

AgriAgent: 面向农业环境控制的端到端大模型智能体系统架构

裘佳颖¹, 刘应昌¹, 高星杰¹, 黄 远², 张红雨¹, 田 芳^{1,3},
李万理^{1*}, 冯在文^{1,3,4*}

(1. 华中农业大学信息学院, 湖北武汉 430070, 中国; 2. 华中农业大学园艺林学学院, 湖北武汉 430070, 中国; 3. 云南现代农业产业研究院有限公司, 云南昆明 650032, 中国; 4. 农业智能技术教育部工程研究中心, 湖北武汉 430070, 中国)

摘要: [目的/意义] 为突破大语言模型在农业生产决策领域应用中存在的物理世界交互能力不足、专业工具调用受限, 以及“幻觉”生成等问题, 构建了一个工具增强型环控智能体系统。[方法] 首先, 基于开放气象数据与作物生长模型构建农业数字孪生评测平台, 提供标准化闭环控制验证环境。其次, 提出了 AgriAgent 架构, 通过“传感器-记忆库-检索器-大语言模型-工具执行器”闭环流程, 实现了田间环境的动态感知、知识检索、推理决策与精准执行的端到端闭环控制。该端到端架构的优势在于实现了“环境感知-动态决策-执行反馈”的全流程无缝衔接, 极大地提升了系统在复杂、动态农业环境下的实时响应能力和决策精度, 解决了传统应用缺乏“感知-决策-执行”闭环能力的问题。为验证系统有效性, 在数字孪生环境中, 对玉米、小米、甜菜、番茄、卷心菜5种作物在常规、扰动、极端3种情景下开展实验。[结果和讨论] 以 Qwen2.5_7B 为核心的 AgriAgent-7B 模型表现最优, 常规情景、扰动情景和极端情景下玉米产量较基线模型分别提升 562.20%、459.70%、463.60%; 小米产量分别提升 646.80%、558.70%、351.20%; 甜菜产量分别提升 228.20%、118.20%、125.40%; 番茄产量分别提升 926.20%、979.55%、1 537.46%; 卷心菜产量分别提升 501.67%、504.65%、1 185.14%。此外, 智能体参数量显著影响决策的质量, 当参数量达到 7 B 时, 智能体展现了卓越的灵活性与适应性, 能够解决大部分复杂的农业生产决策问题。[结论] 本研究验证了工具增强型智能体在农业动态决策中的有效性, 提出的 AgriAgent 端到端架构, 通过知识检索与工具调用闭环, 在农业场景下实现了“环境感知-动态决策-执行反馈”的全周期闭环控制, 为农业智能体研究提供了可靠技术路径和创新范式。

关键词: 工具增强型大语言模型; 农业智能体; 数字孪生; 农业评测平台; DSSAT/作物生长模型

中图分类号: S126; TP18

文献标识码: A

文章编号: SA202507042

引用格式: 裘佳颖, 刘应昌, 高星杰, 黄远, 张红雨, 田芳, 李万理, 冯在文. AgriAgent: 面向农业环境控制的端到端大模型智能体系统架构[J]. 智慧农业(中英文), 2026, 8(2): 220-236. DOI: 10.12133/j.smartag.SA202507042

QIU Jiaying, LIU Yingchang, GAO Xingjie, HUANG Yuan, ZHANG Hongyu, TIAN Fang, LI Wanli, FENG Zaiwen. AgriAgent: End-to-End Large Model Agent System Architecture for Agricultural Environment Control[J]. Smart Agriculture, 2026, 8(2): 220-236. DOI: 10.12133/j.smartag.SA202507042 (in Chinese with English abstract)

0 引言

大语言模型 (Large Language Models, LLMs)

的自然语言理解、知识整合及推理生成能力, 为农业智能化提供新路径。然而, LLMs 存在固有局限: 内容“幻觉”可能导致信息失真^[1,2]; 缺乏物理世界

收稿日期: 2025-07-30

基金项目: 湖北省技术创新计划项目 (2024BBB055); 湖北省科技厅援藏项目 (2024EIA009); 云南省重大科技专项计划 (202502AE090003); 中央高校基本科研业务费专项资金 (2662025XXPY005, 2662023XXQD002); 农业融合出版知识挖掘与知识服务重点实验室开放课题 (2025KMKS01)

作者简介: 1. 裘佳颖, 本科, 研究方向为自然语言处理。E-mail: 19560533416@163.com; 2. 刘应昌, 硕士研究生, 研究方向为自然语言处理。E-mail: liuyingchang@webmail.hzau.edu.cn。裘佳颖和刘应昌并列第一作者

*通信作者: 1. 李万理, 博士, 讲师, 研究方向为自然语言处理、知识图谱、元学习。E-mail: liwanli@mail.hzau.edu.cn; 2. 冯在文, 博士, 副教授, 研究方向为农业 AI 大模型。E-mail: zaiwen.Feng@mail.hzau.edu.cn

copyright©2026 by the authors

(登录 www.smartag.net.cn 免费获取电子版全文)

交互与实时数据获取能力;难以调用专业工具执行精准操作。这些问题制约了其在高准确性、高实时性场景中的应用。

为克服上述问题,工具增强型 LLMs 通过集成外部工具,使 LLMs 具备实时信息获取、精确计算与物理控制能力^[3]。在农业场景中,此类系统可连接物联网设备采集田间数据,结合知识库与生长模型生成灌溉施肥决策^[4],也能为偏远地区小农提供自然语言交互服务^[5]。不过当前研究仍面临 3 重挑战:闭源模型导致本地化落地困难;应用多局限于知识问答,尚未发挥核心决策价值^[6];缺乏农业标准化评测基准,这也阻碍了技术迭代。

近年来,工具增强型 LLM (Tool-Augmented LLMs) 在通用领域迅速发展。通用工具学习旨在赋予 LLMs 理解并使用各种外部工具的能力^[7-9]。但这类模型向农业领域延伸时,面临诸多挑战。首先,农业场景中的“工具”含硬件设备,如土壤传感器、气象站等,对这些硬件的操作需要更精确的参数控制和对实时状态的感知。其次,作物生长受多因子动态耦合影响,要求 LLMs 处理连续多模态数据流并实时决策。再次,农业决策直接影响作物的产量和品质,需要强化鲁棒性与风险评估能力。最后,评估要关注决策实际效益,往往需要结合长时间的模拟或真实的田间试验,评估周期长、成本高。现有通用框架依赖生成式预训练变换器 (Generative Pre-Trained Transformer) 模型进行需求构建和评估,与农业真实场景的适配性存在明显偏差。

与此同时,国内外研究者已开始探索 LLMs 在农业领域的具体应用,并取得了一些初步成果。这些探索多聚焦在信息服务层面,为农业生产提供信息服务和决策支持^[10]。例如,微软团队在其研究中^[11],通过检索增强生成 (Retrieval-Augmented Generation, RAG) 和微调 (Fine-Tuning) 技术,提升了农业回答准确性;农业双向编码器表示预训练语言模型 (Agricultural Bidirectional Encoder Representations from Transformers, AgriBERT)^[12] 等模型则专注于文本处理,在农业文本上进行了预训练,以期更好地理解农业领域的特定术语和语境。然而,这些模型均未扩展到与物理世界的交互,不涉及通过工具控制农业生产过程。

综上,现有应用缺乏“感知-决策-执行”闭环能力,难以满足农业实时精准管理需求。因此,突破物理世界交互壁垒,构建具备全周期决策与执行能力的农业智能体,成为当前领域发展的核心方向。

本研究聚焦三大核心目标:一是构建农业领域评测平台,提供标准化评估环境,以量化大语言模型在农业场景中的工具调用准确性、决策合理性及对产量的影响,从而填补该领域的空白;二是设计 AgriAgent 架构,明确大语言模型在农业环境控制系统中“感知-决策-行动”的结构化流程,实现环境感知、数据处理、工具调用与反馈整合的无缝衔接;三是通过实验验证有效性,对比 AgriAgent 与基线模型的性能,证明评测平台的评估价值与 AgriAgent 的决策优势,为农业领域大模型智能体的应用研究提供方法参考。

1 理论基础

1.1 大语言模型

LLMs 基于变换器架构 (Transformer Architecture),通过大规模文本预训练、监督微调 (Supervised Fine-Tuning) 与人类反馈强化学习 (Reinforcement Learning from Human Feedback) 优化后,展现出卓越的语言理解、知识整合、复杂推理及文本生成能力,奠定了其作为智能决策系统核心的基础,模型核心是学习语言的概率分布,可表示为公式 (1)。

$$P(w_i | w_{<i}; \theta) = \text{Transformer}(w_{<i}; \theta) \quad (1)$$

式中: w_i 为第 i 个单词; $w_{<i} = (w_1, w_2, \dots, w_{i-1})$ 代表历史上下文; θ 为模型参数。尽管学习到了大量语言的概率分布知识,然而 LLMs 在现实任务中仍面临显著局限:

首先,“幻觉 (Hallucination)”现象^[13]可能导致农业领域决策失误。其次,LLMs 缺乏物理世界交互能力,无法直接感知环境的动态变化。再次,LLMs 在访问实时数据库、调用专用应用程序编程接口 (Application Programming Interface, API) 或控制外部设备方面能力有限,且其生成的自然语言指令与外部工具所需的结构化输入之间存在转换障碍。最后,LLMs 的知识依赖预训练数据。对于预训练之后出现的新知识、动态变化的信息,无法及时获取和更新^[14],这限制了其在需要高度时效性决策场景中的应用潜力^[15]。

1.2 工具学习与工具增强型 LLMs

为突破上述局限,工具增强型 LLMs^[16]通过集成外部工具扩展能力边界,目标是构建一个具备工具选择与调用能力的扩展模型 M_{tool} ,表示为公式 (2)。

$$M_{\text{tool}}(x) = \text{Plan}(x; \theta) = [(p_1, o_1), (p_2, o_2), \dots, (p_k, o_k)] \quad (2)$$

式中： x 为模型输入； p_t 为第 t 步动作，可以是语言生成或工具调用； o_t 为传递给工具 p_t 的结构化参数； $\text{Plan}(x; \theta)$ 为多步推理与行动序列。工具增强型LLMs的决策受扩展模型 M_{tool} 输出影响，其形式化表示为 $P(w_t|w_{<t}; \theta)$ ，其中 $w_{<t}$ 包含由 M_{tool} 产生的 $w_{\text{tool}} = \text{Transformer}(M_{\text{tool}}(x))$ 。

其核心协同技术包括：上下文学习 (In-Context Learning, ICL)、指令微调 (Instruction Fine-tuning)^[17]、复杂规划与推理框架^[18] (如推理-行动 (Reason and Act, ReAct) 框架^[19])。该范式赋予LLMs实时信息获取、精准计算分析、物理闭环控制3重优势，形成“感知-决策-执行”完整链路。

1.3 基于LLM的智能体(LLM-Based Agents)

智能体 (Agent) 是能感知环境、决策行动以达成目标的计算实体，核心架构为“感知-规划-行动 (Perception-Planning-Action) 闭环”^[20]，表示为公式 (3)。

$$\text{Agent} = (\mathbf{ST}, \mathbf{O}, \mathbf{A}, \pi, M) \quad (3)$$

式中： $\mathbf{ST}$ 为环境状态空间； $\mathbf{O}$ 为观测空间，即智能体感知的输入； $\mathbf{A}$ 为动作空间，包含当前智能体可执行的操作； π 为策略函数，即代表决策逻辑； M 为大模型中存在的长期记忆知识，辅助智能体决策。其“感知-规划-行动”的闭环核心架构可表示为公式 (4)。

$$\forall \tau \in N: \begin{cases} \mathbf{o}_\tau = O(\mathbf{s}_\tau) \\ \mathbf{a}_\tau = \pi(\mathbf{o}_\tau) \\ \mathbf{s}_{\tau+1} = f(\mathbf{s}_\tau, \mathbf{a}_\tau) \end{cases} \quad (4)$$

式中： $\mathbf{o}_\tau$ 、 $\mathbf{a}_\tau$ 、 $\mathbf{s}_\tau$ 分别为当前时间步 τ 下的观测数据、执行动作、环境状态向量； f 为状态转移函数。智能体系统随着时间的推演，循环执行闭环控制任务序列。

该类智能体在农业应用中仍面临特殊挑战：农业生产环境动态多变，要求智能体具备高度适应性；农业决策依赖农学、土壤学等深层次专业知识，通用LLM领域专业知识不足；农业决策高风险长周期；且存在多源数据整合、工具调用安全及领域微调数据缺乏等部署难题。

1.4 DSSAT 作物生长模型

作物生长模型 DSSAT (Decision Support Sys-

tem for Agrotechnology Transfer) 作为国际公认的综合模拟平台^[21,22]，整合了作物生理生态、土壤过程、气象与管理模块，提供标准化仿真框架。其模拟引擎基于物候发育、光合作用、干物质分配、水分与养分动态等关键过程，能够逐日模拟作物生长，输出叶面积指数、生物量、土壤水分及最终产量等指标。其强解释与预测能力，使其可作为高保真数字孪生环境的核心引擎，为智能体训练、决策验证与性能评估提供领域知识支撑。

2 系统构建与实验实施方法

图1展示了工具增强型智能环控智能体系统的整体架构，系统由数字孪生农场和农业决策系统 AgriAgent 两部分组成，其中数字孪生农场包括：1) 从传感器中收集到的原始数据，以及基于数据增强技术生成的各类仿真数据；2) 基于 DSSAT 核心的作物生长模型；3) 一个可以控制孪生农场内各类设施的控制系统。AgriAgent 通过分析和模拟孪生农场中作物的生长数据做出智能决策，并生成6维控制指令控制农场各种执行器设备，包括：灌溉、补光、通风、加温、施肥，以及CO₂施放，调整环境状态以达到控制作物产量的目的。一次完整的决策控制流程如下所示。

1) 数据收集：各种传感器负责采集数据，包括土壤数据、气候数据、空气数据和作物数据。

2) 数据处理与模拟：采集到的数据进入数字孪生农场，在这里被处理并用于作物生长模型。这些数据经过数据增强技术被转化为孪生数据。例如，原始气候数据经过数据处理和情景扩充转换成常规 (Regular)、扰动 (Perturbed)、极端 (Extreme) 气候，来模拟多场景的气候；专家校准的固定土壤剖面参数作为孪生土壤数据。

3) 作物生长模型用于模拟作物的生长状况，它通过输入孪生数据和田间管理措施来逐日模拟作物生长状况，DSSAT内核以1d为决策步骤推进模拟。

4) 智能决策：每个决策周期 AgriAgent 从数字孪生农场获取环境数据和作物生长数据，系统将该数据与记忆库中的全局信息，以及利用检索增强生成 (Retrieval-Augmented Generation, RAG) 技术从知识库中查询到的相关知识共同载入LLM上下文。最后，模型通过隐式思维链推理，输出JSON格式的决策指令。

5) 决策执行：生成的控制指令被发送回数字孪生农场，用于更新下一轮的模拟数据，形成一个

持续优化的反馈循环。同时,这些指令通过执行器(如灌溉、施肥、光照和通风系统等)转化为实际操作,实现对农场的智能化管理。

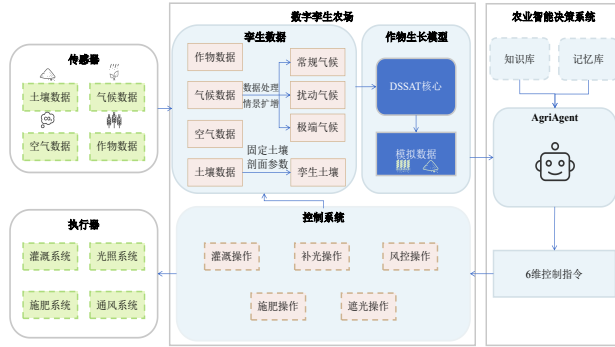


图1 工具增强型环控智能体系系统架构示意图

Fig. 1 System architecture of the tool-augmented environmental control agent

2.1 农业评测平台构建

在数字孪生设施农场建模前,需要构建1个标准化的评估平台,用于评估智能体在农业场景中的工具调用准确性、决策合理性,以及对农业效益的影响。本小节将从数据收集和仿真平台搭建两方面着重介绍标准化评估平台的构建。

2.1.1 数据收集

为确保评测平台任务与真实生产场景保持高度一致,首先需汇集多源、异构且覆盖全年的原始环境数据并进行标准化处理。本研究将Kaggle数据科学竞赛平台(Kaggle)上的首尔历史气象数据集(Seoul Historical Weather Data)作为气候驱动来源,该数据集按日记录了30年来的最高/最低/平均气温、降水量、相对湿度、风速、日照时数等14个关键气象变量,字段齐全且缺失极少,可直接对接DSSAT模型的输入需求,之后对初始数据进行如下处理。

1) 截取1994—2023年共30年的完整记录;将温度从华氏度转换为摄氏度,降水量维持mm/d计量,太阳辐射强度(单位通常为W/m²)转换为每日总太阳辐射量(单位为MJ/(m²/d))。

2) 对偶发缺测样本采用邻近7d滑动均值进行填补。

3) 每1个公历年视为1条独立“气候轨迹”,对应作物的1个完整生长周期,由此生成30条互不重叠的季节样本。

4) 在原始序列基础上构造3种气候情景。

常规情景(Regular)直接采用实测序列,记为

$T_{\text{reg}}(y, i)$ 与 $P_{\text{reg}}(y, i)$,分别代表第 y 年第 i 天的温度与降水。

扰动情景(Perturbed)在常规序列上叠加零均值高斯噪声,以检验策略对短期随机波动的鲁棒性。其生成方法由公式(5)定义。

$$\begin{cases} T_{\text{pert}}(y, i) = T_{\text{reg}}(y, i) + \varepsilon_T, \varepsilon_T \sim N(0, 1.5^2) \\ P_{\text{pert}}(y, i) = P_{\text{reg}}(y, i) + \varepsilon_P, \varepsilon_P \sim N(0, (0.2 \times P_{\text{reg}}(y, i))^2) \end{cases} \quad (5)$$

式中: $T_{\text{pert}}(y, i)$ 为扰动情景下第 y 年第 i 天的温度; $P_{\text{pert}}(y, i)$ 为扰动情景下第 y 年第 i 天的降水;温度噪声 ε_T 的标准差为1.5℃;降水噪声 ε_P 的标准差为当日常规降水值 $P_{\text{reg}}(y, i)$ 的20%。

极端情景(Extreme)是基于统计学方法设计的多因子耦合胁迫(Multi-Factor Coupled Stress, MFCS)情景,具体生成方法如下。

计算年度极端阈值。针对第 y 年的常规气象序列,计算其第95百分位数作为该年的极端事件阈值,如公式(6)所示。

$$\begin{cases} T_{95}(y) = \text{quantile}(T_{\text{reg}}(y, :), 0.95) \\ P_{95}(y) = \text{quantile}(P_{\text{reg}}(y, :), 0.95) \end{cases} \quad (6)$$

式中: $\text{quantile}(X, p)$ 为计算序列 X 的 p 分位数; $T_{95}(y)$ 为第 y 年温度的95百分位阈值; $P_{95}(y)$ 为第 y 年降水的95百分位阈值。

提取极端事件集合。筛选出该年中所有超过阈值的极端日值,构成高温与暴雨的候选事件池,如公式(7)所示。

$$\begin{cases} T_{\text{extreme}}(y) = \{T_{\text{reg}}(y, i) | T_{\text{reg}}(y, i) > T_{95}(y)\} \\ P_{\text{extreme}}(y) = \{P_{\text{reg}}(y, i) | P_{\text{reg}}(y, i) > P_{95}(y)\} \end{cases} \quad (7)$$

式中: $T_{\text{extreme}}(y)$ 为第 y 年高温极端事件集合; $P_{\text{extreme}}(y)$ 为第 y 年暴雨极端事件集合。

构建MFCS情景序列。在作物生长的敏感阶段,随机选取1个连续的10d窗口 $W(y)$ 。在该窗口内,每日的气温与降水值分别从其对应的极端事件集合中有放回地随机抽样赋值,从而确保热浪与洪涝压力同步发生。窗口外时段则保持常规序列值,如公式(8)和公式(9)所示。

$$T_{\text{mfcs}}(y, i) = \begin{cases} \text{sample}(T_{\text{extreme}}(y)), & \text{if } i \in W(y) \\ T_{\text{reg}}(y, i), & \text{if } i \notin W(y) \end{cases} \quad (8)$$

$$P_{\text{mfcs}}(y, i) = \begin{cases} \text{sample}(P_{\text{extreme}}(y)), & \text{if } i \in W(y) \\ P_{\text{reg}}(y, i), & \text{if } i \notin W(y) \end{cases} \quad (9)$$

式中: $T_{\text{mfcs}}(y, i)$ 为多因子耦合胁迫情景下第 y 年第 i 天的温度; $P_{\text{mfcs}}(y, i)$ 为多因子耦合胁迫情

景下第 y 年第 i 天的降水; $\text{sample}()$ 为随机抽样函数。

此设计通过迫使智能体在决策时同时应对高温与强降水的耦合胁迫, 专门用于测试其系统恢复力与在复杂非线性环境下的策略调整能力。

经上述处理后, 30条原始气候轨迹最终派生出90条驱动序列。除气象序列外, 数字孪生模型还需初始土壤数据。由于公开土壤数据缺乏足够长的时间序列记录, 本研究采用经专家校准的固定土壤剖面参数作为所有实验的统一初始条件。

作物方面, 为充分评估 LLM-Based Agent 的通用性, 本研究选用 DSSAT 支持的作物作为测试对象, 以确保覆盖粮、经、蔬等多元作物生态。通过上述设计, 评测平台在保持高保真度的同时, 可提供丰富且可控的气候与土壤初始条件, 为农业决策模型的综合评测奠定坚实数据基础。

2.1.2 极端情景设计

为确保模型在复杂、非线性环境下的鲁棒性与泛化能力, 本研究基于统计学方法设计了多因子耦合胁迫情景来模拟极端情景, 并通过实际灾害案例来支撑情景的有效性。

多因子耦合胁迫情景设计, 为模拟设施农业中更具破坏性的多因子耦合胁迫, 本研究构建了以下3种典型复合情景。

1) 热旱复合胁迫。连续14 d最高温度超过历史95%百分位; 同时降水设为0。模拟夏季设施大棚内高温高湿(因蒸散)与持续性缺水(干旱)并存, 作物遭受光抑制、热胁迫与水分胁迫的复合影响。借鉴了2013年和2022年长江中下游地区夏季持续高温干旱, 导致经济作物遭受的复合热/水胁迫案例。

2) 暴雨-低温-弱光复合胁迫。连续15 d降雨量大于80 mm(代表区域特大暴雨); 且连续5 d最低温度降低和光辐射降低。模拟梅雨季或台风期的内涝、湿渍导致的根系缺氧, 以及低温弱光导致的光合作用受阻环境。参照2021年河南特大暴雨事件, 造成作物长时间湿渍、根系缺氧的复合灾害模式。

3) 周期性灾害反转。20 d热旱复合胁迫后, 紧接着是10 d暴雨-淋溶复合胁迫。专门测试 Agri-Agent 的系统恢复力。要求智能体在环境从极端干旱急剧转为极端湿涝后, 能迅速调整策略以最小化损失并促进作物恢复。

核心仿真引擎 DSSAT 作物生长模型基于机理过程, 能够定量模拟作物在水分平衡、养分动态

和胁迫响应等关键生理生态过程, 确保了设计的 MFCS 情景能够真实地反映对作物生长过程的综合影响。

2.1.3 评测平台搭建

本研究基于 DSSAT v4.8 构建可插拔式农业决策评测平台, 采用“指令层-工具调度层-仿真执行层”3级架构。平台以90条年度气候序列与统一土壤为驱动, 实现农业智能体的闭环验证。

1) 指令层。大模型智能体通过函数调用协议生成 JSON 格式动作指令, 以产量最大化为目标, 通过提示语(Prompt)工程调用预置 API。

2) 工具调度层。执行语法与物理约束校验(如单步灌溉量 ≤ 30 mm、施氮量 ≤ 200 kg/hm²), 超限操作自动裁剪至设备极限并记录日志; 将校验后指令映射至对应 API, 触发执行并返回结果至 Agent 上下文。

3) 仿真执行层: DSSAT 内核以1 d为决策步推进模拟, 保留作物生理过程精度; 提供两类封装 API——环境状态感知(实时数据获取)与环境参数调控(动态修改); 每步模拟结束后返回更新状态, 重启指令层决策循环。

该平台通过“文本交互-工具调用-高频仿真”闭环, 既能在周尺度捕捉灌溉、施氮等操作细节, 又能在年度尺度输出最终产量等农艺收益, 为算法研究提供统一、高保真的评测基准。

2.2 数字孪生农场建模

本节旨在阐述数字孪生农场建模方案, 为后续 LLM-驱动农业智能体的评估提供统一、可复现的实验平台。

2.2.1 建模目标与原则

数字孪生农场的建模遵循以下核心原则。

1) 完整性: 覆盖土壤、气象、作物三大要素, 并支持27类常见作物的全生育期模拟。

2) 可交互性: 将环境状态和控制量分别封装为 Sensor API 与 Actuator API, 便于语言模型通过函数调用实现“感知-决策-执行”闭环。

3) 安全可重复: 采用 DSSAT v4.8 作物生长模型为仿真内核, 在沙盒环境迭代策略, 避免真实温室试错的高成本与风险。

4) 可评估: 以产量、资源利用效率、胁迫指数及能耗等多维指标衡量控制策略的有效性, 并通过统计检验验证显著性。

2.2.2 环境状态空间

在数字孪生农场系统中, 状态向量 ST 是 1 个多维向量, 用于描述在时间步 t 时农场的环境条件和作物状态。它通过传感器数据实时采集, 反映了土壤、气候和植物健康的关键指标。这些信息为农场管理提供了数据支撑, 使管理者或自动化系统能根据当前状态采取优化措施 (如灌溉、施肥或通风), 进而提高作物产量和资源利用效率。数字农场在每个时间步 τ 输出状态向量, 如公式 (10) 所示。

$$ST = [W_{\text{soil}}, N_{\text{total}}, OM, T_{\text{avg}}, T_{\text{max}}, P, WS, CO_2, GS, D_{AP}, S_{\text{water}}, S_{\text{growth}}]_t \quad (10)$$

式中: W_{soil} 为土壤可利用水量, mm; N_{total} 为土壤全氮含量, kg/hm²; OM 为土壤有机质含量, t/hm²; T_{avg} 为平均气温, °C; T_{max} 为最高气温, °C; P 为日降水量; WS 为风速, m/s; CO_2 为二氧化碳浓度, $\mu\text{mol/mol}$; GS 为生育阶段; D_{AP} 为播种后天数, d; S_{water} 为水分胁迫指数; S_{growth} 为综合胁迫指数。

复合环境胁迫指数计算采用“分生理过程适配”策略。最小值法源于 Liebig 最小因子定律 (Liebig's Law of the Minimum)^[23], 认为作物生长受单一最限制因子制约, 适用于光合作用等即时生理过程 (如极端高温、严重缺水会直接阻断光合, “最短板因子”影响占主导); 乘法模型^[24]则适配生物量积累等长期过程, 可通过“乘积放大/衰减”体现多胁迫协同效应 (如轻度水分亏缺与氮素不足的负协同, 乘积结果能反映累积影响), 本研究中光合作用用最小值法、生物量积累用乘法模型, 具体计算如公式 (11) 和公式 (12) 所示。

1) 最小值法:

$$S_{\text{growth}} = \min(S_{\text{water}}, S_{\text{Nitrogen}}, S_{\text{Temperature}}, S_{\text{Photo}}, \dots) \quad (11)$$

式中: S_{Nitrogen} 为氮素胁迫指数; $S_{\text{Temperature}}$ 为温度胁迫指数; S_{Photo} 为光合胁迫指数。

2) 乘法模型:

$$S_{\text{growth}} = S_{\text{water}} \times S_{\text{Nitrogen}} \times S_{\text{Temperature}} \times S_{\text{Photo}} \times \dots \quad (12)$$

2.2.3 控制动作空间

在数字孪生农场中, 自动化控制系统可以调控农场中的控制器, 具体包括的控制动作由 1 个六维向量表示, 如公式 (13) 所示。

$$A_t = [I, L, V, H, F, C]_t \quad (13)$$

式中: t 为时间步, 用于描述在特定时间点或阶段执行的控制策略, 各动作的含义与作用机制如表 1 所示。

表 1 AgriAgent 控制系统建模

Table 1 AgriAgent control system model

设备	控制动作符号	范围/分辨率	主要影响
灌溉系统	I	0~30 mm/h	提升 W_{soil}
补光系统	L	0~200%	提升光合有效辐射
通风/加温	V, H	0/1(开关)	调节 T_{avg}, WS
施肥系统	F	0~50 kg/hm ²	增加 N_{total}
CO ₂ 施放	C	0~800 $\mu\text{mol/mol/h}$	提升室内 CO ₂

2.2.4 数字孪生农场功能展示

图 2 展示了为 AgriAgent 智能决策系统设计的数字孪生农场可视化平台。该平台旨在提供 1 个高保真、可交互的沙盒环境, 以验证和评估智能体在复杂农业场景下的决策能力。



图 2 数字孪生农场功能示意图

Fig. 2 Diagram of digital twin farm function

平台界面分为 3 个核心区域。

1) 环境状态展示区 (左侧): 实时显示由 DS-SAT 作物生长模型模拟生成的环境数据, 包括天气数据 (降水量、气温、风速)、土壤数据 (有机质、总氮、总水分) 和空气数据 (CO₂ 浓度)。这些数据模拟了真实的传感器输入, 为智能体提供了“感知”环境的依据。

2) 操作记录区 (右上): 详细记录了智能体所执行的控制指令, 如灌溉量和施肥量及深度。这部分数据作为“执行”模块的反馈, 让研究人员能直观地追踪智能体的决策行为。

3) 智能体对话区 (底部): 以对话形式呈现智能体的工作流程, 包括其获取环境信息、进行分析推理, 以及最终做出决策的过程。这个模块清晰地

展示了“思考”与“行动”之间的逻辑链，例如，智能体如何根据土壤水分情况决定调用灌溉工具，并通过对话进行反馈。

2.3 AgriAgent 架构

2.3.1 系统组件与功能

为发挥大型语言模型在复杂控制场景中的泛化与推理能力，本研究设计了一个专注于设施农业环境控制的 LLM-based Agent——AgriAgent。其核心推理引擎采用开源大语言模型 Qwen2.5 系列（由阿里巴巴集团发布）。该模型基于 Transformer 架构，具有强大的自然语言理解、工具调用与多步推理能力。其整体信息流如图 3 所示，核心由“传感器-记忆库-检索器-LLM-工具执行器”5 个子模块构成，形成“感知-回忆-检索-思考-行动”的闭环。各组件的核心职责及关键实现要点如表 2 所示。

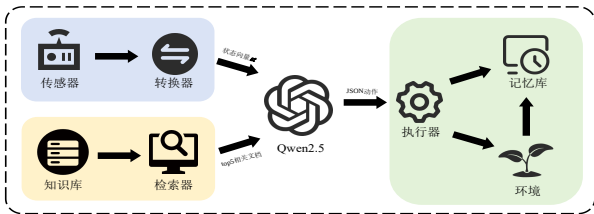


图 3 AgriAgent 架构示意图
Fig. 3 Architectural diagram of AgriAgent

表 2 AgriAgent 系统组件核心职能与关键实现要点
Table 2 Core functions and key implementation points of Agri-Agent system components

模块	功能摘要	实现要点
传感器	从仿真环境读取状态向量 s_t	以 API 形式模拟真实传感器
记忆库	持久化存储全局状态摘要	以文本形式保存在 System prompt 中
格式转换器	将结构化状态拼接为自然语言片段	以固定规则实现
检索器(Retriever)	面向知识库检索作物生理模型、控制经验	以 Faiss 检索实现
LLM	解析环境信息并生成 JSON 格式工具调用	经过预训练和后训练的大语言模型
工具执行器	校验、裁剪并调用修改 DSSAT 的接口	Python 实现

2.3.2 构建知识库

大型语言模型在农业决策中迫切需要作物生理与田间管理知识。为弥补这一空白，本研究从零构建了一个面向中文作物生产场景的小型知识库，其构建工作分为“原始资料采集-数字化与清洗-结构化标注-检索索引”4 个阶段。

在资料采集阶段，本研究以“可公开获取、覆

盖主粮与经济作物、信息粒度细至操作规程”为原则，筛选出 1 部中文农业教材和技术手册，所有 PDF 与扫描版资料统一经 Pix2Text 识别，转换为 UTF-8 编码的 Markdown 文本，总字数约 5.1 M。

数字化后，对文本进行了细粒度清洗：去除目录、页眉页脚及重复页；统一计量单位；利用正则与领域词典纠正常见光学字符识别（Optical Character Recognition, OCR）错字；对噪声过多的段落执行相似度阈值过滤，最终保留 4.4 M 有效字符。

结构化环节中，为便于检索与提示拼装，本研究采用“半结构化标签+段落原文”的混合分块方式。每 1 段正文均绑定 3 个核心元数据：章节题目（Section）、小节题目（Title）、内容（Source）。

检索索引采用稠密向量检索和稀疏向量检索的混合方案：文本向量由 bge-large-zh 生成（768 维）；检索时通过向量召回 5 条候选，为适配 LLM 提示长度限制，每条返回文档会被压缩为固定长度。

2.3.3 推理流程

在数字孪生农场的周度循环中，AgriAgent 持续执行“感知-回忆-检索-思考-执行-写回”6 阶段决策流程。每个周期以传感器从 DSSAT 引擎获取 12 维环境向量为起点，该数据与记忆库中的全局信息共同载入 LLM 上下文。随后进入知识检索阶段：格式转换器基于模板将环境向量与记忆内容生成结构化查询（如“当前作物：番茄；阶段：开花坐果；土壤水势：-55 kPa；目标：水肥调控”），由向量检索器从知识库召回 5 条相关文档（例如，“番茄花期土壤水势维持在 -40~-60 kPa 可降低畸形果率”）。

LLM 基于当前状态、环境摘要及先验知识 3 要素生成不超过 2 k token 的 Prompt，并遵循预置函数调用规范。模型通过隐式思维链推理（如综合水势偏离目标区间 5 kPa、氮素充足及低湿度等因素），输出 JSON 格式决策（如“补灌 8 mm、不施氮、维持 100% 补光”）。格式转换器执行语法与物理约束校验后触发工具调用，同时将执行结果同步写回系统。

整季生长结束后，仿真平台输出最终产量及全周期决策日志，完成农业时序决策闭环。该机制通过显式知识补全，实现“历史-先验-现状-目标”四维感知，支撑作物水肥光热参数的动态优化。

2.4 实验设计

为验证前述 AgriAgent 在实际农业控制任务中

的有效性,本研究按照性能评估的标准流程,在数字孪生温室中对多种作物-气候组合进行了系统评测。本节详细说明实验环境、数据划分、评测指标、对比基线及关键实现细节等设置。

仿真环境与数据源:所有实验均运行于 DS-SAT v4.8 内核的数字孪生平台,决策步长为 7 d。气候驱动采用从 Seoul Historical Weather Data 中提取的 1994—2023 年共 30 条年度序列,经常规、扰动、极端 3 种生成策略扩增为 90 条驱动轨迹。土壤数据来自中国农业科学院新乡实验基地,同时使用 DSSAT 土壤养分模块模拟氮素消耗,这为 AgriAgent 提供了更复杂、更接近真实的决策反馈环境。具体理化性质与特征参数如表 3 所示。

表 3 数字孪生农场土壤的理化性质与特征参数
Table 3 Soil physicochemical properties and characteristic parameters for digital twin farm

土层深度/cm	[0,20)	[20,40)	[40,60)	[60,80)	[80,100)	
黏粒	21.6	21.4	20.8	11.2	10.4	
颗粒组成/%	粉粒	78.4	78.6	78	78.1	69.3
	砂粒	0.0	0.0	1.2	10.7	20.3
土壤容重/(g/cm ³)	1.47	1.44	1.52	1.54	1.46	
田间持水量/(cm ³ /cm ³)	0.306 0	0.274 9	0.319 1	0.375 6	0.328 8	
永久凋萎点/(cm ³ /cm ³)	0.080 6	0.095 2	0.093 1	0.072 3	0.093 1	
土壤有机碳/%	1.0	0.5	0.5	0.4	0.3	
pH	8.47	8.80	8.82	8.84	8.85	
饱和含水量/(cm ³ /cm ³)	0.410 5	0.420 9	0.257 8	0.426 0	0.429 2	
饱和导水率/(cm/d)	45.60	25.13	55.56	7.49	130.30	

测试作物与气候场景:选用玉米(C₄作物)、小米(耐旱作物)、甜菜(块根作物)、番茄(茄果类作物)和卷心菜(叶菜类作物)5种作物作为评价对象,覆盖不同生理类型,验证系统通用性。每种作物在 90 条气候轨迹上分别运行常规(Regular)、扰动(Perturbed)、极端(Extreme)3档难度,共得到 450 条完整生长季样本。

根据 Qwen2.5 技术报告^[25]的实验设计方式,本研究设计 4 类基线算法,尽可能充分地探索大模型在农业智能环境控制场景下的比例定律(Scaling Law)^[26]。

- 1) Baseline: 无任何调控,自然生长。
- 2) AgriAgent-1.5B: 基于开源模型 Qwen2.5-1.5B,参数量为 1.5 B,属于小规模模型,适用于基

础农业决策任务,具备良好的响应速度与部署灵活性。该模型的构建,可以探索大模型决策系统的最小可用架构,验证轻量化模型在实际应用中的可行性与稳定性。

3) AgriAgent-3B: 基于 Qwen2.5-3B,参数量为 3 B,同属小规模模型,但在上下文理解和语言生成方面表现更优,适合多轮交互任务。该模型在保持较低资源消耗的同时,提升了语义建模能力和任务适应性,适用于需要更高语言理解深度的农业咨询与辅助决策场景。

4) AgriAgent-7B: 基于 Qwen2.5-7B,参数量为 7 B,属于中规模模型,在复杂农业环境中展现出更强的推理能力和泛化性能,适用于高维度数据分析与策略生成。该模型在多任务协同、跨领域知识整合,以及复杂语义理解方面具备显著优势,是构建高性能农业智能体的关键支撑。

评价指标:沿用评估标准的既定度量,即作物最终经济产量——作物收获部位(籽粒/块根/果实等)干重(kg/hm²)。

实现细节:检索器采用 bge-arge-zh 向量(768 维)+Facebook AI 相似度搜索(Facebook AI Similarity Search, FAISS);查询召回 Top-5,检索延迟<40 ms。LLM 推理参数为 top_p=0.9、temperature=1。

2.5 测评指标

本研究以作物最终经济产量(特指收获部位生理成熟时的干物质重量)作为核心评估指标。此产量由 DSSAT 模型模拟生成,为干重,并非农学中的籽粒鲜重或标准含水率下的产量,其绝对数值与产量存在显著差异。该指标不仅直接反映农业生产的核心目标——最大化有效收获物,还能综合衡量智能体通过工具调用实施生育期管理决策的效果,而高产量则表明智能体成功识别了作物需求、有效应对了环境胁迫并优化了资源利用。

尽管资源利用效率、作物胁迫指数及生产成本等辅助指标能更全面评估系统性能(如环境可持续性与经济效益),但本研究聚焦于验证工具增强型 LLM 的农业决策基础能力。经济产量因其直接性(关联核心目标)、综合性(涵盖全生育期决策)与易量化性,被确立为首要评估依据。

3 结果与分析

3.1 主实验结果

本实验评估了基线模型(Baseline)及不同规

模的AgriAgent模型（1.5B/3B/7B）在5种作物（玉米、小米、甜菜、番茄、卷心菜）与3种环境情景（常规Regular、扰动Perturbed、极端Extreme）下的农业决策性能。结果以作物收获部位干重（kg/hm²）衡量，核心数据如图4和表4所示。

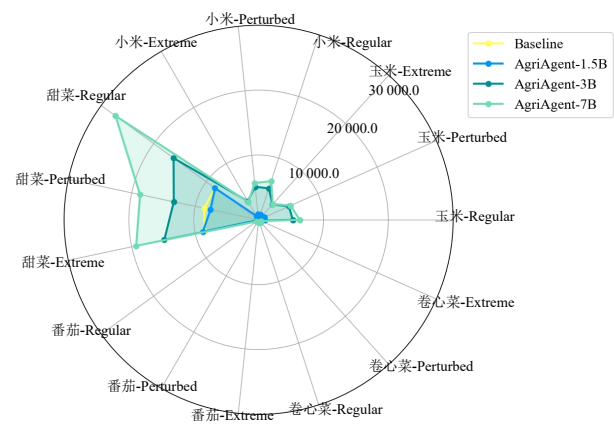


图4 AgriAgent模型(1.5B, 3B, 7B)与基线模型在5种作物和3种情景下的性能对比(以最终作物干物质产量(kg/hm²)衡量,数据为30个独立气候轨迹的平均值)

Fig. 4 Performance comparison of the AgriAgent models (1.5B, 3B, 7B) with the baseline model for five crops and three scenarios

3.1.1 核心发现

1) Baseline 模型。在常规情景下，所有作物均维持基础产量水平。然而，在极端情景下，所有作物产量均出现显著下降，尤其是玉米和卷心菜分别下降约39.4%和60.4%，验证了MFCS情景对作物生长的显著胁迫效应。番茄在极端情景下产量下降43.9%，进一步说明无主动调控策略难以应对复杂环境变化。

2) AgriAgent-1.5B模型。在常规、扰动及极端情景下，玉米、小米、番茄、卷心菜的模拟产量与Baseline完全一致。仅甜菜在各情景下的产量与Baseline接近（常规情景差异<0.1%）。这一结果客观反映了小规模模型的性能局限性，其未能学习到超越Baseline的调控策略，决策鲁棒性不足。

3) AgriAgent-3B模型。在所有作物和情景中，AgriAgent-3B均显著优于Baseline与1.5B模型。例如，在极端情景下，玉米、小米、甜菜、番茄和卷心菜的产量分别提升450.2%、375.1%、73.7%、1256.3%和785.1%，展现出中等规模模型在多作物环境下的初步适应能力。

4) AgriAgent-7B模型。AgriAgent-7B在所有5

表4 不同参数量AgriAgent模型在5种作物与3种情景下的平均最终产量干重

Table 4 Average final yield dry weight of the AgriAgent model under five crops and three scenarios conditions with different parameter amounts

单位: kg/hm ²				
作物	模型	常规	扰动	极端
玉米	Baseline	990.10±144.55	997.60±142.00	600.67±524.62
	AgriAgent-1.5B	990.10±144.55	997.60±142.00	600.67±524.62
	AgriAgent-3B	5440.80±1520.62	5287.33±1471.01	3304.53±2950.12
	AgriAgent-7B	6557.83±950.12	5584.17±1200.45	3385.20±2900.34
小米	Baseline	867.20±268.49	894.63±268.79	721.97±426.74
	AgriAgent-1.5B	867.20±268.49	894.63±268.79	721.97±426.74
	AgriAgent-3B	5281.03±1174.62	5291.87±1428.79	3431.40±1948.76
	AgriAgent-7B	6474.03±1400.56	5892.70±1200.78	3257.83±1800.45
甜菜	Baseline	8587.90±6250.11	8853.70±6459.85	8845.07±7099.96
	AgriAgent-1.5B	8592.87±6265.19	7835.5±6465.73	9081.30±7113.27
	AgriAgent-3B	16755.30±5911.93	13818.70±6247.26	15365.40±7188.85
	AgriAgent-7B	28186.00±2745.67	19315.33±2400.50	19938.23±2800.25
番茄	Baseline	39.43±57.93	35.20±135.13	22.13±91.36
	AgriAgent-1.5B	39.43±57.93	35.20±135.13	22.13±91.36
	AgriAgent-3B	367.97±178.01	352.67±142.79	300.03±194.80
	AgriAgent-7B	404.63±120.34	380.00±150.67	362.17±200.12
卷心菜	Baseline	88.87±6.52	84.97±7.32	35.20±25.99
	AgriAgent-1.5B	88.87±6.52	84.97±7.32	35.20±25.99
	AgriAgent-3B	398.53±127.86	336.60±158.31	311.57±142.41
	AgriAgent-7B	534.70±80.12	513.77±70.45	452.27±90.23

种作物和3种情景下均取得最优表现。常规情景、扰动情景和极端情景下玉米产量较基线模型分别提升562.20%、459.70%、463.60%；小米产量分别提升646.80%、558.70%、351.20%；甜菜产量分别提升228.20%、118.20%、125.40%；番茄产量分别提升926.20%、979.55%、1 537.46%；卷心菜产量分别提升501.67%、504.65%、1 185.14%。充分体现了大模型在复杂决策中的灵活性与抗逆性。

3.1.2 分析与讨论

实验结果表明, AgriAgent-7B模型在5种作物与3种气候情景下均展现出卓越的适应性及决策一致性。其在极端情景中通过动态感知与策略调整, 显著缓解了环境胁迫对作物生长的影响, 特别是在番茄和卷心菜等经济作物上表现尤为突出。多作物、多情景的全面验证, 充分体现了AgriAgent架构在设施农业环境控制中的高度灵活性与可扩展性。通过集成工具增强的“感知-决策-执行”闭环, AgriAgent不仅在单一作物上实现产量突破, 更在多种作物系统中展现了强大的通用决策能力, 为智能农业系统的跨作物部署提供了可靠的技术路径。

3.2 调控过程分析

图5展示了在AgriAgent-7B模型决策调控下, DSSAT模型模拟的玉米、小米和甜菜3种作物收获部位干重(Harvest Weight, Harwt)随时间的变化曲线。这些曲线直观地反映了智能体在不同生长阶段通过动态调整管理决策(如灌溉、施肥等)对最终收获物重量的影响。

玉米: 玉米的产量在4月底至5月初出现了1个显著的增长, 随后在6月中旬又达到了新的高峰, 最终在6月下旬稳定在高位。这表明LLM可能在该时期根据DSSAT的模拟反馈, 执行了关键性的管理操作, 从而有效提升了收获物最终重量。同时, 在5月中旬玉米的产量出现了1个明显的下降, 说明LLM在该时期可能做出了错误的管理决策(如不合理的灌溉、施肥等), 导致产量的下降。

小米: 小米的产量曲线呈现出平稳增长的趋势, 尤其是在4月下旬和6月中旬出现了两次明显的跃升。这可能与智能体在这些时间点进行了精准的灌溉或施肥决策有关, 满足了作物关键生长期的水分和养分需求。

甜菜: 甜菜的产量从4月开始持续稳定增长, 特别是在5月下旬和6月初, 增速明显加快。这表明智能体成功地调控了甜菜的生长, 动态调整管理

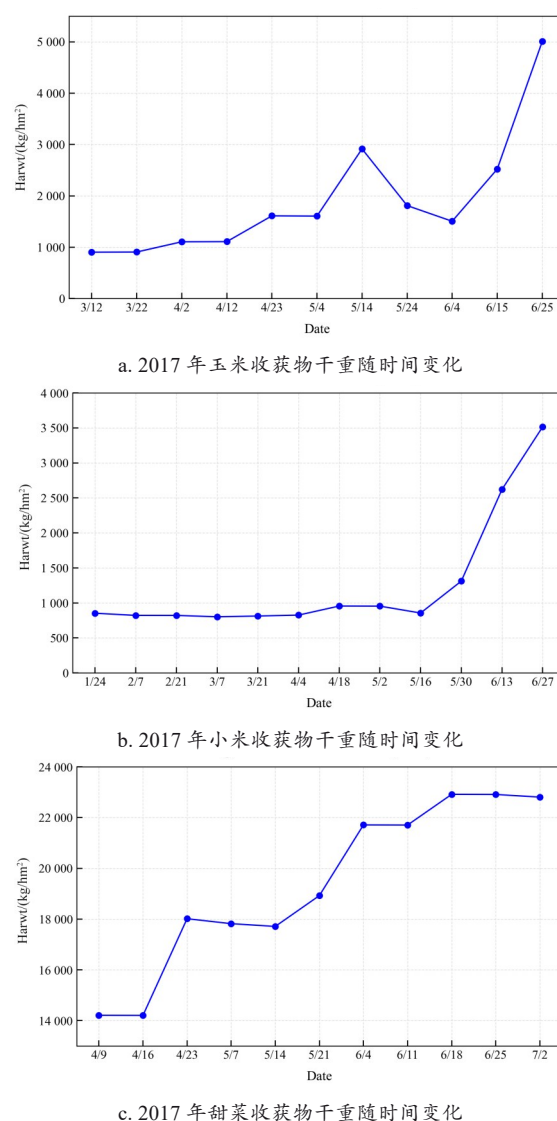


图5 AgriAgent-7B模型调控下玉米、小米和甜菜收获物干重随时间的变化曲线(常规情景)

Fig. 5 Time-dependent changes in harvested dry weight of corn, millet, and sugar beets under the AgriAgent-7B model (conventional scenario)

决策确保了在关键块根膨大期有充足的资源, 从而实现了产量的持续攀升。

这些动态变化曲线共同证明了大模型决策在数字孪生环境中对作物生长的精细化调控能力。通过动态决策, 大模型能够识别作物的生长瓶颈, 并及时作出决策, 最终引导DSSAT模拟出更高的收获物重量。

3.3 样例分析

在图6的案例中, 工具增强型LLM驱动的设施农场环控助手通过标准化流程执行环境监测与调控任务。

system: You are Owen, created by Alibaba Cloud. 你是一个农场专业助手位于武汉, 每天启动一次, 每天运行你直到你不再调用工具就停止.....
农场信息: 1. 农场位于武汉, 湖北省, 中国. 2.....
书本相关知识: (第 1 章土壤水分运动基本原理 1.2 土壤水势 中: ϵ 为单位体积溶液中含有的溶质质量 (即溶液浓度); M 为溶质的摩尔质量; g 为重力加速度; R 为普适气体常量.....)

assistant: 为了让环境最适宜作物生长, 获得最大的农产品收获, 我们首先需要了解当前植物的生长情况。我将查询当前的植物的实时信息。
'tool_calls': [{'type': 'function',
'function': {'name': 'get_plant_data', 'arguments': {}}}]

tool: {'name': 'get_plant_data', 'content': {'growth_stress': {'value': 0.0, 'unit': 'none'}, 'growth_stage': {'value': '出苗期', 'unit': 'none'}, 'water_stress': {'value': 0.0, 'unit': 'none'}, 'days_after_planting': {'value': 9, 'unit': 'day'}}}

assistant: 根据获取的信息, 目前作物处于出苗期, 已经播种 9 天。当前作物没有生长胁迫和水分胁迫的情况。接下来, 我们需要检查土壤湿度和其他环境因素, 以确保作物能够健康成长。
首先, 我会检查土壤中的水分含量。

'tool_calls': [{'type': 'function',
'function': {'name': 'get_soil_data', 'arguments': {}}}]

tool: {'name': 'get_soil_data',
'content': {'organic_matter': {'value': '0%', 'unit': 't/hm²'}, 'total_nitrogen': {'value': '18851.7%', 'unit': 'kg/hm²'}, 'total_water': {'value': '687%', 'unit': 'mm'}}}

assistant: 当前土壤中的水分含量为 687mm, 总氮含量为 18.8517 万 t/hm², 有机质含量为 0t/hm²。考虑到作物正处于出苗期, 需要充足的水分来支持幼苗的生长, 而当前的土壤水分含量看起来是足够的。接下来, 我将检查当前的天气和空气状况, 以进一步确认是否需要调整环境条件。现在, 我将获取天气和空气的相关数据。
'tool_calls': [{'type': 'function',
'function': {'name': 'get_weather_data', 'arguments': {}}}]

tool: {'name': 'get_weather_data',
'content': {'rainfall': {'value': 10.0, 'unit': 'mm/d'}, 'temperature_max': {'value': 14.5, 'unit': '°C'}, 'temperature_avg': {'value': 12.3, 'unit': '°C'}, 'wind_speed': {'value': 5.2, 'unit': 'm/s'}}}

assistant: 根据获取的数据, 当前的天气情况如下: -日降雨量: 10.0mm/d-平均气温: 12.3 摄氏度-最高气温: 14.5°C-风速: 5.2m/s 这些数据表明, 当前的降雨量对于补充土壤水分是有利的, 但需要注意的是, 平均气温较低, 可能会对作物的生长产生一定影响。此外, 适度的风速有助于空气流通, 有利于作物的健康成长。

接下来, 我会检查空气中的二氧化碳浓度, 以便更好地了解作物的生长环境。

'tool_calls': [{'type': 'function',
'function': {'name': 'get_air_data', 'arguments': {}}}]

tool: {'name': 'get_air_data',
'content': {'co2': {'value': 384.4, 'unit': 'ppm'}}}

assistant: 当前空气中二氧化碳浓度为 384.4 μmol/mol。这个数值处于正常范围内, 有利于作物的光合作用。综上所述, 当前的环境条件对于作物的生长来说是较为理想的。不过, 考虑到气温较低可能会影响作物的生长速度, 我们可以适当调整温室内的环境条件来促进作物的健康生长。为了确保作物在出苗期能够得到充足的光照, 我们可以适当增加补充光系统的强度。假设我们将补充光强度增加至 120%, 以增强光照效果。

同时, 由于风速适中且不会造成过大的温度变化, 我们不需要立即调整温室的通风系统。
如果未来气温持续低, 我们可以考虑适当提高温室内的温度以促进作物生长。现在, 我将控制补充光系统以增加光照强度。
'tool_calls': [{'type': 'function',
'function': {'name': 'control_supplementary_lighting',
'arguments': {'lighting_intensity': 120}}}]

tool: {'control_supplementary_lighting',
'content': {'status': 'True'}}}

assistant: 补充光系统的控制已成功执行, 我们将补充光强度增加到了 120%。这将有助于提高作物在当前较低光照条件下的光合作用效率。至此, 我们已经根据当前的环境条件进行了必要的调整, 以确保作物能够在适宜的环境中生长。如果未来有任何新的环境变化或作物生长状况的变化, 我会继续监控并及时调整该环境, 以达到最佳的农作物生长效果。

图 6 AgriAgent 在设施环境控制中的“感知-决策-执行”闭环流程样例: 基于实时环境数据的工具调用决策

Fig. 6 An example of AgriAgent's "perception-decision-execution" closed-loop process in facility environmental control: tool invocation decisions based on real-time environmental data

该助手依托标准化工作流程展开环境监测与调控。首先, 实时获取作物生育阶段 (出苗期) 及胁迫状态 (无生长/水分胁迫)。其次, 同步采集土壤水分、氮含量、有机质含量、降水量、气温、风速及大气 CO₂ 浓度等多源数据。综合分析显示, 当前环境整体适宜作物生长, 但存在低温胁迫风险, 可能抑制出苗期生长速率。基于光合作用补偿机制, 系统决策将补光强度提升至 120% 以缓解低温效应; 执行调控后持续监测环境参数与作物响应, 形成“感知-决策-执行”闭环优化机制。

该案例验证了工具增强型 LLM 在多源数据融合、动态决策生成与自动控制闭环方面的技术优势。未来研究需进一步优化多重环境因子耦合作用下的决策鲁棒性, 并扩展数据源以增强系统适应性。

3.4 工具增强对决策幻觉的抑制效果分析

为验证工具增强对 LLM “幻觉”问题的抑制

效果, 本研究设计了基于真实气象数据的幻觉量化评估方案。通过对比有/无天气工具条件下 AgriAgent 的决策质量, 系统评估工具增强对农业决策准确性的提升效果。

3.4.1 幻觉量化指标设计

本研究提出两种幻觉评估体系, 兼顾幻觉发生率与影响程度, 实现决策幻觉的精准量化。

1) 加权幻觉率: 考虑幻觉影响程度的综合指标, 计算如公式 (14) 所示。

$$H_{\text{weighted}} = \frac{\sum_{i=1}^n w_i \times h_i}{N_{\text{total}}} \times 100\% \quad (14)$$

式中: w_i 为幻觉权重 (高影响为 1.0, 中高影响为 0.5, 中等影响为 0.2); h_i 为对应类型幻觉发生次数; N_{total} 为总决策步骤数。

2) 高影响幻觉率: 仅统计对产量和作物生存有直接影响的高权重幻觉占比。

依据影响程度分级, 结合天气条件与决策行为的关联, 识别出农业决策中的主要幻觉类型如下。

1) 高影响幻觉 (权重 1.0): 直接影响产量和作物生存的关键决策错误, 无缓冲空间, 此类幻觉的典型场景、对应的天气触发条件及产量损伤机制如表 5 所示。

表 5 高影响幻觉 (权重 1.0) 类型及核心特征: 决策错误、天气条件与产量影响关联

Table 5 Types and core characteristics of high-impact hallucinations (weight = 1.0): correlation between decision errors, weather conditions and yield impacts

决策错误	天气条件	对产量的影响
极端高温下灌溉不足	温度 > 作物临界高温	作物热胁迫, 光合作用受阻, 直接减产
极端高温下大量施肥	温度 > 作物临界高温 + 5 °C	烧苗, 根系损伤, 不可逆伤害
干旱条件灌溉不足	降水少 + 温度高	水分胁迫, 生长停滞, 严重减产
低湿度高温灌溉不足	湿度低 + 温度高	过度蒸腾, 水分失衡, 萎蔫死亡

2) 中高影响幻觉 (权重 0.5): 可能显著影响产量, 但存在一定缓冲调整空间, 此类幻觉的典型场景、对应的天气触发条件及产量损伤机制如表 6 所示。

3) 中等影响幻觉 (权重 0.2): 主要影响作物健康状态, 对产量的直接影响较弱, 此类幻觉的典型场景、对应的天气触发条件及产量损伤机制如表 7 所示。

本小节的幻觉量化与类型分析基于 2017 年单

表6 中高影响幻觉(权重0.5)类型及核心特征:决策错误、天气条件与产量影响关联

Table 6 Types and core characteristics of medium-high impact hallucinations (weight = 0.5): correlation between decision errors, weather conditions and yield impacts

决策错误	天气条件	对产量的影响
降水充足仍大量灌溉	降水>适宜范围上限	根系缺氧,病害滋生,中等减产
极端低温过量灌溉	温度<作物临界低温	冻害风险,生长延迟,可能减产
强降水大量施肥	降水>20 mm	肥料流失,浪费资源,轻度减产

表7 中等影响幻觉(权重0.2)类型及核心特征:决策错误、天气条件与产量影响关联

Table 7 Types and core characteristics of medium-impact hallucinations (weight = 0.2): correlation between decision errors, weather conditions and yield impacts

决策错误	天气条件	对产量的影响
高湿度过量灌溉	湿度>85%	真菌病害风险,品质下降
大风高温灌溉不足	风速大+温度高	轻微水分胁迫,恢复较快
低辐射大量施肥	辐射<阈值	肥料利用率低,资源浪费

一年份数据,选择该年份的核心原因如下:2017年包含完整的气象数据序列,可支撑逐日决策对比分析。并且逐日决策幻觉对比需要大量计算资源,单年份分析已能充分验证工具增强抑制幻觉的核心假设。

3.4.2 消融实验结果分析

通过对比 AgriAgent-7B 在有/无天气工具条件下的表现(表8),验证工具增强对幻觉的抑制效果。

表8 工具增强对决策幻觉的抑制效果(7B模型)
Table 8 Suppression effect of tool augmentation on decision hallucinations (7B model)

作物	是否调用天气工具	决策数	加权幻觉率/%	高影响幻觉率/%	产量/(kg/hm ²)
玉米	有	16	0.0	0.0	6 885
玉米	无	22	2.3	0.0	6 182
甜菜	有	35	4.9	2.9	28 460
甜菜	无	31	8.1	6.5	22 571

由表8可知。工具增强显著降低幻觉率:在甜菜作物中,引入天气工具后加权幻觉率从8.1%降至4.9%,降幅达39.5%;高影响幻觉率从6.5%降至2.9%,降幅达55.4%。幻觉率与产量负相关:幻觉率的降低直接转化为产量提升。甜菜在无工具条件

下高影响幻觉率为6.5%,对应产量22 571 kg/hm²;有工具条件下高影响幻觉率降至2.9%,产量提升至28 460 kg/hm²,增幅26.1%。

3.5 研究局限性分析与讨论

尽管本研究构建的 AgriAgent 系统在数字孪生环境中取得了显著的决策效果,但本研究必须深入讨论 DSSAT 作物生长模型所固有的局限性,以及这些局限性对研究结论外推和实际应用带来的不确定性。

1) 地表水文过程模拟的简化及其影响。DS-SAT 模型的核心优势在于对土壤-植物-大气连续体中水平衡、养分动态和作物生理过程的模拟。然而,其在模拟极端降水事件下的精细化地表水文过程方面存在不足。

无法模拟地表积水深度:模型采用“剖面排水”函数来模拟水分再分布,缺乏对地表微地形、排水设施效能及由此产生的地表积水深度变化的量化能力,无法为智能体提供“积水深度”这一关键感知信息,也使得智能体无法生成“紧急排水”等针对性控制指令。

对地表径流过程的处理简化:DSSAT对降水的处理优先考虑入渗,超出土壤入渗能力的水分被简单地处理为“径流损失”,但并未模拟这一径流产生的动态过程、汇流路径,以及对农田养分和土壤的冲刷效应。这种简化使得模型无法反映真实世界中暴雨导致的养分瞬时淋失和土壤侵蚀等问题,可能导致 AgriAgent 在施肥决策上的评估偏差。

2) 多重胁迫协同效应的量化偏差。DSSAT 模型通过乘法模型和最小律来耦合不同胁迫因子对作物生长的综合影响。尽管这种方法能够反映多重胁迫的叠加效应,但其本质上是一种简化后的经验性耦合。

真实的植物生理过程表明,多种胁迫因子之间存在复杂的非线性协同效应。DSSAT 所采用的简化耦合方式,可能系统性低估了这种复合胁迫对作物造成的实际伤害,导致在极端情景下对智能体决策效能的评估过于乐观。

并且 DSSAT 主要模拟作物的生理过程和最终产量,不包含生物力学模型,无法预测极端天气引发的植株倒伏、茎秆折断等物理性损伤。这使得 AgriAgent 的决策优化目标主要集中在生理产量上,而未能将作物的“结构性安全”作为显式的决策变量,可能带来实际生产中的潜在风险。

4 真实场景迁移性、系统鲁棒性与落地可行性分析

本研究构建的数字孪生环境的数据源、物理过程和执行机制均与真实的设施农业环境存在本质差异。因此,为确保研究成果具备实际落地价值和普适性,本章深入探讨“数字孪生-真实农场”的迁移挑战,并提出具体的工程和算法适配方案,为AgriAgent的未来实地部署奠定理论和技术基础。

4.1 传感器数据真实性和可靠性挑战

数字孪生评测平台通常假设传感器输入完全可靠。然而,在真实农业场景中,传感器信号受环境扰动、设备性能与通信条件等多种因素影响,数据不确定性显著提升。智能体策略能否在现实场景中保持鲁棒性,首先取决于其对复杂传感器数据的感知与处理能力。

综合分析,真实农场中的传感器数据主要面临以下3类挑战。

1) 数据噪声与误差:真实农田中的传感器数据(如土壤水分、温度、气体浓度等)易受环境干扰、硬件磨损及电源波动影响,产生短期高频噪声。这与本研究所使用的理想化无噪声环境形成显著对比。

2) 传感器漂移:受传感器元件老化、长期浸润或校准偏差影响,传感器读数会随时间出现系统性偏移。这种漂移会在长期运行中逐步积累,进而影响AgriAgent决策的稳定性与精确性。

3) 数据缺失:通信中断、电源故障或设备损坏会导致数据流出现间歇性中断或大片缺失。若缺失数据未被有效处理,可能造成环境状态估计不稳定,进而引发错误或空白决策。

4.1.1 鲁棒的数据预处理与状态估计

针对上述问题,AgriAgent的感知层需具备更强的数据完整性保障与异常抑制能力,超越基础的滑动均值或线性插补方法。本研究提出如下多层鲁棒感知策略。

1) 数据清洗与插补。针对大块缺失或长时间中断的情况,可采用基于深度学习的时序插补模型,如时间融合变换器(Temporal Fusion Transformer, TFT)^[27]或混合变换器(MixTransformer)^[28]。这些模型能够利用多传感器通道间的相关性与时间依赖关系,推断出更准确的缺失值,从而保证输入数据的连续性与可靠性。

2) 传感器漂移校正。为修正漂移与随机噪声,

可在感知层引入扩展卡尔曼滤波(Extended Kalman Filter, EKF)^[29]实时状态估计算法,以DSSAT作物生长模型作为状态演化方程,通过递归滤波动态校正传感器读数中的系统性误差与噪声扰动。

3) 异常值检测。对于短期极端异常值,可采用基于最大相关熵准则(Maximum Correntropy Criterion, MCC)^[30]的鲁棒回归模型或基于生成对抗网络(Generative Adversarial Network, GAN)^[31]的时序异常检测模型,有效识别并抑制极端异常对决策过程的干扰,从而提升数据的整体可信度。

4.1.2 仿真到真实的域适应训练

除数据层面的鲁棒性外,AgriAgent的策略层还需具备应对“仿真-现实差距(Sim-to-Real Gap)”的域适应能力。为增强模型在真实环境中的泛化性与稳定性,本研究提出以下策略。

1) 情景扩增(Scenario Augmentation)。在仿真训练阶段,通过在常规驱动序列中叠加零均值高斯噪声,以模拟传感器测量扰动与环境随机性,为模型预训练引入多样化干扰样本。

2) 域随机化(Domain Randomization, DR)。借鉴机器人学中的DR策略^[32],在数字孪生评测平台中随机化物理属性与不确定性参数,包括噪声类型、幅度、采样延迟与执行器响应时间等。通过“带噪仿真”环境下的策略优化,AgriAgent能在数据不完美、反馈不精确的条件下保持稳定决策,从而提升其向真实农场场景迁移的可行性与鲁棒性。

4.2 极端环境胁迫的系统鲁棒性与作物结构安全

4.2.1 作物结构性风险评估与监测

为弥补DSSAT在作物结构性损伤模拟上的不足,AgriAgent系统需在未来部署中通过模型互补的方式,集成对结构性风险的评估和实时监测。

1) 集成生物力学预警:农业生物力学是一个跨学科领域。这些专业知识可以构建到RAG知识库中。当智能体获取实时气象预警(强风、暴雨)时,可以检索相应的生物力学知识,进行倒伏风险的概率评估。

2) 实时结构监测:在真实农场部署中,应引入无人机(Unmanned Aerial Vehicle, UAV)或固定式视觉系统作为AgriAgent的“结构传感器”。先进的计算机视觉(Computer Vision, CV)^[33]算法,如基于卷积神经网络(Convolutional Neural Network, CNN)^[34]的模型,能够通过分析高分辨率图像,实

时检测作物的异常和倒伏现象,并实现高精度和低推理延迟的实时分类。视觉AI还能识别早期生理胁迫,为AgriAgent提供即时的结构状态反馈。

4.2.2 极端天气情景下的动态风险评估与控制调整

AgriAgent应具备元决策能力,能够根据环境的实时风险动态调整其优化目标。在极端天气预警情景下,LLM必须利用其推理能力和RAG知识库中的风险管理规程,将目标动态切换为“结构安全优先/最小化胁迫指数”。例如,针对强风/暴雨预警,智能体推理的首要任务是防止倒伏,可选择执行预防性适度灌溉,增加土壤密度和根土结合强度,增强作物的物理稳定性。

4.3 执行器容错与执行不确定性

在数字孪生环境中,执行器被理想化为“完美”的组件,但在真实农业生产场景中,执行器常受到通信延迟、环境扰动、机械磨损及能耗限制等因素影响,可能出现执行延迟、性能衰减甚至失效的情况,影响智能体的稳定性与控制精度。

为此,为确保AgriAgent在真实场景中的可落地性与鲁棒性,本研究提出了一套系统化的“闭环状态反馈-故障检测诊断-跨系统参数补偿”容错控制框架(Fault-tolerant Control, FTC)^[35]。

4.3.1 闭环状态反馈与执行延迟处理

在容错控制体系中,闭环反馈机制是实现自适应调节与动态修正的基础。AgriAgent的决策过程遵循“感知-决策-执行-验证”的强化闭环逻辑。系统需持续监测执行结果,并依据反馈信号判断实际效果与期望目标的偏差。与此同时,执行延迟是影响控制稳定性的重要因素。无论是LLM在推理生成控制指令的过程,还是物理执行器执行动作的过程,均存在固有的时间延迟。为提升系统的实时响应能力,AgriAgent采用异步通信协议,如消息队列遥测传输协议(Message Queuing Telemetry Transport, MQTT),来实现指令的并行传输与状态反馈。同时,LLM在决策推理阶段将延迟特性显式纳入建模过程,从而优化时序控制与系统调度。

4.3.2 故障检测、诊断与容错控制框架(FTC)

容错控制的核心在于实时识别故障类型、位置及程度,并据此快速采取补救措施。对于传感器故障,可通过数据冗余校验与模型预测偏差分析来识别异常;对于执行器故障,则可通过监测其电气参数变化,或借助流量计等辅助传感设备直接验证其

物理输出量来判断。

一旦检测到故障,AgriAgent将立即启动FTC策略。首先执行故障隔离,以防止故障扩散;随后,LLM核心将调用RAG模块中的故障知识库进行推理与策略重构,利用系统内剩余健康的执行器动态调整控制策略,从而维持系统的安全与稳定运行。

4.3.3 多因子耦合下的跨系统参数动态补偿

AgriAgent作为多工具增强型智能体,其优势也体现在多因子耦合场景下的跨系统补偿机制。当某一控制回路发生失效时,系统不会被动等待修复,而是通过LLM的知识推理主动进行软性补偿。具体而言,LLM智能体基于深度农学知识与实时环境数据,能够推断出作物受到的胁迫类型,并相应调整其他健康执行器的参数,以维持整体生态平衡和作物生理稳定性。

示例:灌溉执行器失效时的动态补偿。

1) 推理阶段:当灌溉系统发生故障,导致土壤含水量无法提升时,AgriAgent能识别作物正面临水分胁迫。

2) 补偿策略:此时,LLM决策不再单纯依赖灌溉系统,而是通过降低蒸腾速率以节约水分。例如,可适当调节通风与加温策略,或增加CO₂释放浓度,从而在不显著损失核心产量的前提下,缓解水分胁迫对作物生理过程的不利影响。

4.4 硬件成本、轻量化模型压缩与边缘计算部署

AgriAgent的落地可行性依赖于其经济效益和系统效率。LLM智能体的部署面临硬件成本和推理延迟的挑战。

4.4.1 硬件成本与模块化部署策略

智能农业系统的总体拥有成本(Total Cost of Ownership, TCO)是农户采纳的关键障碍。基于物联网(Internet of Things, IoT)^[36]和人工智能(Artificial Intelligence, AI)的系统部署成本为49.4~370.7美元/hm²。为了提高经济可行性,本研究建议采用模块化部署策略。农户可以根据预算逐步引入系统组件。例如,首先部署低成本传感器,随后逐步引入高级执行器和LLM边缘推理单元。通过精准农业技术优化投入品可以带来显著的成本节约。

4.4.2 轻量化模型压缩与边缘计算部署

AgriAgent的核心模型AgriAgent-7B参数量较大,推理速度较慢,难以直接部署在资源受限的农

场边缘设备上。农业环境控制要求实时或近实时的决策响应。

1) 边缘计算部署: 传统的云端LLM推理模式会导致较高的通信延迟。边缘计算的引入是解决这一矛盾的必然选择。通过将核心推理逻辑保留在本地, 边缘计算能够显著降低延迟, 确保时间敏感型控制指令的及时执行, 并允许智能体在网络连接不稳定时也能独立工作^[37]。

2) 轻量化模型压缩: 为实现边缘部署, 必须进行模型压缩。

a) 量化(Quantization)^[38]: 将模型权重和激活值从高精度的浮点数转换为低精度的整数。量化可以显著减小模型体积, 降低计算复杂度和内存需求, 同时大幅降低推理延迟和功耗。

b) 知识蒸馏(Knowledge Distillation, KD)^[39]: KD旨在利用AgriAgent-7B的丰富知识, 训练一个参数量更少、效率更高的“学生模型”。通过模仿教师模型的输出行为, 学生模型能够在大幅降低规模的同时, 保持核心决策精度。

5 结论与展望

本研究聚焦于工具增强型大语言模型在设施农业环境控制中的应用, 系统性地提出并验证了面向闭环决策的端到端智能体架构AgriAgent。通过理论分析、平台构建与多情景实验, 本研究得出以下核心结论:

1) 架构有效性得到验证。本研究提出的Agri-Agent架构, 通过“传感器-记忆库-检索器-大语言模型-工具执行器”的闭环流程, 成功实现了农业场景下“环境感知-动态决策-执行反馈”的端到端闭环控制。该架构有效弥合了LLMs与物理世界交互的鸿沟, 为解决农业决策中普遍存在的“幻觉”、工具调用受限及实时性不足等问题提供了可行技术路径。

2) 模型规模与决策性能正相关。实验结果表明, 智能体的决策质量与其参数量相关。以Qwen2.5_7B为核心的AgriAgent-7B模型在常规、扰动及极端3种气候情景下, 对玉米、小米、甜菜、番茄、卷心菜5种作物的产量均实现了提升(极端情景下提升幅度为125.40%~1537.46%), 展现出卓越的决策灵活性、环境适应性及应对复杂非线性胁迫的能力。

3) 工具增强有效抑制决策幻觉。消融实验证明, 通过集成实时数据获取工具(如天气API),

工具增强机制能降低决策过程中的加权幻觉率与高影响幻觉率, 并将幻觉率的降低直接转化为作物产量的有效提升, 验证了该机制对于保障农业决策可靠性的关键作用。

4) 系统展现了通用性与扩展潜力。AgriAgent在多种生理类型作物(C₄作物、耐旱作物、块根作物、茄果类、叶菜类)及多样化气候情景下均表现出一致的性能优势, 证明了该架构在设施农业环境控制中具有良好的通用性与可扩展性, 为智能农业系统的跨作物部署提供了创新范式。

展望未来, 本研究构建的智能体系统首先需完成从数字孪生环境向真实农场的迁移, 这包括在实地部署传感器网络以采集真实数据、优化模型, 并系统测试容错控制在长期运行中的可靠性, 以完成从仿真到现实的验证。其次, 需从商业价值与气候适应性两方面开展评估, 即通过量化经济收益证明其商业可行性, 并基于多年度特别是极端气候数据验证系统的稳定性与鲁棒性。最后, 面向广泛落地, 需探索轻量化部署与算法增强, 以适应资源受限及更复杂的生产条件。通过持续的技术迭代与实践, 该智能体系统有望从创新的技术架构, 发展为推动农业生产智能化、精准化与可持续发展的核心驱动力。

利益冲突声明: 本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献:

- [1] HUANG L, YU W J, MA W T, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions[EB/OL]. 2023: arXiv: 2311.05232, 2023.
- [2] XU Z W, JAIN S, KANKANHALLI M. Hallucination is inevitable: An innate limitation of large language models[EB/OL]. arXiv: 2401.11817, 2024.
- [3] SAMUEL D J, SKARGA-BANDUROVA I, SIKOLIA D, et al. AgroLLM: Connecting farmers and agricultural practices through large language models for enhanced knowledge transfer and practical application[EB/OL]. arXiv: 2503.04788, 2025.
- [4] ZHU H Y, QIN S, SU M, et al. Harnessing large vision and language models in agriculture: A review[EB/OL]. arXiv: 2407.19679, 2024.
- [5] GE C, ZHANG G J, WANG Y J, et al. Research status and development trends of artificial intelligence in smart agriculture[J]. Agriculture, 2025, 15(21): 2247.
- [6] SELVA KUMA S, KHAN A K M A, BANDAY I A, et al. Overcoming LLM challenges using RAG-driven precision in coffee leaf disease remediation[EB/OL]. arXiv: 2405.01310, 2024.
- [7] LI M H, ZHAO Y X, YU B W, et al. API-bank: A comprehensive benchmark for tool-augmented LLMs[EB/OL]. arXiv: 2304.08244, 2023.
- [8] QIN Y J, LIANG S H, YE Y N, et al. ToolLLM: Facilitating large language models to master 16000+ real-world APIs[EB/OL]. arXiv: 2307.16789, 2023.
- [9] YAO S Y, ZHAO J, Y U D, et al. ReAct: Synergizing reasoning

- and acting in language models[EB/OL]. arXiv: 2210.03629, 2022.
- [10] 王婷, 王娜, 崔运鹏, 等. 基于人工智能大模型技术的果蔬农技知识智能问答系统[J]. 智慧农业(中英文), 2023, 5(4): 105-116.
WANG T, WANG N, CUI Y P, et al. Agricultural technology knowledge intelligent question-answering system based on large language model[J]. Smart Agriculture, 2023, 5(4): 105-116.
- [11] CHEN X J, CHEN T Y, ZHAO J B, et al. AgriBERT: A joint entity relation extraction model based on agricultural text[M]// Knowledge Science, Engineering and Management. Singapore: Springer Nature Singapore, 2024: 254-266.
- [12] LIM B, ARİK S Ö, LOEFF N, et al. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting[J]. International Journal of Forecasting, 2021, 37(4): 1748-1764.
- [13] KANG H Q, LIU X Y. Deficiency of large language models in finance: An empirical examination of hallucination[EB/OL]. arXiv: 2311.15548, 2023.
- [14] WALLAT J, JATOWT A, ANAND A. Temporal blind spots in large language models[EB/OL]. arXiv: 2401.12078, 2024.
- [15] 冯在文, 许多, 田芳, 等. 设施农业智能决策大模型关键技术与发展方向[J]. 农业工程学报, 2025, 41(13): 50-61.
FENG Z W, XU D, TIAN F, et al. Key technologies and development trends of intelligent decision-making large models for facility agriculture[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(13): 50-61.
- [16] HALEVY A, DWIVEDI-YU J. Learnings from data integration for augmented language models[EB/OL]. arXiv: 2304.04576, 2023.
- [17] BALAGUER A, BENARA V, DE FREITAS CUNHA RL, et al. RAG vs fine-tuning: Pipelines, tradeoffs, and a case study on agriculture[EB/OL]. arXiv: 2401.08406, 2024.
- [18] WEI J, WANG X Z, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]// Proceedings of the 36th International Conference on Neural Information Processing Systems. New York, USA: ACM, 2022: 24824-24837.
- [19] WU Z H. AutoNo: A ReAct-based highly robust autonomous agent framework[EB/OL]. arXiv: 2504.04650, 2025.
- [20] CHEN R Q, DONG N Q, LIU W, et al. Empowering and assessing the utility of large language models in crop science[C]// Advances in Neural Information Processing Systems 37. Vancouver, Canada: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024: 52670-52722.
- [21] MEHRABI F, SEPASKHAH A R. Winter wheat yield and DSSAT model evaluation in a diverse semi-arid climate and agronomic practices[J]. International Journal of Plant Production, 2020, 14(2): 221-243.
- [22] ALDERMAN P D. A comprehensive R interface for the DSSAT cropping systems model[J]. Computers and Electronics in Agriculture, 2020, 172: 105325.
- [23] LING M, CHEN J, MÖLLER T, et al. Document domain randomization for deep learning document layout extraction[EB/OL]. arXiv: 2105.14931, 2021.
- [24] DEHGHAN MOROOZEH A, FARHADI BANSOULEH B, GHOBADI M, et al. Assessment of DSSAT and AquaCrop models to simulate soybean and maize yield under water stress conditions[J]. Spanish Journal of Agricultural Research, 2023, 21(3): e1201.
- [25] YANG A, YANG B, ZHANG B, et al. Qwen2.5 technical report [EB/OL]. arXiv: 2412.15115, 2025.
- [26] RAI N, DAEUN, CHOI, et al. Advancing site-specific disease and pest management in precision agriculture: From reasoning-driven foundation models to adaptive, feedback-based learning [EB/OL]. arXiv: 2510.24650, 2025.
- [27] CUI YT, SONG TH, WU GS, et al. MixFormerV2: Efficient fully transformer tracking[EB/OL]. arXiv: 2305.15896, 2023.
- [28] ARPAIA P, BUZIO M, DI CAPUA V, et al. Drift-free integration in inductive magnetic field measurements achieved by Kalman filtering[J]. Sensors, 2022, 22(1): 182.
- [29] BECKER P, FREYMUTH N, NEUMANN G. KalMamba: Towards efficient probabilistic state space models for RL under uncertainty[EB/OL]. arXiv: 2406.1513, 2024.
- [30] QIN Z, TAO J, YANG L, et al. Proportionate recursive maximum correntropy criterion adaptive filtering algorithms and their performance analysis[EB/OL]. arXiv: 2210.12311, 2022.
- [31] TOBIN J, FONG R, RAY A, et al. Domain randomization for transferring deep neural networks from simulation to the real world[EB/OL]. arXiv: 1703.06907, 2017.
- [32] SIAM SI, AHN H, LIU L, et al. Artificial intelligence of things: A survey[EB/OL]. arXiv: 2410.19998, 2024.
- [33] LIN F M. Vision language models: A survey of 26K papers[EB/OL]. arXiv: 2510.09586, 2025.
- [34] JUNAIDI A, HASHIM S Z M, OTHMAN M SBIN, et al. Deep learning and edge computing in agriculture: A comprehensive review of recent trends and innovations[J]. IEEE Access, 2025, 13: 137464-137490.
- [35] KIM D, LEE J D, BANG H, et al. Reinforcement learning-based fault-tolerant control for quadrotor with online transformer adaptation[EB/OL]. arXiv: 2505.08223, 2025.
- [36] KAPLAN J, MCCANDLISH S, HENIGHAN T, et al. Scaling laws for neural language models[EB/OL]. arXiv: 2001.08361, 2020.
- [37] GE J K, YU Z H, GONG X W, et al. Evaluation of irrigation modes for greenhouse drip irrigation tomatoes based on AquaCrop and DSSAT models[J]. Plants, 2023, 12(22): 3863.
- [38] ZHANG T, SHRIVASTAVA A. LeanQuant: Accurate and scalable large language model quantization with loss-error-aware Grid[EB/OL]. arXiv: 2407.10032, 2024.
- [39] FANG LY, YU XW, CAI JZ, et al. Knowledge distillation and dataset distillation of large language models: Emerging trends, challenges, and future directions[J]. Artificial Intelligence Review, 2025, 59(1): 17.

AgriAgent: End-to-End Large Model Agent System Architecture for Agricultural Environment Control

QIU Jiaying¹, LIU Yingchang¹, GAO Xingjie¹, HUANG Yuan², ZHANG Hongyu¹,
TIAN Fang^{1,3}, LI Wanli^{1*}, FENG Zaiwen^{1,3,4*}

(1. College of Informatics, Huazhong Agricultural University, Wuhan 430070, China; 2. College of Horticulture and Forestry, Huazhong Agricultural University, Wuhan 430070, China; 3. Yunnan Modern Agricultural Industry Research Institute, Kunming 650032, China; 4. Engineering Research Center of Agricultural Intelligent Technology, Ministry of Education, Wuhan 430070, China)

Abstract: [Objective] Large language models (LLMs) have demonstrated strong capabilities in natural language understanding, knowledge integration, and complex reasoning, offering new opportunities for intelligent decision-making in agriculture. However, their di-

rect application in agricultural production and facility environment control remains challenging due to strong physical constraints and high operational risks. The lack of real-world interaction and executable decision grounding limits the practical effectiveness of conventional LLMs in such scenarios. To address these challenges, a tool-augmented LLM-based agricultural intelligent agent system, termed AgriAgent, was proposed, and a digital-twin-based evaluation platform for agricultural decision-making was developed. By integrating a high-fidelity digital twin environment with an end-to-end agent architecture, the decision-making performance of agricultural intelligent agents with different parameter scales was systematically evaluated across multiple crops and climate scenarios. **[Methods]** A high-fidelity agricultural digital twin evaluation platform was constructed using the decision support system for agrotechnology transfer (DSSAT) v4.8 crop growth model as the core simulation engine to model crop growth under diverse environmental conditions and management strategies. Meteorological driving data were obtained from the Seoul Historical Weather Data dataset. Through data cleaning, missing-value imputation, unit normalization, and time-series reconstruction, the raw meteorological data were transformed into standardized inputs compatible with DSSAT. Three climate scenarios representing different environmental complexities were designed, including a regular scenario, a perturbed scenario, and an extreme scenario. The regular scenario employed historical observations, the perturbed scenario introduced stochastic disturbances to simulate short-term climate variability, and the extreme scenario incorporated multi-factor coupled stresses such as high temperatures and excessive precipitation during sensitive growth stages. In total, 90 annual climate driving sequences were generated. Fixed soil profile parameters calibrated by domain experts were applied across all simulations to minimize confounding effects. Within this digital twin environment, a tool-augmented agricultural intelligent agent, AgriAgent, was implemented using a modular architecture consisting of a sensor module, memory module, retriever, large language model, and tool executor, forming a closed-loop decision-making framework. In each decision cycle, the agent perceived environmental and crop state information, including soil moisture and nutrient status, meteorological conditions, crop growth stages, and stress indicators. State summaries and historical decisions were stored in memory, while agronomic knowledge was retrieved through a retrieval-augmented generation mechanism. Based on integrated information, the LLM generated structured environmental control commands in JSON format, which were validated and constrained by the tool executor before updating the DSSAT environment. The system supported irrigation, supplementary lighting, ventilation, heating, fertilization, and CO₂ enrichment. Five representative crops: maize, millet, sugar beet, tomato, and cabbage were simulated under the three climate scenarios over complete growing seasons, resulting in 450 crop-scenario combinations. An unmanaged DSSAT simulation served as the baseline. AgriAgent models with three parameter scales (1.5B, 3B, and 7B), built on the Qwen2.5 series, were evaluated. Crop economic yield expressed as dry matter at physiological maturity was adopted as the evaluation metric. **[Results and Discussions]** The results showed that AgriAgent consistently outperformed the baseline across all crops and climate scenarios, with model scale exerting a significant influence on decision-making performance. AgriAgent-7B achieved the best overall performance under regular, perturbed, and extreme scenarios, demonstrating strong generalization ability and environmental adaptability. By dynamically adjusting water, nutrient, light, and thermal management strategies, the agent effectively mitigated environmental stresses even under multi-factor coupled extreme climate conditions. Under extreme scenarios, AgriAgent-7B increased yields by 463.60% for maize, 351.20% for millet, 125.40% for sugar beet, 1 537.46% for tomato, and 1 185.14% for cabbage compared with the baseline. Particularly large gains were observed for high-value crops such as tomato and cabbage, highlighting the advantages of the proposed framework for precision-controlled facility agriculture. In contrast, AgriAgent-1.5B exhibited performance comparable to the baseline, while AgriAgent-3B achieved moderate improvements but remained inferior to the 7B model. These findings indicate a clear scaling effect, suggesting that larger models possess stronger capabilities in multi-source information integration, long-term temporal reasoning, and adaptation to complex environments. **[Conclusions]** This study developed a digital-twin-based agricultural decision evaluation platform and proposed a tool-augmented, end-to-end agricultural intelligent agent named AgriAgent. Experiments across multiple crops and climate scenarios verified the effectiveness and robustness of the proposed framework for dynamic agricultural decision-making. The results demonstrate that integrating knowledge retrieval, reasoning, and tool execution within a closed-loop LLM-based agent enables stable, reliable, and adaptive environmental control, providing a feasible technical pathway and standardized evaluation paradigm for intelligent agriculture.

Keywords: tool-augmented large language model; agricultural agent; digital twin; agricultural evaluation platform; DSSAT/crop growth model

Foundation items: Hubei Provincial Technology Innovation Plan Project (2024BBB055); Tibet Assistance Project of Department of Science and Technology of Hubei Province (2024EIA009); Yunnan Provincial Major Science and Technology Special Project Plan (202502AE090003); Fundamental Research Funds for the Chinese Central Universities (2662025XXPY005, 2662023XXQD002); Open Project of Key Laboratory of Knowledge Mining and Knowledge Service for Agricultural Integrated Publishing (2025KMKS01)

First author: 1. QIU Jiaying, E-mail: 19560533416@163.com; 2. LIU Yingchang, E-mail: liuyingchang@webmail.hzau.edu.cn

***Corresponding author:** 1. LI Wanli, E-mail: liwanli@mail.hzau.edu.cn; 2. FENG Zaiwen, E-mail: zaiwen.Feng@mail.hzau.edu.cn