is likely to lead to the emergence of more powerful and versatile multimodal models (Section 2.1.3).

2.1.3. Multimodal foundation models (MFMs)

MFMs represent a new class of AI models that leverage and integrate multiple types of data inputs, like text, video, audio, and images, which is a distinct departure from their non-multimodal counterparts that are constrained to processing a singular data type. By capitalizing on the richness and diversity of these data types, MFMs can attain a more holistic understanding of the context. This, in turn, empowers them to execute more nuanced and precise tasks, such as translating text into visual representations, identifying objects in images based on textual prompts, and creating image captions. As a result, these models can fuel a vast array of real-world applications (Bommasani et al., 2021; Moor et al., 2023; Mai et al., 2023).

One notable example is OpenAI's CLIP (Contrastive Language-Image Pretraining, Radford et al., 2021), which has garnered significant attention for its ability to learn joint representations of images and text. It utilizes a strategy known as contrastive learning (Zhao et al., 2021; Jaiswal et al., 2020), which entails learning to associate accurate text-image pairs while dissociating inaccurate ones. Consequently, CLIP excels at tasks such as zero-shot learning, wherein it can classify new images based on textual descriptions and vice versa. Its multimodal nature renders it highly advantageous in diverse domains, such as computer vision, natural language processing, and digital content creation. Building on this foundation, OpenCLIP (Ilharco et al., 2021), a

model trained on up to 2B image-text pairs, is proposed to investigate scaling laws (Kaplan et al., 2020) for CLIP (Radford et al., 2021) model with public datasets. OpenCLIP identifies power scaling laws through various downstream tasks, such as zero-shot retrieval, classification, linear probing, and end-to-end fine-tuning, concluding that the training distribution plays a key role in scaling laws. The new Gemini 1.5[3] developed by Google delivers dramatically enhanced performance with a more efficient architecture compared to its previous version (Team et al., 2023a).

However, a majority of FMs are typically engineered for either understanding-based or generation-based tasks. To address this issue, BLIP (Li et al., 2022), a new Vision-Language Pre-training (VLP) framework, is introduced for both vision-language understanding and generation tasks. To effectively utilize the noisy web data by bootstrapping the captions, a captioner is employed to generate synthetic captions, subsequently, a filter is adopted to remove the noisy ones. BLIP achieves state-of-the-art results across various vision-language tasks, such as image-text retrieval, image captioning, and VQA. Similarly, OpenFlamingo (Alayrac et al., 2022) is trained on large-scale multi-modal web corpora with arbitrarily interleaved text and images for a variety of image and video tasks via few-shot learning, such as image captioning and multiple-choice VQA (Alayrac et al., 2022). KOSMOS-1 (Huang et al., 2023), a Multimodal Large Language Model (MLLM), is subsequently introduced to process multimodal inputs, adhere to instructions, and perform in-context learning for language tasks and multimodal tasks, such as multimodal dialogue, image captioning, VQA, and image recognition with descriptions. Impressively, KOSMOS-1 demonstrates exemplary performance across various settings, such as zero-shot, few-shot, and multimodal chain-of-thought prompting, on various tasks without necessitating any gradient updates or fine-tuning. Built on previous versions, GPT-4 (OpenAI, 2023), a large-scale and multimodal model, is introduced by OpenAI to accept both image and text inputs and produce text outputs.

The future of MFMs is set to be exciting, with further advancements anticipated in several areas. They will likely include more sophisticated integration of various data modalities and enhanced transfer learning for broader applicability across different domains (Bommasani et al., 2021). Furthermore, efforts are expected to be directed towards minimizing data bias from different domains, improving fairness, and developing energy-efficient AI models (Bommasani et al., 2021).

2.1.4. Reinforcement learning foundation models (RLFMs)

Over the past decade, the field of decision-making has undergone a significant revolution due to the emergence of RL algorithms integrated with advanced DL/ML algorithms. These models, *e.g.*, deep RL (DRL) algorithms like DQN (Mnih et al., 2015) and PPO (Schulman et al., 2017), have enabled RL agents to learn complex behaviors and achieve remarkable performance in various domains, such as game playing (Mnih et al., 2013, 2015; Vinyals et al., 2017), robot control (Kober et al., 2013; Lee et al., 2020), and agricultural applications (Bu and Wang, 2019; Binas et al., 2019; Gandhi, 2022; Gautron et al., 2022). By interacting with the environment, these models learn optimal decision-making through trial and error. However, due to their reliance on learning tasks from scratch without comprehensive knowledge of vision, language, or other datasets, these methods often face challenges in terms of generalization and sample efficiency (Yang et al., 2023c).

There are increasing attempts to combine FMs and sequential decision-making to tackle complex real-world problems with better generalization. Gato Reed et al. (2022), a multi-modal, multi-task, and multi-embodiment generalist RL policy, is proposed by Google Deep-Mind with a 1.2B parameter decoder-only transformer architecture, trained on 604 distinct tasks with varying modalities, observations, and

---

[3] Gemini 1.5 website: https://deepmind.google/technologies/gemini/#introduction.

action specifications. Gato is able to play Atari, caption images, chat, and stack blocks with a real robot arm, using the same network and weights with little fine-tuning. Subsequently, AdA (Team et al., 2023b) proposes by Google DeepMind with 265M parameters that learn new tasks in an extended and 3D environment simulation as fast as humans based on a modified transformer architecture to store more information and enable efficient training.

Leveraging world knowledge from FMs, sequential decision making can solve tasks faster and generalize better (Yang et al., 2023c). FMs for decision making are still in their infancy and still facing significant challenges including the discrepancy in data modalities, uncertainties surrounding environment and task structures, and the absence of certain components in existing FMs and decision-making paradigms. Further advancements are necessary to address these challenges and unlock the full potential of FMs in decision making (Yang et al., 2023c; Reed et al., 2022).

3. Agricultural foundation models (AFMs)

In this section, the potential for creating FMs in agriculture is first examined, *i.e.*, AFMs. Then current applications of these models in the realm of smart agriculture are presented, followed by a discussion on emerging areas where these FMs are most applicable.

3.1. Developing FMs for agriculture

Fig. 1 presents a comprehensive overview of the pipeline for developing (multimodal) FMs in agriculture. This process principally encompasses stages such as data collection, dataset curation, model training, and downstream fine-tuning.

The development of MFMs for agricultural applications commences with the imperative phase of data collection and curation (Bommasani et al., 2021). This process necessitates the assembly of heterogeneous data spanning a range of modalities, including visual (*e.g.*, images and videos indicating the status of crops, pests, and disease symptoms), textual (*e.g.*, human instructions and pest identification documents), meteorological (*e.g.*, temperature, precipitation, and wind speed measurements), and possibly auditory data (*e.g.*, distinctive insect or animal sounds). Given the magnitude of the data involved, meticulous annotation and curation are indispensable to discriminate meaningful patterns and correlations. Such data labeling could encompass tasks such as the classification of crop and weed species, the annotation of disease symptoms present on plant foliage, or the categorization of weather conditions.

Subsequent to data curation is the model training phase. Capitalizing on contemporary advancements in ML and AI, such as transformer architectures (Vaswani et al., 2017) and convolutional neural networks (Gu et al., 2018), the model is trained to process and interpret multimodal inputs. These AI models are designed to associate visual depictions with their corresponding textual labels, comprehend textual descriptions, and generate an inclusive understanding of the agricultural ecosystem. Notably, the training of these models may be computationally demanding, which necessitates the use of advanced hardware and optimization algorithms to enhance performance, which will be discussed in Section 4.3.

The culmination of this process is the implementation of these trained models in downstream agricultural applications, such as plant health monitoring (Xu et al., 2022b), crop and weed management (Hu et al., 2021; Dang et al., 2023; Rai et al., 2023), fruit picking (Chu et al., 2021), and precision livestock farming (Yang et al., 2023a). For instance, in relation to plant health, the models can be utilized to identify early indicators of disease or pest infestations, thus facilitating prompt interventions to reduce crop loss. For weed management, the models can assist in distinguishing weeds from crops, enabling precision farming equipment to selectively eradicate unwanted flora.

Furthermore, in the domain of precision livestock farming, these models can be harnessed to monitor livestock health and behavior, optimize feeding schedules, and predict potential diseases. In the context of downstream tasks, transfer learning or fine-tuning (Weiss et al., 2016; Zhuang et al., 2020) of the FMs becomes paramount to adapt the generalized learning to specific agricultural applications, thereby enhancing model performance and efficiency. Through these varied applications, MFMs hold the potential to significantly transform the agricultural industry, thereby enhancing productivity, promoting sustainability, and ultimately boosting profitability.

3.2. Agricultural applications

In this subsection, the existing applications of integrating FMs are discussed and explored in Section 2 regarding the field of smart agriculture.

3.2.1. LFMs in agriculture

LFMs, also referred to as LLMs and detailed in Section 2.1.1, have transformed the realms of design, development, and deployment by enabling human-like interactions. Researchers are now increasingly integrating these LFMs into various stages of design and development for agricultural applications, as highlighted in Stella et al. (2023), Lu et al. (2023). Specifically, in Stella et al. (2023), the authors discuss the process of integrating LLMs into the design phase of robotic systems. They specifically illustrate the procedure to design a robotic gripper optimized for tomato picking. During the initial ideation phase, researchers consult LLMs, such as ChatGPT (OpenAI, 2022), to understand potential challenges and opportunities within the field. Based on this information, they select the most promising and intriguing pathways, subsequently narrowing down the design possibilities through further dialogue with the LLM. This collaborative interaction covers a wide spectrum of knowledge domains and various levels of abstraction, encompassing everything from conceptual thinking to technical implementation. Throughout this process, the human collaborator leverages the AI's knowledge to access information beyond their personal expertise. During the subsequent phase of the design process, which is more technically focused, these general directions need to be actualized into a tangible, fully functional robot. As of now, LLMs are not capable of producing complete computer-aided design (CAD) models, assessing code, or autonomously fabricating a robot. However, recent progress in AI research has indicated that these algorithms can provide substantial assistance in executing software (Chen et al., 2021), mathematical reasoning (Wolfram, 2023), and even in the generation of shapes (Ramesh et al., 2022). Although future AI methodologies will be capable of handling these tasks, at the moment, the technical implementation remains a team effort between AI models and humans. Finally, the human would fine-tune the code suggested by the LLM, finalize the CAD design, and oversee the actual fabrication of the robot. Figure 2(b) in Stella et al. (2023) shows the primary outputs generated by the LLM, as well as the real-world application of the AI-designed robotic gripper used for crop harvesting.

The authors in Lu et al. (2023) have deliberated on the prospective use cases of LLMs in organizing unstructured metadata, converting metadata from one format to another, and identifying potential errors during the data collection phase. They also envision the future generation of LLMs as exceptionally powerful tools for data visualization (Bubeck et al., 2023), offering invaluable assistance to researchers in extracting meaningful insights from vast volumes of phenotypic data. Furthermore, in Tzachor et al. (2023), the authors utilize GPT to revolutionize agricultural extension services by simplifying scientific knowledge and delivering personalized, location-specific, and data-driven agricultural recommendations. In Shutske (2023), Generative AI and LLMs, particularly ChatGPT, are investigated to elevate innovation and enhance agricultural safety, occupational health, and health promotion.
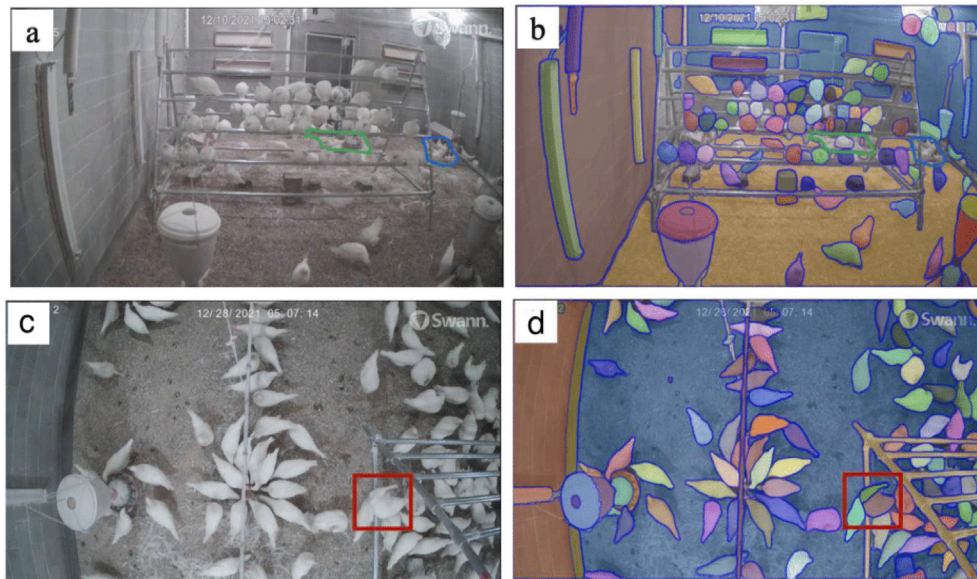
Fig. 3. Side-view (a,b) and top-view (c,d) views of chicken segmentation using SAM model in a research cage-free house proposed in Yang et al. (2023a).

Moreover, researchers have leveraged LLMs for various plant science applications. For example, in Yang et al. (2024), PLLaMa, an open-source LLM based on LLaMa-2, is developed specifically for plant science-related topics. Additionally, in Kuska et al. (2024), an AI-Chatbots for Agriculture is created to aid in interpreting decision support models in plant disease management.

3.2.2. VFMs in agriculture

VFMs (as detailed in Section 2.1.2), encompassing algorithms for core computer vision tasks like classification, detection, segmentation, and generative modeling, have revolutionized traditional computer vision methodologies. Trained on extensive image datasets, these VFMs can be efficiently adapted to novel domains, such as agriculture, with minimal fine-tuning and a limited set of labeled images. Recently, the agricultural community has shown a growing interest in VFMs, as seen in recent studies (Yang et al., 2023a; Williams et al., 2023).

In Yang et al. (2023a), SAM (Section 2.1.2) is employed for chicken segmentation tasks in a zero-shot manner using both part-based segmentation and the use of infrared thermal images (Fig. 3). Experimental results show that SAM outperformed the other VFMs algorithms, such as SegFormer and SETR, in both whole and partial chicken segmentation. For example, SAM achieves a mean Intersection over Union (mIoU) of 94.80% in the chicken segmentation tasks, while SegFormer and SETR only achieve mIoUs of 43.22% and 42.90%, which greatly shows the effectiveness of SAM on the poultry segmentation tasks. In addition, a novel single-bird tracking algorithm, integrating SAM, YOLOX (Ge et al., 2021) and ByteTracker (Zhang et al., 2022b), is proposed for tracking the activities for individual birds across video frames by extracting the location information (*i.e.*, bounding boxes provided by SAM) in each image and then tracking the birds over time with YOLOX and ByteTracker. Despite its promising performance, the tracking algorithm lacked end-to-end capability. The object tracking FMs (Yang et al., 2023b) show impressive performance in video object tracking and segmentation, requiring very little human participation, thereby holding the potential to serve as an end-to-end method.

In Williams et al. (2023), an automatic leaf segmentation pipeline, named "Leaf Only SAM", is introduced for zero-shot segmentation of potato leaves. When compared with a fine-tuned Mask R-CNN (He et al., 2017) model on the annotated potato leaf dataset, Leaf Only SAM demonstrates an average recall and precision of 63.2% and 60.3%, respectively. These results are lower than the Mask R-CNN model's average recall and precision of 78.7% and 74.7%, respectively. However, it is important to note that Leaf Only SAM does not necessitate any fine-tuning or an annotated dataset, thereby showing notable zero-shot generalization capabilities. These findings suggest that the direct application of FMs in a zero-shot manner may not always result in high performance due to potential distribution shifts (Chawla et al., 2021). This will be further discussed in Section 4.4.

3.2.3. MFMs in agriculture

MFMs (Section 2.1.3) introduce a novel class of AI models. These leverage and incorporate various types of data inputs including text, video, audio, and images. The agricultural community has begun to delve into multimodal learning within agricultural applications (Bender et al., 2020; Garg et al., 2021; Parr et al., 2021; Cao et al., 2023a; Dhakshayani and Surendiran, 2023) to enhance performance through the utilization of multimodal data. For instance, Bender et al. (2020) open-sources a multimodal dataset designed for agricultural robotics, which is gathered from cauliflower and broccoli fields, to facilitate robotics and machine learning research activity in agriculture. The dataset includes a variety of data such as stereo color, thermal and hy-perspectral imagery, and crucial environmental details like weather and soil conditions. In Cao et al. (2023a), a multi-modal language model based on image-text-label information is proposed for cucumber disease recognition. They effectively merge label information across multiple domains by employing image-text multimodal contrastive learning and image self-supervised contrastive learning. This methodology helps measure the distance of samples within the common image-text-label space. The experimental results reveal that this novel approach attains a recognition accuracy rate of 94.84% on a modestly sized multimodal cucumber disease dataset.

MFMs have also been discussed in the agriculture domain. Lu et al. (2023) investigates the potential of GPT-4, a multimodal language and vision model (as referenced in ChatGPT, Section 2.1.1), in achieving AGI. In Tan et al. (2023), GPT-4 is applied for geographical, environmental, agricultural, and urban planning applications, such as crop type identification.

However, it is noteworthy that current applications are predominantly based on text-image data, and their use is often confined to question-answering tasks. Indeed, there is a conspicuous gap in applications related to agricultural robotics where inputs such as images, text, or voice (*e.g.*, human instructions), and depth information (*e.g.*, from LiDAR or laser sensors) are used. These robots, which are typically

Table 2
Applications of LLMs and FMs in agriculture.

Model type	Reference	Problem	Method	Applications
LFMs	Stella et al. (2023)	Robotic picking	ChatGPT	Agriculture robots for crop harvesting, parasite control, and packaging
	Tzachor et al. (2023)	Agricultural recommendations	GPT	Agricultural extension services
	Lu et al. (2023)	Data processing	GPT	Precision livestock, precision crops, and agricultural robotics
	Yang et al. (2024)	Plant science	LLAMA	Question answering for plant science-related tasks
	Shutske (2023)	Agricultural safety & health	ChatGPT	Case studies for agricultural safety and health topics
	Kuska et al. (2024)	Ai-Chatbots	ChatGPT	Plant disease management
VFMs	Yang et al. (2023a)	Segmentation and tracking	SAM	Chicken segmentation and tracking
	Williams et al. (2023)	Segmentation	SAM	Potato leaves segmentation
MFMs	Cao et al. (2023a)	Disease recognition	Image-text data	Cucumber disease recognition
	Tan et al. (2023)	Remote Sensing	GPT4	Crop type identification

Note: As of the date of this review, there are no recorded applications of RLFMs in agriculture.

designed for tasks like fruit harvesting or crop monitoring (Tao and Zhou, 2017), could greatly benefit from the application of MFMs, which will be further discussed in Section 3.3.5.

3.2.4. RLFMs in agriculture

DRL (Arulkumaran et al., 2017; Henderson et al., 2018) is an emerging field that fuses deep learning with reinforcement learning to tackle decision-making tasks. Its impressive performance has been demonstrated across a broad spectrum of applications, from computer games (Mnih et al., 2013, 2015), to intelligent transportation systems (Chu et al., 2019; Chen et al., 2023), and robot control (Gu et al., 2017; Lee et al., 2020). Recently, the agricultural community has shown heightened interest in DRL (Binas et al., 2019; Gandhi, 2022), leading to a host of innovative applications such as crop management (Gautron et al., 2022; Din et al., 2022), agriculture robots (Yang et al., 2022b), and smart irrigation systems (Zhou, 2020). However, due to their reliance on learning tasks from scratch without comprehensive knowledge of vision, language, or other datasets, these methods often face challenges in terms of generalization and sample efficiency (Yang et al., 2023c).

Although there currently exist no published papers specifically focusing on the application of RLFMs (Section 2.1.4) in agricultural scenarios, the field nonetheless presents vast untapped opportunities for exploration and innovation. For instance, Gato Reed et al. (2022) and AdA (Team et al., 2023b), discussed in Section 2.1.4, exhibit promising potential for handling decision tasks that involve multimodal, multi-task contexts, paving the way for innovative agriculture applications in the future.

Table 2 summarizes the applications of LLMs and FMs in agriculture.

3.3. Emerging opportunities

In this subsection, some emerging areas in the agricultural space are discussed where these FMs are **most** applicable.

3.3.1. Smart crop management

To achieve sustainable farming, smart crop management (Eli-Chukwu, 2019; Jaramillo et al., 2020) has evolved as a novel research direction. It integrates advanced methodologies from AI, IoT, and big data into crop management tasks, including plant growth monitoring (Lakshmi and Gayathri, 2017), disease detection (Xu et al., 2022b), yield monitoring (Khanal et al., 2018; Kayad et al., 2019), and harvesting (Meshram et al., 2021). Yet, developing such algorithms for precision crop management is complex and challenging, particularly when considering factors like diverse crop varieties, environmental dynamics, and limited or hard-to-acquire training datasets. Furthermore, these systems often necessitate diverse data inputs sourced from varying domains. For example, while a deep learning method is developed on satellite imagery to assess crop health, it simultaneously might need soil sensor data to understand underground moisture levels and market data to predict the optimal time for harvest based on pricing trends.

Designing such algorithms requires deep research insights, making it challenging to smoothly incorporate knowledge from unrelated fields.

In light of the aforementioned challenges, FMs offer a promising solution. These models, trained on datasets spanning multiple domains, can be seamlessly adapted to new areas with minimal fine-tuning. They hold significant potential as tools for smart crop management. For instance, LFMs (Section 2.1.1) empower researchers to tap into knowledge beyond their personal expertise, facilitating software development and design (Stella et al., 2023). Moreover, MFMs (Section 2.1.3) integrate visual, textual, and sometimes auditory data, providing a comprehensive analytical lens, proving invaluable in deciphering intricate agricultural situations, notably in precision crop management.

3.3.2. Smart plant breeding

Smart plant breeding, a breakthrough approach, combines data from genotype, environment, and their interaction to optimize crop varieties (Xu et al., 2022c; Zhang et al., 2023e). Employing sophisticated "omics" technologies, AI, and big data, it constructs effective models to predict plant traits and performance (phenotype). The objective of this "smart breeding" is to promote genetic gains, shorten breeding cycles, and develop plants that are acclimatized to specific and evolving environments (Xu et al., 2022c). However, the accuracy in phenotype prediction is intrinsically tied to the processing of multidimensional or multimodal data, sourced from spatiotemporal omics, encompassing genotype, environment, and their interaction. Such a data-rich environment poses significant challenges to conventional machine learning methodologies (Xu et al., 2022c).

In response to this complexity, MFMs (Section 2.1.3) may emerge as a potential solution. These models, designed to handle and learn from varied data types across multiple domains, could significantly improve the accuracy and efficiency of predicting plant phenotypes, thereby pushing the field of smart plant breeding toward unprecedented possibilities.

3.3.3. Smart livestock farming

Smart livestock farming is an innovative, automated approach that fuses cutting-edge technologies such as the Internet of Things (IoT), deep learning, and technology for intelligent control of livestock production (Islam et al., 2019; Astill et al., 2020; Zhang et al., 2021a; Farooq et al., 2022). This includes the automation of critical processes such as food and water supply, egg collection, livestock activity monitoring, and maintenance of precise environmental conditions. The primary objective of this innovative methodology is to enhance operational efficiency in livestock farms, leading to considerable savings in labor, maintenance, and material costs (Choukidar and Dawande, 2017). Smart livestock farming, while innovative and promising, also presents challenges. The development and implementation of such an intelligent system require a multidisciplinary understanding that spans the realms of advanced technologies, animal science, environmental science, and data analytics, among others (Islam et al., 2019). This integrative approach involves handling and processing voluminous and varied data across multiple domains. Overcoming these challenges

would require strategic collaborations, expertise across domains, and harnessing cutting-edge tools in AI and data analytics.

Given these complex challenges, FMs emerge as a potential solution to facilitate the realization of smart livestock farming (Yang et al., 2023a). These comprehensive models, built on large-scale multimodal data and incorporated knowledge from various domains, can effectively manage the complications of multi-dimensional information and bridge the gaps between different disciplines (Bommasani et al., 2021). FMs, with their ability to learn and generalize from large datasets, could provide valuable insights, enhance decision-making, and enable more precise management, thereby contributing significantly to the advancement of intelligent and efficient livestock farming systems.

3.3.4. Precision aquaculture farming

Precision aquaculture farming (O'Donncha and Grant, 2019; Føre et al., 2018; Antonucci and Costa, 2020), also known as smart aquaculture or precision fish farming, is a modern approach to fish and shellfish farming that integrates advanced technologies like IoT and data-driven techniques to optimize production efficiency, environmental sustainability, and animal welfare. It is often characterized by hundreds of interconnected sensors that store and serve data to gather essential information such as temperature, oxygen, weather, chlorophyll, fish tags, imagery data, and 3D sensor data, which can be utilized for monitoring fish variables (species, health, etc.), behavioral analysis, feeding decisions, and water quality prediction (Yang et al., 2021). Recently, DL has emerged as a promising solution to address the problems in precision aquaculture farming. For example, in Måløy et al. (2019), a dual-stream recurrent network was adopted to classify fish feeding status for salmon, yielding 80% feeding classification accuracy. However, these DL approaches rely heavily on large-scale and labeled datasets for training (Yang et al., 2021). Underwater conditions can be adverse, with poor illumination and low visibility in turbid water, as well as cluttered backgrounds, making it difficult to acquire high-quality data. Besides, it is very challenging to integrate, handle, and process numerous and varied data across multiple domains.

Fortunately, FMs have become an encouraging solution to facilitate the realization of precision aquaculture farming, representing the unique capability to handle diverse data from multiple domains through few-short or zero-short learning. For example, generation-based FMs can be harnessed for data creation, while RLFMs can be utilized for decision-making in fish farming. The application of FMs could significantly enhance the accuracy and efficiency of associated applications in precision aquaculture farming.

3.3.5. Agricultural robots

Agricultural robots, often referred to as "AgriRobots", are at the forefront of transforming modern agriculture (Bechar and Vigneault, 2016). AgriRobots can automate various farming tasks such as planting, weeding, harvesting, and crop health monitoring (Fountas et al., 2020). For instance, AI-guided robots can identify and remove weeds with precision, significantly reducing the use of herbicides. In smart orchards, AI-guided robots equipped with a variety of sensors, including image sensors, LiDAR, and proximity sensors, autonomously pick ripe fruits, thereby optimizing harvesting efficiency and minimizing labor costs. Similarly, drones equipped with advanced imaging sensors can monitor fields for signs of disease or pests, enabling early intervention and reducing yield losses. Similar to the contexts of smart breeding and smart livestock farming, the development and implementation of AgriRobots require multidisciplinary knowledge including robotics, AI, plant science, environmental science, and data analytics. This mixed approach involves handling and processing a significant amount of data from various domains, making it a complex endeavor (Bechar and Vigneault, 2016; Fountas et al., 2020).

The above challenges can be guided effectively by the potential application of FMs. These models, trained on large-scale multimodal data, can provide the robust predictive capabilities, which are necessary for advanced farming operations. For instance, RLFM (Section 2.1.4) could enable precision control of AgriRobots (Reed et al., 2022; Team et al., 2023b), while LFM (Section 2.1.1) could provide valuable insights during the design process of AgriRobots (Stella et al., 2023). By incorporating insights from diverse domains, FMs can offer an integrated solution, driving the development of more effective and efficient agricultural robots, and thus, revolutionizing the future of agriculture.

However, constructing an AFM from scratch presents significant challenges. Firstly, initial data collection requires a large and diverse dataset, which requires laborious efforts and potentially high costs (Lu and Young, 2020). The subsequent data labeling and curation require meticulousness and expertise. Additionally, the process of training models demands substantial computational resources, thus presenting impediments for individuals or entities with constrained resources. Lastly, fine-tuning the model for specific agricultural applications can be challenging, requiring a balance between generalizability and specificity. A detailed discussion of these challenges will be presented in Section 4.

4. Challenges and outlook

In this section, both the strengths and weaknesses of FMs are first summarized. Then, open challenges with training, deployment, and distribution shifts are identified and discussed.

4.1. Pros and cons of FMs

FMs represent a significant paradigm shift in the field of AI. In contrast to traditional AI models that are trained on domain-specific data for specialized tasks, FMs are pre-trained on a vast scale of data, spanning multiple domains and modalities (Bommasani et al., 2021; Moor et al., 2023; Mai et al., 2023). The training process allows these models to learn a broad base of knowledge and skills, much like the foundation of a building that supports all the structures above it. Once trained, these FMs can accomplish versatile tasks with just a little fine-tuning. This is analogous to the way humans transfer knowledge and skills from one area to another (Moor et al., 2023). They have led to breakthroughs in a variety of fields, from natural language processing to computer vision and beyond. The pros of FMs are summarized as follows:

1. **Pre-trained Knowledge**: By training on vast and diverse datasets, FMs possess a form of "general intelligence" that encompasses knowledge of the world, language, vision, and their specific training domains.
2. **Fine-tuning Flexibility**: FMs demonstrate superior performance to be fine-tuned for particular tasks or datasets, saving the computational and temporal investments required to train extensive models from scratch.
3. **Data Efficiency**: FMs harness their foundational knowledge, exhibiting remarkable performance even in the face of limited task-specific data, which is effective for scenarios with data scarcity issues.

Given the distinct advantages of FMs over traditional AI models, they exhibit substantial promise for agricultural applications. Some compelling examples of applying LLMs and FMs in the agricultural domain can be found in Section 3. However, despite their progress, FMs are not without their unique set of challenges:

1. **Computational Cost**: The foundational training of these models demands considerable computational resources, often beyond the reach of individuals or firms with constrained budgets.
2. **Generalization v.s. Specialization**: Although FMs offer commendable generalization capabilities across diverse tasks, there exist scenarios where task-specific models deliver superior performance.

3. **Deployment Difficulties**: The considerable size of many FMs poses challenges for deployment on edge devices or other systems with restricted computational and storage capabilities.
4. **Transparency and Interpretability**: The complexity and scale of FMs can hinder clear understanding and interpretation, making them opaque in certain applications.
5. **Ethical Concerns**: Given their expansive training datasets, there is potential for these models to inadvertently perpetuate and magnify existing societal biases.

4.2. Data collection

The process of data collection for training FMs in agricultural applications presents unique challenges due to the inherent variability and diversity in agriculture, influenced by factors such as crop types, growth stages, soil conditions, weather patterns, farming practices, etc. These factors result in a complex and highly diverse dataset, making its collection and standardization a challenging task (Lu and Young, 2020; Xu et al., 2023b). Firstly, it is critical to ensure the collected data is accurate, comprehensive, and adequately captures the wide variability found in different farming environments or diverse range of growth phases. However, acquiring high-quality data can be time-consuming, labor-intensive, and expensive, especially considering the need for ground truth labeling in supervised learning scenarios (Li et al., 2023a) and the variety of growth stages of plants or livestock. To tackle these challenges, generative models (Lu et al., 2022) that create high-fidelity images and label-efficient learning methods to save on tedious labeling costs (Li et al., 2023a; Zhang et al., 2023a) show promises. In addition, overcoming the time-consuming and labor-intensive data collection in the field can be achieved by leveraging advancements in autonomous data collection technologies, such as UAVs or UGVs equipped with sensors, as well as remote sensing techniques and the IoT devices (Mueller-Sim et al., 2017; del Cerro et al., 2021; Xu et al., 2022a). These technologies automate data-gathering processes and reduce the need for manual intervention. Secondly, data privacy and ownership pose challenges as farms are typically privately owned. Farmers may be reluctant to share data due to concerns over privacy or potential commercial exploitation. Novel frameworks, such as federated learning (Zhang et al., 2021b), can address these concerns, paving the way for broader access to high-quality, diverse agricultural datasets. Temporal dynamics represent another significant hurdle. Agricultural processes inherently change over time, influenced by daily fluctuations, seasonal variations, or annual changes. This requires the collection of time-series data (*e.g.*, data across years), introducing another layer of complexity. To handle temporal dynamics, techniques such as time-series analysis (Ostrom, 1990), dynamic models based on Long Short-Term Memory (LSTM) (Graves and Graves, 2012), and data augmentation (Lu et al., 2022) can be employed to better account for the inherent temporal dynamics in agricultural data.

In conclusion, while data collection for training FMs in agriculture is a challenging endeavor, addressing these issues is crucial for developing robust and effective models capable of advancing the state of agricultural technology and practice.

4.3. Training efficiency

Training FMs for agriculture applications presents significant challenges, particularly with respect to long training times and substantial costs, *e.g.*, requiring thousands of GPU hours and millions of training data. However, innovative solutions are being devised to overcome these challenges. One such approach is the integration of an existing LLM with a relatively lightweight visual prompt generator (VPG) (Li et al., 2023b). Though this approach can reduce the computational costs, the tuning of the VPG component within the VL-LLM still represents a significant computational demand Zhang et al. (2023b). A promising solution discussed in the literature involves the transfer of

an existing VPG from any existing FMs to the target FMs. This allows for the leveraging of pre-existing model structures and parameters, thereby bypassing the need for extensive training from scratch. The key to optimizing this process lies in maximizing transfer efficiency across different LLM sizes and types. A significant development in this regard is the VPGTrans approach (Zhang et al., 2023b), a two-stage transfer framework designed to speed up the transfer learning process, which demonstrated that their method can facilitate VPG transfer with a speed-up of more than 10× and uses only a fraction of the training data compared to building a VPG from scratch.

In summary, while training FMs for agricultural applications indeed presents substantial challenges due to long training times and high costs, advancements like the VPG transfer technique offer promising solutions. These techniques pave the way for more efficient and cost-effective model training processes, thereby making the use of FMs more accessible and practical in the field of agriculture.

4.4. Distribution shift

The issue of distribution shift presents a significant challenge when applying FMs to agricultural applications. Distribution shift occurs when the data that the model encounters during deployment differs significantly from the data used during its training phase (Wiles et al., 2022). In the context of agriculture, the environmental conditions under which data is collected can vary greatly across different regions and seasons. These variations can include differences in crop types, soil conditions, weather patterns, and farming practices, all of which can cause considerable shifts in data distribution (Wiles et al., 2022). For example, Williams et al. (2023) demonstrated that the direct application of FMs in a zero-shot manner for leaf segmentation tasks resulted in unsatisfactory performance, likely due to potential distribution shifts (Chawla et al., 2021).

To combat the challenges posed by distribution shifts when training and deploying FMs, researchers have been continuously investigating these issues (Chawla et al., 2021; Cvejoski et al., 2022; Bommasani et al., 2021). For instance, Cvejoski et al. (2022) demonstrated that LLMs may experience performance degradation of up to approximately 88% when predicting the popularity of future posts from subreddits whose topic distribution evolves over time. To mitigate these distribution shift issues, they introduced neural variational dynamic topic models and attention mechanisms. These techniques infer temporal language model representations for regression tasks, resulting in only about 40% performance drops in the worst-case scenarios and a mere 2% in the best cases. Furthermore, Wang et al. (2023) demonstrated that the integration of techniques such as multi-task learning, continual learning, and distillation with VFMs like SAM and CLIP can significantly alleviate distribution shift issues. These techniques have been shown to significantly mitigate performance degradation caused by distribution shifts, thus demonstrating promising prospects for the application of FMs in the diverse and dynamic field of agriculture.

4.5. Other challenges in real-world deployment

Deploying FMs for agricultural applications presents a unique set of challenges, particularly regarding inference speed and model size. Firstly, the inherent complexity and size of these models can pose substantial computational demands. Large-scale FMs often require significant processing power and memory for both training and inference, which may not always be readily available in on-field agricultural settings (Bommasani et al., 2021). This can limit the deployment of these models in real-time applications, such as monitoring crops or automating irrigation systems. Secondly, inference speed is crucial for many agricultural tasks that require immediate action based on the model's outputs (Bommasani et al., 2021). For instance, in precision agriculture, time-sensitive tasks like pest detection or predicting weather impacts on crop health demand quick and accurate responses.

However, due to the size and complexity of FMs, achieving high-speed inference can be a challenge.

To overcome these hurdles, future research and development efforts need to focus on model optimization techniques (Zhong et al., 2023; Zhu et al., 2023), such as model compression (*e.g.*, pruning and quantization Cheng et al., 2017; Choudhary et al., 2020) and efficient network design (Wan, 2022), which could potentially reduce the model size without compromising performance. Actually, there are already some attempts to deploy FMs. For instance, FQ-ViT (Lin et al., 2022) introduces a fully quantized vision transformer that simplifies inference by utilizing only 1/4 of the memory with minimal accuracy degradation (approximately 1%). Additionally, SqueezeLLM (Kim et al., 2023) demonstrates that utilizing only 3/16 of the parameters significantly reduces the perplexity gap by up to 2.1 times and achieves up to 2.3 times speedup compared to baseline models.

Additionally, the utilization of edge computing solutions could expedite the inference process by enabling data processing closer to its source, consequently enhancing the speed and efficiency of FMs in agricultural applications (Chen and Ran, 2019). For example, in agricultural applications, edge devices like drones or field sensors can analyze data locally, enabling real-time decision-making without relying on distant servers. Recently, MLC LLM (Girshick et al., 2023) offers a universal solution, allowing any language model to be deployed directly across a diverse range of hardware backends and native applications. This solution also provides a productive framework for individuals to further optimize model performance for their specific use cases. On the other hand, MobileSAM (Zhang et al., 2023c) and MobileSAMv2 (Zhang et al., 2023d) make SAM (Kirillov et al., 2023) mobile-friendly by replacing the heavyweight image encoder with a lightweight one. This lightweight SAM is more than 60 times smaller yet performs on par with the original SAM (Kirillov et al., 2023).

4.6. Building capacity for FM adoption in agriculture

Effective utilization of FMs in agriculture requires specialized skills and multi-discipline collaborations (Mai et al., 2023). However, there are several challenges in the skill transfer of applying FMs in agricultural domains, including domain adaptation techniques due to agricultural researchers' potential lack of familiarity with FM applications, understanding unique characteristics of agricultural data such as high variability, and interpreting the complex nature of FMs.

Addressing the skill gap and providing targeted training is crucial for empowering agricultural researchers and users to leverage FMs optimally. Investing in training and capacity-building initiatives, such as workshops, webinars, specialized training programs focusing on FM utilization in agriculture, and educational resources such as online tutorials and case studies, is essential for enabling effective FM utilization in agriculture, driving innovation, and addressing agricultural challenges.

5. Conclusion

In conclusion, this study has investigated the potential of FMs in the field of smart agriculture. These models hold the promise of transforming the agricultural sector by reducing reliance on extensive labeled datasets and enhancing generalization and effectiveness. By categorizing and reviewing various types of LLMs and FMs (Section 2), their applicability in smart crop management, smart plant breeding, smart livestock farming, precision aquaculture farming, and agricultural robotics are revealed (Section 3). However, it is crucial to address challenges and risks such as data collection, model training, distribution shift, real-world deployment, and research gaps before real-world implementation (Section 4). Further exploration and utilization of LLMs and FMs have the potential to enhance productivity, sustainability, and decision-making in agriculture.

In the era of LLMs and FMs, all agricultural fields successfully explored using traditional ML/DL approaches, as discussed in Section 3

and beyond, are ready for exploration by employing FMs. Integrating LLMs and FMs into the existing framework is also a promising direction to improve efficiency and effectiveness, as evident in the studies by Stella et al. (2023), Tzachor et al. (2023). These integrations can leverage the strengths of both frameworks, potentially leading to more robust and adaptable agricultural systems. Furthermore, the development of large-scale agricultural datasets would be an approaching direction to accelerate the development and application of LLMs and FMs in the agriculture field (Lu and Young, 2020). Such datasets can facilitate more comprehensive training of models, enabling them to capture the nuances and complexities of agricultural processes more accurately. Additionally, there is a growing need for interdisciplinary collaboration between experts in agriculture, computer science, and data science. By fostering collaboration across these domains, innovative solutions can be developed to address the unique challenges and opportunities in agriculture using LLMs and FMs.

Ultimately, this study aims to inspire and guide researchers to unlock the untapped potential of LLMs and FMs in smart agriculture and serves as a stepping stone toward developing smart agriculture with next-level intelligence and more sophisticated capabilities.

CRediT authorship contribution statement

Jiajia Li: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Mingle Xu:** Writing – review & editing, Formal analysis. **Lirong Xiang:** Writing – review & editing, Formal analysis. **Dong Chen:** Writing – original draft, Investigation, Conceptualization. **Weichao Zhuang:** Writing – review & editing. **Xunyuan Yin:** Writing – review & editing. **Zhaojian Li:** Writing – review & editing, Supervision, Resources.

Declaration of competing interest

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

Data availability

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

References

Abbas, A., Jain, S., Gour, M., Vankudothu, S., 2021. Tomato plant disease detection using transfer learning with C-GAN synthetic images. Comput. Electron. Agric. 187, 106279.

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al., 2023. Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., Ring, R., Rutherford, E., Cabi, S., Han, T., Gong, Z., Samangooei, S., Monteiro, M., Menick, J., Borgeaud, S., Brock, A., Nematzadeh, A., Sharifzadeh, S., Binkowski, M., Barreira, R., Vinyals, O., Zisserman, A., Simonyan, K., 2022. Flamingo: A visual language model for few-shot learning. arXiv abs/2204.14198.

Anthropic, 2024. The claude 3 model family: Opus, Sonnet, Haiku. URL https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf.

Antonucci, F., Costa, C., 2020. Precision aquaculture: A short review on engineering innovations. Aquacult. Int. 28 (1), 41–57.

Arulkumaran, K., Deisenroth, M.P., Brundage, M., Bharath, A.A., 2017. Deep reinforcement learning: A brief survey. IEEE Signal Process. Mag. 34 (6), 26–38.

Astill, J., Dara, R.A., Fraser, E.D., Roberts, B., Sharif, S., 2020. Smart poultry management: Smart sensors, big data, and the Internet of Things. Comput. Electron. Agric. 170, 105291.

Barham, P., Chowdhery, A., Dean, J., Ghemawat, S., Hand, S., Hurt, D., Isard, M., Lim, H., Pang, R., Roy, S., et al., 2022. Pathways: Asynchronous distributed dataflow for ML. In: Proceedings of Machine Learning and Systems, vol. 4, pp. 430–449.

Bechar, A., Vigneault, C., 2016. Agricultural robots for field operations: Concepts and components. Biosyst. Eng. 149, 94–111.

Bender, A., Whelan, B., Sukkarieh, S., 2020. A high-resolution, multimodal data set for agricultural robotics: A Ladybird's-eye view of Brassica. J. Field Robotics 37 (1), 73–96.

Binas, J., Luginbuehl, L., Bengio, Y., 2019. Reinforcement learning for sustainable agriculture. In: ICML 2019 Workshop Climate Change: How Can AI Help.

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., et al., 2021. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Bowman, S.R., Angeli, G., Potts, C., Manning, C.D., 2015. A large annotated corpus for learning natural language inference. arXiv preprint arXiv:1508.05326.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. In: Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901.

Bu, F., Wang, X., 2019. A smart agriculture IoT system based on deep reinforcement learning. Future Gener. Comput. Syst. 99, 500–507.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S., et al., 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712.

Cao, Y., Chen, L., Yuan, Y., Sun, G., 2023a. Cucumber disease recognition with small samples using image-text-label-based multi-modal language model. Comput. Electron. Agric. 211, 107993.

Cao, Y., Xu, X., Sun, C., Cheng, Y., Du, Z., Gao, L., Shen, W., 2023b. Segment any anomaly without training via hybrid prompt regularization. arXiv preprint arXiv:2305.10724.

Chawla, S., Singh, N., Drori, I., 2021. Quantifying and alleviating distribution shifts in foundation models on review classification. In: NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications. URL https://openreview.net/forum?id=OG78-TuPcvL.

Chen, D., Hajidavalloo, M.R., Li, Z., Chen, K., Wang, Y., Jiang, L., Wang, Y., 2023. Deep multi-agent reinforcement learning for highway on-ramp merging in mixed traffic. IEEE Trans. Intell. Transp. Syst..

Chen, D., Lu, Y., Li, Z., Young, S., 2022. Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. Comput. Electron. Agric. 198, 107091.

Chen, D., Qi, X., Zheng, Y., Lu, Y., Huang, Y., Li, Z., 2024. Synthetic data augmentation by diffusion probabilistic models to enhance weed recognition. Comput. Electron. Agric. 216, 108517.

Chen, J., Ran, X., 2019. Deep learning with edge computing: A review. Proc. IEEE 107 (8), 1655–1674.

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

Cheng, Y., Wang, D., Zhou, P., Zhang, T., 2017. A survey of model compression and acceleration for deep neural networks. arXiv preprint arXiv:1710.09282.

Choudhary, T., Mishra, V., Goswami, A., Sarangapani, J., 2020. A comprehensive survey on model compression and acceleration. Artif. Intell. Rev. 53, 5113–5155.

Choukidar, G.A., Dawande, N., 2017. Smart poultry farm automation and monitoring system. In: 2017 International Conference on Computing, Communication, Control and Automation. ICCUBEA, IEEE, pp. 1–5.

Christiano, P.F., Leike, J., Brown, T., Martic, M., Legg, S., Amodei, D., 2017. Deep reinforcement learning from human preferences. In: Advances in Neural Information Processing Systems, vol. 30.

Chu, P., Li, Z., Lammers, K., Lu, R., Liu, X., 2021. Deep learning-based apple detection using a suppression mask R-CNN. Pattern Recognit. Lett. 147, 206–211.

Chu, P., Li, Z., Zhang, K., Chen, D., Lammers, K., Lu, R., 2023. O2RNet: Occluder-occludee relational network for robust apple detection in clustered orchard environments. Smart Agric. Technol. 100284. http://dx.doi.org/10.1016/j.atech.2023.100284, URL https://www.sciencedirect.com/science/article/pii/S2772375523001132.

Chu, T., Wang, J., Codecà, L., Li, Z., 2019. Multi-agent deep reinforcement learning for large-scale traffic signal control. IEEE Trans. Intell. Transp. Syst. 21 (3), 1086–1095.

Crowson, K., Biderman, S., Kornis, D., Stander, D., Hallahan, E., Castricato, L., Raff, E., 2022. Vqgan-clip: Open domain image generation and editing with natural language guidance. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVII. Springer, pp. 88–105.

Cvejoski, K., Sánchez, R.J., Ojeda, C., 2022. The future is different: Large pre-trained language models fail in prediction tasks. arXiv preprint arXiv:2211.00384.

Dang, F., Chen, D., Lu, Y., Li, Z., 2023. YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. Comput. Electron. Agric. 205, 107655.

Dang, F., Chen, D., Lu, Y., Li, Z., Zheng, Y., 2022. DeepCottonWeeds (DCW): A novel benchmark of YOLO object detectors for weed detection in cotton production systems. In: 2022 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, p. 1.

del Cerro, J., Cruz Ulloa, C., Barrientos, A., de León Rivas, J., 2021. Unmanned aerial vehicles in agriculture: A survey. Agronomy 11 (2), 203.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dhakshayani, J., Surendiran, B., 2023. M2F-Net: A deep learning-based multimodal classification with high-throughput phenotyping for identification of overabundance of fertilizers. Agriculture 13 (6), 1238.

Din, A., Ismail, M.Y., Shah, B., Babar, M., Ali, F., Baig, S.U., 2022. A deep reinforcement learning-based multi-agent area coverage control for smart agriculture. Comput. Electr. Eng. 101, 108089.

Du, N., Huang, Y., Dai, A.M., Tong, S., Lepikhin, D., Xu, Y., Krikun, M., Zhou, Y., Yu, A.W., Firat, O., et al., 2022. Glam: Efficient scaling of language models with mixture-of-experts. In: International Conference on Machine Learning. PMLR, pp. 5547–5569.

Eli-Chukwu, N.C., 2019. Applications of artificial intelligence in agriculture: A review. Eng. Technol. Appl. Sci. Res. 9 (4).

Espejo-Garcia, B., Mylonas, N., Athanasakos, L., Fountas, S., Vasilakoglou, I., 2020. Towards weeds identification assistance through transfer learning. Comput. Electron. Agric. 171, 105306.

Farooq, M.S., Sohail, O.O., Abid, A., Rasheed, S., 2022. A survey on the role of iot in agriculture for the implementation of smart livestock environment. IEEE Access 10, 9483–9505.

Føre, M., Frank, K., Norton, T., Svendsen, E., Alfredsen, J.A., Dempster, T., Eguiraun, H., Watson, W., Stahl, A., Sunde, L.M., et al., 2018. Precision fish farming: A new framework to improve production in aquaculture. Biosyst. Eng. 173, 176–193.

Fountas, S., Mylonas, N., Malounas, I., Rodias, E., Hellmann Santos, C., Pekkeriet, E., 2020. Agricultural robotics for field operations. Sensors 20 (9), 2672.

Gandhi, R., 2022. Deep reinforcement learning for agriculture: Principles and use cases. In: Data Science in Agriculture and Natural Resource Management. Springer, pp. 75–94.

Garg, S., Pundir, P., Jindal, H., Saini, H., Garg, S., 2021. Towards a multimodal system for precision agriculture using IoT and machine learning. In: 2021 12th International Conference on Computing Communication and Networking Technologies. ICCCNT, IEEE, pp. 1–7.

Gautron, R., Maillard, O.-A., Preux, P., Corbeels, M., Sabbadin, R., 2022. Reinforcement learning for crop management support: Review, prospects and challenges. Comput. Electron. Agric. 200, 107182.

Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J., 2021. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

Ghazi, M.M., Yanikoglu, B., Aptoula, E., 2017. Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing 235, 228–235.

Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., He, K., 2018. Detectron. https://github.com/facebookresearch/detectron.

Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., He, K., 2023. MLC LLM. https://github.com/mlc-ai/mlc-llm.

Goëau, H., Bonnet, P., Joly, A., 2022. Overview of PlantCLEF 2022: Image-based plant identification at global scale. In: CLEF 2022-Conference and Labs of the Evaluation Forum, vol. 3180, (no. 153), pp. 1916–1928.

Gondchawar, N., Kawitkar, R., et al., 2016. IoT based smart agriculture. Int. J. Adv. Res. Comput. Commun. Eng. 5 (6), 838–842.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Commun. ACM 63 (11), 139–144.

Graves, A., Graves, A., 2012. Long short-term memory. In: Supervised Sequence Labelling with Recurrent Neural Networks. Springer, pp. 37–45.

Gu, S., Holly, E., Lillicrap, T., Levine, S., 2017. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In: 2017 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 3389–3396.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al., 2018. Recent advances in convolutional neural networks. Pattern Recognit. 77, 354–377.

Hasan, A.M., Sohel, F., Diepeveen, D., Laga, H., Jones, M.G., 2021. A survey of deep learning techniques for weed detection from images. Comput. Electron. Agric. 184, 106067.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., Meger, D., 2018. Deep reinforcement learning that matters. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, (no. 1).

Hu, K., Wang, Z., Coleman, G., Bender, A., Yao, T., Zeng, S., Song, D., Schumann, A., Walsh, M., 2021. Deep learning techniques for in-crop weed identification: A review. arXiv preprint arXiv:2103.14872.

Huang, S., Dong, L., Wang, W., Hao, Y., Singhal, S., Ma, S., Lv, T., Cui, L., Mohammed, O.K., Liu, Q., et al., 2023. Language is not all you need: Aligning perception with language models. arXiv preprint arXiv:2302.14045.

Ilharco, G., Wortsman, M., Carlini, N., Taori, R., Dave, A., Shankar, V., Namkoong, H., Miller, J., Hajishirzi, H., Farhadi, A., Schmidt, L., 2021. Open clip.

Islam, M.M., Tonmoy, S.S., Quayum, S., Sarker, A.R., Hani, S.U., Mannan, M.A., 2019. Smart poultry farm incorporating GSM and IoT. In: 2019 International Conference on Robotics, Electrical and Signal Processing Techniques. ICREST, IEEE, pp. 277–280.

Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F., 2020. A survey on contrastive self-supervised learning. Technologies 9 (1), 2.

Jaramillo, S., Graterol, E., Pulver, E., 2020. Sustainable transformation of rainfed to irrigated agriculture through water harvesting and smart crop management practices. Front. Sustain. Food Syst. 4, 437086.

Jenkins, J.R., Fuchs, L.S., Van Den Broek, P., Espin, C., Deno, S.L., 2003. Sources of individual differences in reading comprehension and reading fluency. J. Educ. Psychol. 95 (4), 719.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90.

Kang, M., Zhu, J.-Y., Zhang, R., Park, J., Shechtman, E., Paris, S., Park, T., 2023. Scaling up gans for text-to-image synthesis. arXiv preprint arXiv:2303.05511.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., Amodei, D., 2020. Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T., 2020. Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8110–8119.

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., Pirotti, F., 2019. Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. Remote Sens. 11 (23), 2873.

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S., 2018. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. Comput. Electron. Agric. 153, 213–225.

Kim, S., Hooper, C., Gholami, A., Dong, Z., Li, X., Shen, S., Mahoney, M., Keutzer, K., 2023. SqueezeLLM: Dense-and-sparse quantization. arXiv.

Kirillov, A., He, K., Girshick, R., Rother, C., Dollár, P., 2019. Panoptic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9404–9413.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al., 2023. Segment anything. arXiv preprint arXiv:2304.02643.

Kober, J., Bagnell, J.A., Peters, J., 2013. Reinforcement learning in robotics: A survey. Int. J. Robot. Res. 32 (11), 1238–1274.

Kuska, M.T., Wahabzada, M., Paulus, S., 2024. AI-chatbots for agriculture-where can large language models provide substantial value? Available at SSRN 4685971.

Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., Epstein, D., Polosukhin, I., Devlin, J., Lee, K., et al., 2019. Natural questions: A benchmark for question answering research. Trans. Assoc. Comput. Linguist. 7, 453–466.

Lakshmi, K., Gayathri, S., 2017. Implementation of IoT with image processing in plant growth monitoring system. J. Sci. Innov. Res. 6 (2), 80–83.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521 (7553), 436–444.

Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V., Hutter, M., 2020. Learning quadrupedal locomotion over challenging terrain. Sci. Robot. 5 (47), eabc5986.

Li, J., Chen, D., Qi, X., Li, Z., Huang, Y., Morris, D., Tan, X., 2023a. Label-efficient learning in agriculture: A comprehensive review. Comput. Electron. Agric. 215, 108412.

Li, J., Chen, D., Yin, X., Li, Z., 2024. Performance evaluation of semi-supervised learning frameworks for multi-class weed detection. arXiv:2403.03390.

Li, Z., Guo, R., Li, M., Chen, Y., Li, G., 2020. A review of computer vision technologies for plant phenotyping. Comput. Electron. Agric. 176, 105672.

Li, J., Li, D., Savarese, S., Hoi, S., 2023b. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. arXiv preprint arXiv:2301.12597.

Li, J., Li, D., Xiong, C., Hoi, S., 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In: International Conference on Machine Learning. PMLR, pp. 12888–12900.

Li, J., Magar, R.T., Chen, D., Lin, F., Wang, D., Yin, X., Zhuang, W., Li, Z., 2023c. SoybeanNet: Transformer-based convolutional neural network for soybean pod counting from Unmanned Aerial Vehicle (UAV) images. arXiv preprint arXiv:2310.10861.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer, pp. 740–755.

Lin, Y., Zhang, T., Sun, P., Li, Z., Zhou, S., 2022. Fq-vit: Post-training quantization for fully quantized vision transformer. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22. pp. 1173–1179.

Liu, Y., Wang, Y., Li, Y., Li, Q., Wang, J., 2022. SETR-YOLOv5n: A lightweight low-light lane curvature detection method based on fractional-order fusion model. IEEE Access 10, 93003–93016.

Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., Tang, J., 2021. Self-supervised learning: Generative or contrastive. IEEE Trans. Knowl. Data Eng. 35 (1), 857–876.

Lu, Y., Chen, D., Olaniyi, E., Huang, Y., 2022. Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review. Comput. Electron. Agric. 200, 107208.

Lu, G., Li, S., Mai, G., Sun, J., Zhu, D., Chai, L., Sun, H., Wang, X., Dai, H., Liu, N., et al., 2023. AGI for agriculture. arXiv preprint arXiv:2304.06136.

Lu, Y., Young, S., 2020. A survey of public datasets for computer vision tasks in precision agriculture. Comput. Electron. Agric. 178, 105760.

Ma, J., Wang, B., 2023. Segment anything in medical images. arXiv preprint arXiv:2304.12306.

Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., et al., 2023. On the opportunities and challenges of foundation models for geospatial artificial intelligence. arXiv preprint arXiv:2304.06798.

Måløy, H., Aamodt, A., Misimi, E., 2019. A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture. Comput. Electron. Agric. 167, 105087.

Mendigoria, C.H., Aquino, H., Concepcion, R., Alajas, O.J., Dadios, E., Sybingco, E., 2021. Vision-based postharvest analysis of musa acuminata using feature-based machine learning and deep transfer networks. In: 2021 IEEE 9th Region 10 Humanitarian Technology Conference. R10-HTC, IEEE, pp. 01–06.

Meshram, V., Patil, K., Meshram, V., Hanchate, D., Ramkteke, S., 2021. Machine learning in agriculture domain: A state-of-art survey. Artif. Intell. Life Sci. 1, 100010.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M., 2013. Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al., 2015. Human-level control through deep reinforcement learning. nature 518 (7540), 529–533.

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, 1419.

Moor, M., Banerjee, O., Abad, Z.S.H., Krumholz, H.M., Leskovec, J., Topol, E.J., Rajpurkar, P., 2023. Foundation models for generalist medical artificial intelligence. Nature 616 (7956), 259–265.

Moysiadis, V., Sarigiannidis, P., Vitsas, V., Khelifi, A., 2021. Smart farming in Europe. Comput. Sci. Rev. 39, 100345.

Mueller-Sim, T., Jenkins, M., Abel, J., Kantor, G., 2017. The robotanist: A ground-based agricultural robot for high-throughput crop phenotyping. In: 2017 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 3634–3639.

Nadkarni, P.M., Ohno-Machado, L., Chapman, W.W., 2011. Natural language processing: An introduction. J. Am. Med. Inform. Assoc. 18 (5), 544–551.

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., Chen, M., 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv preprint arXiv:2112.10741.

O'Donncha, F., Grant, J., 2019. Precision aquaculture. IEEE Internet Things Mag. 2 (4), 26–30.

OpenAI, 2022. Introducing ChatGPT. https://openai.com/blog/chatgpt.

OpenAI, 2023. GPT-4 technical report. arXiv:2303.08774.

Ostrom, C.W., 1990. Time Series Analysis: Regression Techniques, (no. 9), Sage.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al., 2022. Training language models to follow instructions with human feedback. Adv. Neural Inf. Process. Syst. 35, 27730–27744.

Parr, C.S., Lemay, D.G., Owen, C.L., Woodward-Greene, M.J., Sun, J., 2021. Multimodal AI to improve agriculture. IT Prof. 23 (3), 53–57.

Pathak, H., Igathinathane, C., Zhang, Z., Archer, D., Hendrickson, J., 2022. A review of unmanned aerial vehicle-based methods for plant stand count evaluation in row crops. Comput. Electron. Agric. 198, 107064.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning. PMLR, pp. 8748–8763.

Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al., 2019. Language models are unsupervised multitask learners. OpenAI blog 1 (8), 9.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. 21 (1), 5485–5551.

Rai, N., Zhang, Y., Ram, B.G., Schumacher, L., Yellavajjala, R.K., Bajwa, S., Sun, X., 2023. Applications of deep learning in precision weed management: A review. Comput. Electron. Agric. 206, 107698.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M., 2022. Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125.

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S.G., Novikov, A., Barth-Maron, G., Gimenez, M., Sulsky, Y., Kay, J., Springenberg, J.T., et al., 2022. A generalist agent. arXiv preprint arXiv:2205.06175.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695.