doi:10.6041/j.issn.1000-1298.2026.22.010

融合 GraphRAG 与作物模型的极端气候下小麦生产决策智能体

周　昱[1,2]　刘义轩[1,3]　康　敏[1,4]　顾佳心[1,2]　王柯鑫[1,2]　张怡静[1,2]
刘　兵[1,4]　刘蕾蕾[1,4]　汤　亮[1,4]　曹卫星[1,4]　朱　艳[1,5]　肖浏骏[1,4]
(1. 南京农业大学智慧农业学院(人工智能学院), 南京 211800; 2. 国家信息农业工程技术中心, 南京 211800;
3. 智慧农业教育部工程研究中心, 南京 211800; 4. 农业农村部农作物系统分析与决策重点实验室, 南京 211800;
5. 江苏省农业科学院, 南京 210014)

摘要: 极端气候事件对冬小麦生产决策时效性和可靠性提出双重挑战,决策过程既需农业领域知识支撑,又需对生育进程等连续过程做出定量判断。通用大语言模型在辅助决策中表现出一定潜力,但仍存在领域知识覆盖不全、数值推理稳定性有限及复杂决策过程可追溯性不足等问题。为此,本文将图检索增强生成(GraphRAG)与作物生长模型 CropGrow 结合,构建了冬小麦极端气候辅助决策智能体。以农业知识图谱为证据来源,通过模型上下文协议(MCP)将 CropGrow 封装为数值计算服务,并由协调模块统一完成问题解析、证据检索、作物模型调度和结果校验,使知识证据与机理计算在同一决策流程中联动。在 450 组极端气候问答和 111 组田间生育期观测数据上的结果表明,与同基座 Qwen3 - Max 相比,专家核心实体加权召回率由 30.69% 提升至 37.15%,风险相关无支撑表达密度由 8.63 个/千字降至 5.49 个/千字;调度 CropGrow 后,开花期和成熟期预测均方根误差由 20.88 d 降至 5.79 d,决定系数由 0.027 升至 0.896。整体上,智能体的协同调度提高了冬小麦极端灾害情景下知识覆盖和生育期预测能力,为快速形成灾害应对方案提供了可实现的技术路径。

关键词: 大语言模型; 农业智能体; 图检索增强生成; 作物生长模型; 冬小麦; 极端气候
中图分类号: TP391.1; S126　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)22-0122-11　　**OSID**:

Integrating GraphRAG and Crop Models to Develop Decision-making Agent for Wheat Production under Extreme Climates

Zhou Yu[1,2]　Liu Yixuan[1,3]　Kang Min[1,4]　Gu Jiaxin[1,2]　Wang Kexin[1,2]　Zhang Yijing[1,2]
Liu Bing[1,4]　Liu Leilei[1,4]　Tang Liang[1,4]　Cao Weixing[1,4]　Zhu Yan[1,5]　Xiao Liujun[1,4]
(1. *College of Smart Agriculture* (*College of Artificial Intelligence*), *Nanjing Agricultural University*, *Nanjing* 211800, *China*
2. *National Engineering and Technology Center for Information Agriculture*, *Nanjing* 211800, *China*
3. *Engineering Research Center of Smart Agriculture*, *Ministry of Education*, *Nanjing* 211800, *China*
4. *Key Laboratory for Crop System Analysis and Decision Making*, *Ministry of Agriculture and Rural Affairs*, *Nanjing* 211800, *China*
5. *Jiangsu Academy of Agricultural Sciences*, *Nanjing* 210014, *China*)

Abstract: Extreme climate events pose dual challenges to the timeliness and reliability of decision-making in winter wheat production. Effective decision-making requires not only support from domain-specific agricultural knowledge but also quantitative assessment of continuous processes such as crop phenological development. Although general-purpose large language models (LLMs) have demonstrated considerable potential for decision support, their application remains constrained by insufficient coverage of domain knowledge, limited robustness in numerical reasoning, and inadequate traceability in complex decision-making processes. To address these limitations, integrateing graph retrieval-augmented generation (GraphRAG) with the crop model (CropGrow) to develop an intelligent decision-support

收稿日期: 2026-08-14　修回日期: 2026-08-28
基金项目: 国家重点研发计划项目(2024YFD2301300)、国家自然科学基金面上项目(32471995)和江苏省优秀青年基金项目(BK20250198)
作者简介: 周昱(2003—),男,博士生,主要从事智慧农业研究,E-mail: zhouyunjau@stu.njau.edu.cn
通信作者: 肖浏骏(1988—),男,副教授,博士,主要从事智慧农业研究,E-mail: liujunxiao@njau.edu.cn

agent for winter wheat production under extreme climate events. The agent employed an agricultural knowledge graph as its evidence base and encapsulates CropGrow as a numerical computation service through the model context protocol (MCP). A coordination module orchestrated query interpretation, evidence retrieval, crop model invocation, and result validation, thereby enabling knowledge-based evidence and process-based mechanistic computation to operate synergistically within a unified decision-making workflow. Evaluation using 450 extreme-climate question-answer pairs and 111 sets of field-observed phenological data showed that, compared with the same backbone model, Qwen3 - Max, the expert-weighted recall of core entities was increased from 30.69% to 37.15%, while the density of unsupported risk-related statements was decreased from 8.63 to 5.49 per 1 000 words. After integrating CropGrow, the root mean square error (RMSE) for flowering and maturity date predictions was decreased from 20.88 days to 5.79 days, while the coefficient of determination (R^2) was increased from 0.027 to 0.896. Overall, the coordinated orchestration of the proposed agent improved both domain knowledge coverage and phenological prediction performance under extreme climate disaster scenarios, providing a feasible technical framework for the rapid formulation of disaster-response strategies in winter wheat production.

Key words: large language model; agricultural agent; graph retrieval-augmented generation (GraphRAG); crop growth model; winter wheat; extreme climate

0　引言

全球气候变化背景下,极端高温、霜冻、阶段性干旱及持续涝渍等灾害事件发生频率和影响范围均呈增加趋势,对作物生产构成严重威胁[1-3]。与平均气候态的缓慢演变不同,极端天气事件突发性强、持续时间不确定且区域差异显著,极易扰乱作物生育进程、破坏群体结构协调性,并导致产量与品质协同下降,进而威胁粮食生产稳定性[4-6]。冬小麦生长发育对气候具有明显的阶段性响应,不同生育阶段对气象灾害的敏感性存在差异[7]。因此,极端气候条件下辅助决策不仅需识别灾害类型和发生阶段,还需结合区域环境、作物生育进程和田间条件提供可追溯的机理证据、风险阈值与针对性措施。

农业生产决策具有较强的知识依赖性,面向智能农业生产与可持续管理的问答系统需对农业领域信息进行有效检索、组织与利用[8]。大语言模型逐渐应用于农业问答、种植咨询、病虫害诊断和农事管理等任务[9-13]。在美国内布拉斯加大学林肯分校 TAPS 项目开展的玉米生产试验中,ChatGPT - 4o 参与田间管理决策,AI 辅助管理地块的产量在全部 31 个参赛地块中排名第 8,为生成式人工智能参与实际农场管理决策提供了实证案例[11]。

但通用大语言模型在农业专业知识覆盖、事实可靠性和幻觉控制等方面仍存在局限[10,12,14]。检索增强生成(RAG)可通过外部知识补充模型上下文[15],农业知识图谱则能够以实体和关系形式组织作物、生育阶段、环境胁迫、受害机制和调控措施等领域知识[16-18]。在此基础上,知识图谱增强检索生成方法可利用图结构关系扩展检索上下文,增强相关知识之间的关联组织[19]。但知识检索主要解决领域知识获取与组织问题,对逐日环境驱动下连续过程计算仍缺乏显式机理约束。

生育期预测、播期计算及灾害影响分析等农业任务具有明显的连续过程特征。CropGrow 过程模型能够依据结构化品种、环境和管理输入开展作物生长发育及产量形成模拟[20-22];而 DSSAT 和 APSIM 则对农作系统模型输入准备、参数组织和结果解释具有一定专业门槛[23-24]。大语言模型在自然语言理解、任务分解和结果组织方面具有优势,而作物模型可为连续数值过程提供机理约束,两者在农业辅助决策中具有较强互补性。模型上下文协议(MCP)以及任务规划、工具调用等智能体技术的发展,为大语言模型调用农业知识库、气象数据和作物模型提供了实现基础[25-29]。

农业领域已开始探索大语言模型智能体,如对话式植物表型分析多智能体系统[30],多智能体 AI 研究也从协同、动态编排与工作流等角度总结了系统结构与关键问题[31];但现有工作普遍缺乏领域证据与机理模型的耦合验证,尤其缺乏针对极端气候条件下辅助决策智能体研究。本文聚焦南京农业大学“慧耕耘”智能体中衔接知识服务与生长预测的知识-机理协同技术,以冬小麦极端气候为场景,探索 GraphRAG 知识引擎与 CropGrow 量化引擎的融合机制,通过冬小麦极端气候问答评价其领域知识覆盖、因果链组织和风险无支撑表达情况,并利用田间生育期观测数据评价机理模型接入后的连续数值任务性能,在此基础上以播期推荐案例验证知识查询、参数组织、模型调用和结果生成之间的流程衔

接,以期为农业智能体由知识问答向机理约束下辅助决策拓展提供方法基础。

1 材料与方法

1.1 研究对象与系统架构

面向冬小麦极端气候知识服务、生育期预测和播期推荐任务,构建由协调智能体、知识引擎和量化引擎组成的知识-机理协同框架。协调智能体统一采用 Qwen3 - Max 作为基模型,负责意图识别、任务分解、工具选择、执行状态管理和结果组织;知识引擎基于 GraphRAG 完成实体、关系、文本单元、社区报告和原始证据检索;量化引擎通过 MCP 封装 CropGrow、气象数据接口和田间数据接口,用于冬小麦生育期预测及播期计算。为保证智能体基模型对照的公平性,各功能模块以标准化接口接入协调智能体,不对基座模型或检索模型进行独立预训练和微调。

系统接收用户请求后,首先识别作物、区域、生育阶段、灾害类型和任务目标,并根据任务需求调用知识检索、气象数据获取、参数组织和机理计算等工具。工具执行结果返回后,由协调智能体进行证据与量化结果的一致性检查,并形成包含计算结果、判断依据、适用条件和风险提示的最终输出。为避免系统持续迭代产生版本差异,所有评价均基于同一阶段的知识图谱快照、模型回答及 CropGrow 模拟结果。系统总体架构如图 1 所示。

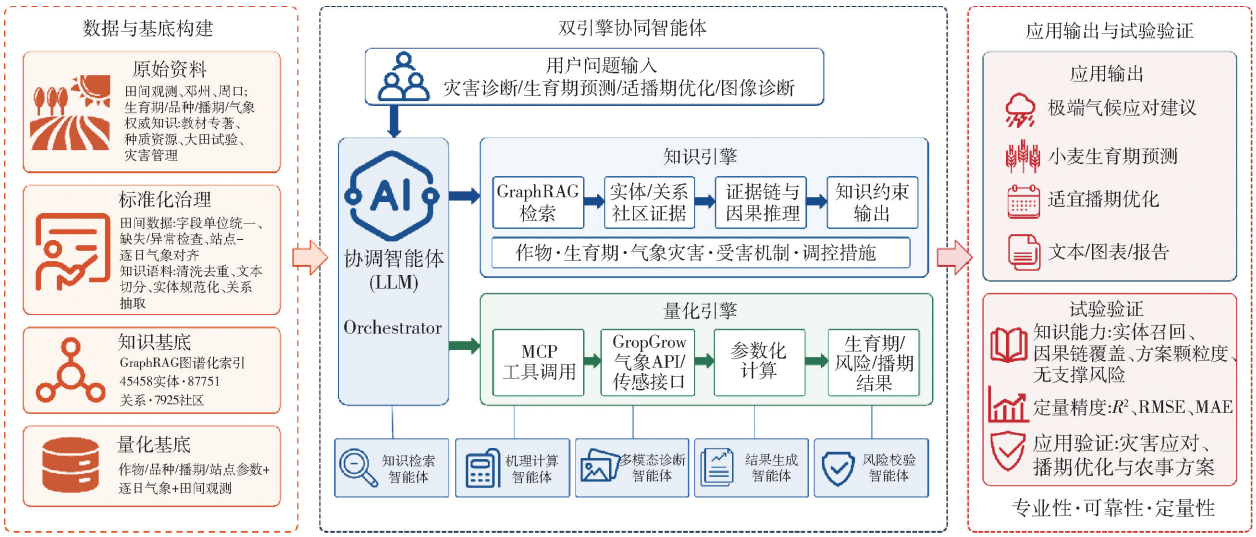


图 1 融合知识引擎与量化引擎的智能体总体架构

Fig. 1 Overall architecture of agent integrating knowledge and quantitative engines

1.2 研究数据来源

研究数据包括知识语料、田间生育期观测数据和气象数据。知识语料由 282 本农业教材与专著,以及种质资源、大田试验和农业气象灾害管理等相关资料组成,经人工筛选和清洗后用于知识抽取、实体关系组织及农业知识图谱构建。

田间生育期数据于 2025 年在河南省邓州市智慧农业中心获取。试验设置郑麦 1860、郑麦 136、周麦 36、周麦 37 和百农 307 共 5 个冬小麦品种,以及 N0、N90、N180、N270 和 N360 共 5 个施氮水平,施氮量分别为 0、90、180、270、360 kg/hm^2,其余田间管理措施保持一致。每个品种-施氮组合设置 5 次重复,共 125 个试验小区,并逐小区记录开花期和成熟期。经生育期观测记录及对应气象序列完整性检查,共 111 组样本具有完整的开花期、成熟期观测和逐日气象序列,纳入生育期预测评价,其余 14 组数据因记录不完整被排除。

气象数据来源于中国地面气候资料日值数据集(V3.0)。使用的气候要素包括日平均气温、日最高气温、日最低气温、日照时数和日降水量。

1.3 农业知识库、GraphRAG 知识引擎与证据检索

知识引擎为智能体生成结果提供可追溯的农业领域证据。知识资源覆盖水稻、小麦及其他作物,评价范围主要包括冬小麦生育阶段、气象灾害、环境驱动、受害机制、阈值条件和调控措施。原始语料经清洗、去重、文本切分和实体规范化后,在农业本体约束下抽取作物、生育期、气象灾害、环境因子、受害机制和调控措施等实体及其关系,并进行关系融合和图谱构建[17]。在 GraphRAG 流程中,根据知识图谱中实体关系建立图结构索引,并采用 Leiden 算法[32]进行主题社区划分和社区报告生成。查询阶段,根据用户问题识别相关实体,并检索与其关联的关系、文本单元和原始证据;对于涉及多个知识主题的问题,可进一步结合社区信息扩展检索上下文。检索结果以结构化形式返回协调智能体,用于约束后回答生成。

当前知识图谱快照包含 45 458 个实体、87 751 条关系和 7 925 个社区，按构建流程所用分词器统计语料规模为 3 920 275 token。实体与关系类型分布见表 1 和表 2，其中“其他未归并”关系占 17.83%。

表 1　知识图谱实体类型统计

Tab. 1　Statistics on entity types in knowledge graph

相关实体组	实体数量	总频次	平均频次	占全部实体比例/%	代表实体
生育期器官与产量性状	8 876	16 962	1.91	19.53	叶片；根系；分生孢子；光合作用；拔节期；籽粒；冬孢子；籽粒产量
栽培管理与工程调控	6 603	10 932	1.66	14.53	轮作；肥料；复种；茎叶喷雾；药剂防治；喷雾；灌溉；间混套作
病虫草害与防控投入品	6 509	15 519	2.38	14.32	除草剂；杂草；乙草胺；草甘膦；稗草；百草枯；蚜虫；莠去津
区域设备与支撑主体	6 285	13 346	2.12	13.83	中国；青藏高原；长江流域；东北地区；南方；北方
作物品种与种质	4 935	14 011	2.84	10.86	小麦；玉米；水稻；大豆；棉花；冬小麦；春小麦；油菜
气象灾害与环境胁迫	4 451	8 475	1.90	9.79	温度；干旱；药害；湿害；冻害；水分；气温；气候变化
其他	4 137	5 715	1.38	9.10	土壤；产量；作物；栽培措施；病菌；药剂；磷肥；禾本科杂草
土壤肥水与地力	3 662	7 809	2.13	8.06	氮肥；钾；有机肥料；磷；氮素；土壤有机质；氮

表 2　知识图谱关系类型统计

Tab. 2　Statistics on relationship types in knowledge graph

语义关系类型	数量	平均权重	占全部关系比例/%
危害影响	18 512	9.61	21.10
其他未归并	15 646	7.59	17.83
防治控制	13 146	8.28	14.98
栽培管理与施用	8 552	7.86	9.75
分类归属	7 648	8.41	8.72
气象环境作用	6 483	7.41	7.39
促进提高	5 176	7.74	5.90
生育阶段与时间	4 565	7.47	5.20
地区分布与适宜性	2 393	7.58	2.73
组成成分与来源	2 208	7.74	2.52
监测评价与试验	2 184	7.38	2.49
机构文献与生产登记	868	7.70	0.99
互作关联	370	7.25	0.42

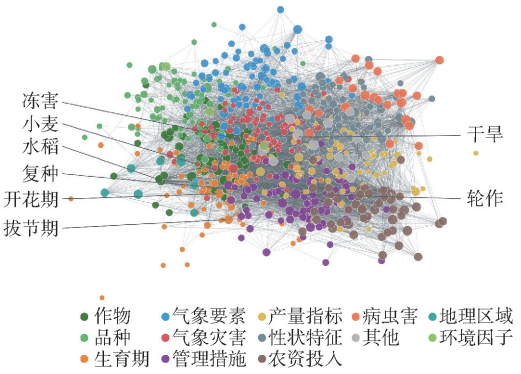


图 2　GraphRAG 构建的农业知识图谱示例（部分）

Fig. 2　Example of an agricultural knowledge graph built with GraphRAG (partial)

在灾害知识组织中，图谱不仅提供实体名称，还连接气象条件、生理响应和调控措施。以冬小麦扬花期高温热害为例，系统可沿“作物–生育期–灾害条件–生理损伤–产量风险–应对措施”路径检索相关实体与原文证据，用于约束后续回答。该路径表示知识组织和检索过程，不等同于田间管理效果验证。知识图谱局部结构示例见图 2。

在灾害推理中，图谱不仅提供实体名称，还提供从气象条件到生理响应再到调控方案的因果链。查询阶段，GraphRAG 将用户问题编码至向量空间，从实体、文本单元和社区报告等索引中按余弦相似度检索相关证据。

根据相似度定位问题中作物、生育期和灾害等实体最相关的几个节点，扩展其邻域关系，形成局部知识上下文。对于需要跨主题综合的问题，可进一步调用社区报告进行全局信息汇总。最终，检索得到的实体、关系、原文证据和社区摘要共同构成大语言模型生成回答的上下文。

以黄淮海地区冬小麦低温与涝渍问题为例，展示 GraphRAG 返回的内部实体、关系标识及回答组织形式。该示例只说明回链至图谱元素的过程，内部编号不等同于原始文献页码，也不作为图谱准确率或减灾效果的独立验证。

智能体回复如下（其中［Data：Relationships（…）；Entities（…）］是 GraphRAG 内部实体和关系编号）：

黄淮海地区冬小麦在越冬期及返青拔节期可能遭遇低温冻害与田间渍水风险，2 类灾害可先后发生或在特定天气过程中叠加，需结合生育阶段、土壤墒情和田间积水状态分支处置。防控顺序应为先识别现场主要限制因子，再组织防冻、排涝与灾后恢复措施。

防冻管理应根据苗情和土壤墒情实施。在墒情适宜且无田间积水时，可结合当地栽培规程评估灌水、中耕镇压等措施，以改善近地层小气候和增强麦苗抗寒性［Data：Relationships（44515，30100，

30111)；Entities（25546，17596，17420）]。同时，可通过选用抗冻品种和适期播种培育壮苗，降低弱苗受冻风险［Data：Entities（26079，18104）；Relationships（31249）]。寒潮前应避免无条件追肥，防止徒长削弱抗性［Data：Entities（17420）；Relationships（29709）]。

涝渍防控以“排”为主。降水偏多导致田间积水时，应通过开沟排水、疏浚沟渠等措施减轻根系缺氧；灌浆至成熟期的持续渍水还可能影响籽粒灌浆并降低穗粒重［Data：Entities（19768，13259）；Relationships（34653，34667，22173）]。由于冻害和涝渍的管理目标不同，涉及灌水的防冻措施只能在田间无积水且土壤墒情适宜时讨论。

灾后恢复应先判断积水、土壤墒情与受害程度。田间有积水时先排涝；排水完成且墒情允许时，再结合苗情评估灌水或追施速效氮肥等恢复措施［Data：Relationships（29762，29763，29764）；Entities（17435）]。对具体植物生长调节剂的使用，若缺少剂量、适用时期和区域登记依据，则只作为候选证据提示，不生成可直接执行的用药处方。

综上，黄淮海地区冬小麦的低温与涝渍复合风险应遵循“现场诊断—优先排除主要限制因子—分阶段恢复”的原则，并将图谱证据、当地栽培规程与现场墒情共同作为措施边界。

具体知识图谱推理路径如图 3 所示。



图 3 农业知识图谱推理路径示例

Fig. 3 Example of agricultural knowledge graph reasoning path

1.4 CropGrow 量化引擎与 MCP 调用

量化引擎采用作物生长模型 CropGrow[20] 完成连续环境驱动下的作物发育计算。为实现智能体对模型和数据服务的标准化调用，通过 MCP 将 CropGrow、气象数据接口、田间数据接口和播期算法封装为工具服务。协调智能体生成参数请求后，MCP 服务器依次执行字段校验、模型调用和结果返回。该结构化调用方式与基础模型工具学习中的任务理解、工具选择、参数生成和工具执行等基本范式相一致[25]。

系统以 JSON－RPC 格式传递作物、品种参数、播期、站点、逐日气象序列和目标指标。CropGrow 依据温度、辐射和水分等环境因子计算发育进程，输出开花期和成熟期等结果；上层模块仅负责语义解析、字段映射和结果组织，不改变 CropGrow 的计算机理。

1.5 角色分工与任务编排

系统采用流程化角色分工和工具编排方式实现不同功能模块之间的协同。协调智能体负责识别用户任务并决定是否调用 GraphRAG、CropGrow 及其他工具；知识检索角色返回相关实体、关系、文本单元和原始证据；机理计算角色负责调用 CropGrow 和气象接口；各角色通过预设结构化接口交换结果，并最终由协调智能体完成汇总与输出[26－28]。该架构将知识检索、参数组织、机理计算和结果生成分隔为可记录步骤，用于追溯证据来源、输入参数和工具返回结果。

1.6 对照设置、评价指标与统计分析

评价包括不同大语言模型之间描述性比较和同基座模型条件下系统级比较。对比模型包括构建的智能体系统（Our－Agent）、Doubao－Seed、DeepSeek－V3 和 Qwen3－Max，测试问题覆盖霜冻机理、低温涝渍、生育期灾害、高温干旱和干热风等 9 类冬小麦极端气候相关任务。共设置 450 组相互独立的问题，且所有问答均相互独立，不存在上下文连续传递。对照组不调用知识检索、机理模型、任务编排或风险校验，因此该比较反映本智能体系统相对于同基座模型的总体差异。评价任务与专家实体知识角色对应关系见图 4。

1.6.1 定量预测精度评价

生育期预测评价以 111 组有效田间样本开花期和成熟期观测结果为依据，每组样本包含 2 个生育时期。采用决定系数 R^2 和均方根误差（RMSE）评价预测结果与田间观测之间的总体拟合程度及日期误差。

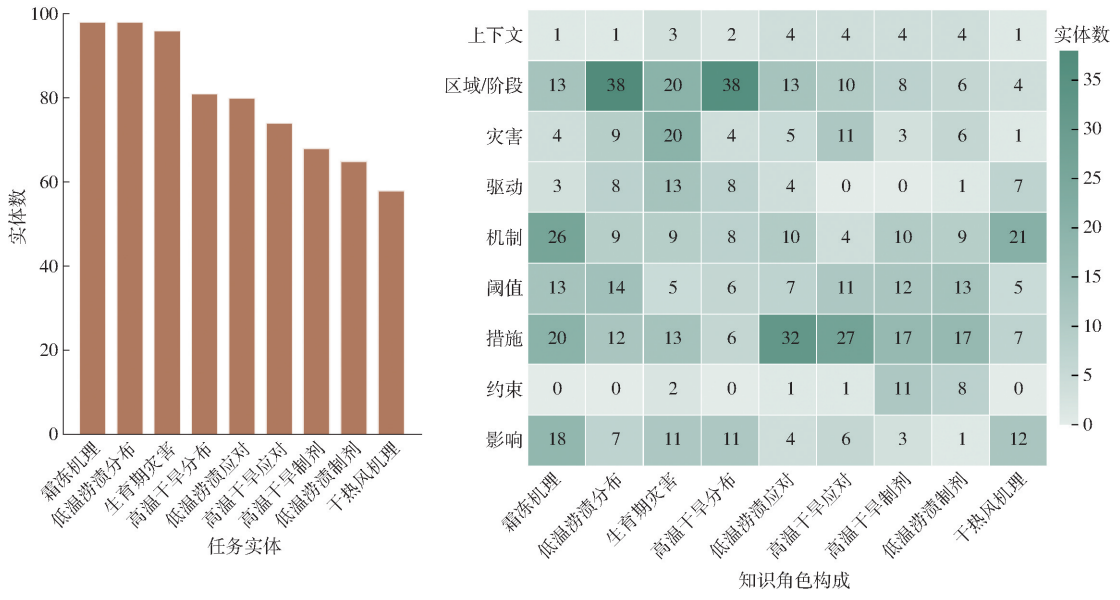


图 4　专业实体表任务知识角色组成

Fig. 4　Task-by-knowledge-role composition of expert entity table

1.6.2　基于专家实体与裁判模型的自动化评估

针对冬小麦极端气候问题构建专家实体表，并根据知识作用将实体划分为上下文、区域/阶段、灾害、驱动、机制、阈值、措施、约束和影响等类别，并在专家实体表中预先赋权，各模型共用同一权重。模型回答经分词、同义词归一化和正则匹配后与专家实体表比对，计算专业核心、专业主体、专业加共识和路径要点 4 类关键实体加权召回率。

回答可执行性按照方案颗粒度划分为 3 个等级：等级 1 表示以原则性描述为主、缺乏具体操作；等级 2 表示措施方向正确但缺少定量参数；等级 3 表示包含明确剂量、时期、浓度、频次或环境阈值等可执行信息，并统计不同等级回答比例。

因果链评价以专家实体表中“灾害–驱动–机制–影响–措施”等关键角色序列为参考，将回答中与参考角色及关系方向一致的链定义为匹配预期因果链，其余计为未匹配扩展链。因果链匹配率为匹配预期因果链数量占识别链总数的比例。未匹配扩展链可能属于合理知识拓展，也可能包含缺乏证据支持的推断，因此需结合证据检索结果进行解释。

无支撑风险密度用于补充召回率评价。评价规则将不能与专家实体表、知识图谱或题目输入证据建立匹配的风险敏感实体或表达单元计为潜在风险无支撑知识单元，并按回答字符数标准化为“个/千字”。该指标本质上是与内部证据资源一致性的风险代理量，其结果受图谱覆盖、规则识别和回答长度影响，不等同于经农业专家确认的事实错误率或幻觉率。计算式为

$$HRD_m = 1\,000\,\frac{Nu_m}{L_m}$$

式中　HRD_m——模型 m 无支撑风险密度

Nu_m——规则识别潜在无支撑风险单元数

L_m——模型 m 输出总字符数

乘以 1 000 为“个/千字”，HRD_m 越低表明单位长度文本中与内部证据资源未匹配的风险单元较少。

自动化评价在模型名称匿名化后，按统一标准评价专业性、完整性、可操作性和准确性，并统计每千字的风险敏感无支撑知识单元数。原试验以 GPT－5.2 作为自动裁判，评分可能受裁判模型偏好影响；因此，将其定位为探索性辅助证据，主要结论优先依据规则化实体匹配、同问题配对统计和 CropGrow 观测对比，后续仍需由异构裁判与农业专家盲评复核。

2　结果与分析

2.1　领域知识覆盖度与响应颗粒度分析

不同模型关键实体加权召回率如图 5 所示。在相同专家实体表和匹配规则下，本智能体系统在专业核心、专业主体、专业加共识和路径要点 4 类指标上加权召回率均高于同基座模型。其中，本智能体专家核心实体加权召回率为 37.15%，Qwen3－Max 为 30.69%，DeepSeek－V3 为 28.00%。结果表明，在既定评价范围内，本智能体够覆盖更多预设的冬小麦极端气候领域知识。由于该指标受到回答长度、实体表范围和权重设置影响，其结果主要用于反映知识覆盖程度。

按照预设方案颗粒度评价标准进行了评估，本智能体系统等级 3 回答比例为 89.3%，明显高于

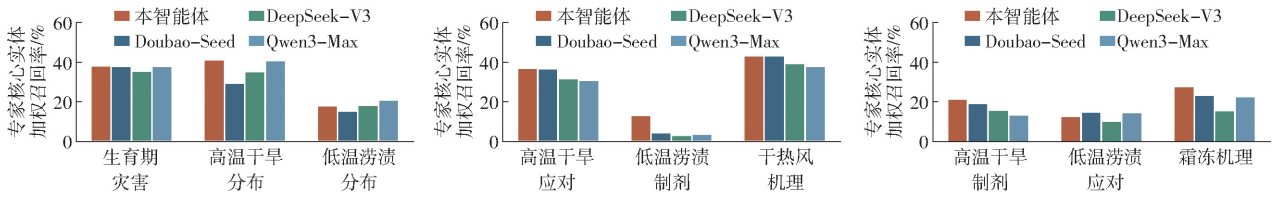


图 5 不同模型关键实体加权召回率

Fig. 5 Weighted key-entity recall among models

DeepSeek－V3 和 Qwen3－Max，但低于 Doubao－Seed 的 98.2%。综合看来，本智能体系统能够提供更多包含时期、剂量、阈值等具体参数的信息，但平均回答长度也增加。颗粒度等级结果见图 6。

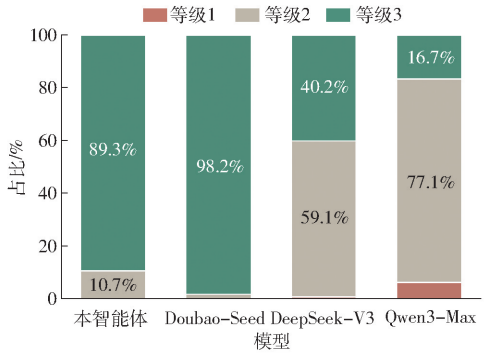


图 6 不同模型回复颗粒度等级

Fig. 6 Response-granularity levels among models

2.2 因果推理链覆盖与无支撑表达风险代理分析

农业灾害知识服务不仅要求命中实体，还需将实体组织为可追溯的因果链。本智能体平均匹配预期因果链为 23.44 条，DeepSeek－V3 和 Qwen3－Max 分别为 11.80 条和 11.21 条，表明本智能体系统在回答中组织了更多与专家预设结构一致的农业灾害知识链。本智能体系统因果链匹配率为 57.7%，略低于 Qwen3－Max 的 58.8%，表明本智能体在增加匹配预期因果链数量的同时，也生成了更多扩展性关系。未匹配扩展链并不必然表示错误，需要结合检索证据和风险无支撑知识单元共同评价。结果见图 7。

风险无支撑知识单元评价结果显示，本智能体的风险无支撑知识单元密度低于 DeepSeek－V3 和 Qwen3－Max，表明在既定内部证据资源和识别规则下，本智能体系统单位长度文本中无法获得证据匹配的风险相关表达较少。结果见图 8。

2.3 生育期预测定量精度评估

为评价智能体处理连续数值任务能力，调用 CropGrow 对 111 组冬小麦样本的开花期和成熟期进行模拟，并与田间观测结果进行比较，结果见图 9。

由图 9 可知，111 组有效样本的开花期与成熟期评价结果显示，本智能体系统调用 CropGrow 后的 RMSE 为 5.79 d、R^2 为 0.896；同基座 Qwen3－Max

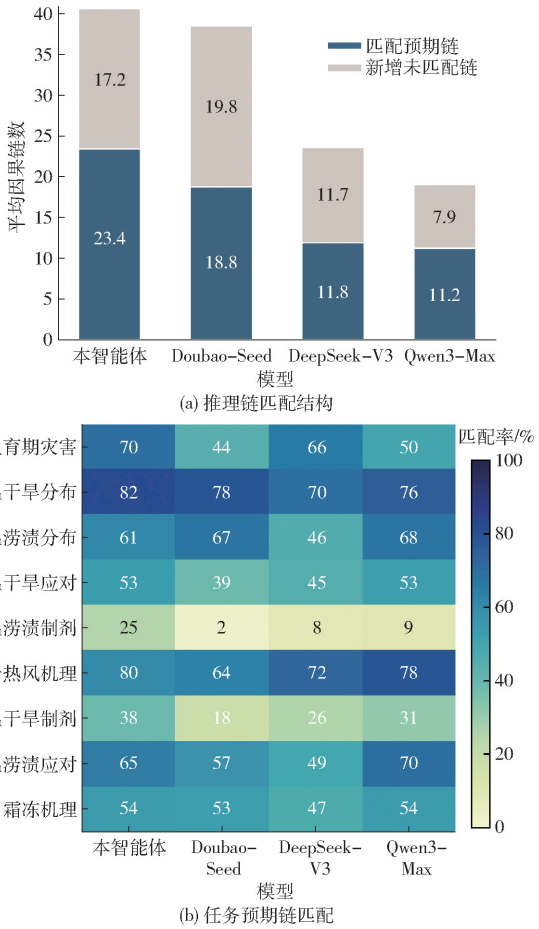


图 7 因果链覆盖与未匹配扩展链分析

Fig. 7 Analysis of causal-chain coverage and unmatched extensions

模型 RMSE 为 20.88 d，R^2 为 0.027。在本研究站点、年度及样本范围内，接入 CropGrow 后生育期日期预测误差明显降低，合并数据的总体拟合程度提高。

2.4 冬小麦播期推荐流程案例

为验证知识检索、参数组织和计算工具之间的流程衔接能力，以 2025 年江苏省徐州市冬小麦播期推荐为案例。案例输入包括目标区域、年份、品种类型、冬前积温需求和越冬安全条件。系统首先查询区域与品种相关知识，并将形成冬前壮苗所需的大于等于 0℃积温作为计算约束；随后调用气象数据和播期计算方法，分别采用叶龄积温法和温度趋势法获得播期范围，并对结果进行综合。具体过程见表 3。

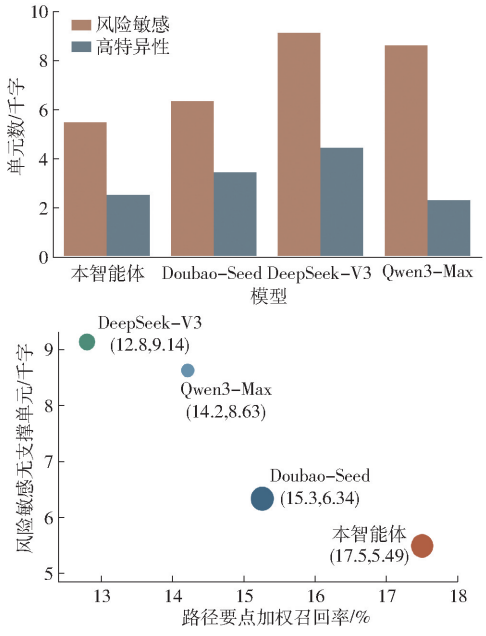


图 8　领域知识覆盖与无支撑表达密度代理分析

Fig. 8　Proxy analysis of knowledge coverage and unsupported-expression density

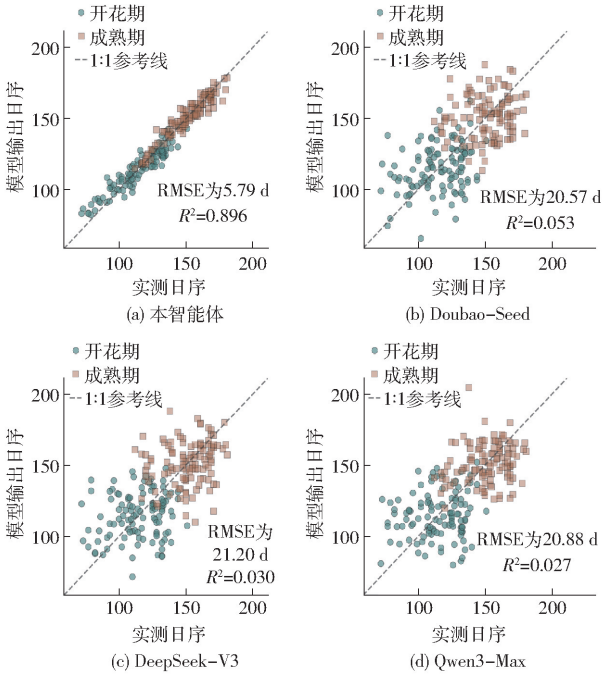


图 9　不同模型的冬小麦生育期预测结果

Fig. 9　Winter wheat phenology predictions among models

由表 3 可知，叶龄积温法计算得到播期为 10 月 11—16 日，温度趋势法得到播期为 10 月 9—27 日，系统综合形成 10 月 11—16 日推荐窗口。该案例表明系统已实现区域识别、知识查询、气象参数组织、播期计算及结果生成之间的流程衔接。由于未设置同期田间播期对照试验，该结果仅用于验证计算流程，不作为最适播期、增产或减灾效果的田间验证结论。

表 3　冬小麦播期推荐流程案例

Tab. 3　Workflow case of sowing-date recommendation for winter wheat

Round 1：知识查询 Round 2：参数核对	1、综合分析依据 目标区域与年份 地区：江苏省徐州市（属黄淮冬麦区，典型半冬性小麦种植区） 年份：2025 年 关键科学指标——积温需求 根据作物机理与栅培知识库资料，冬小麦从播种到越冬、形成主茎 5 叶 1 心壮苗所需大于等于 0℃ 积温为 495 ~ 570℃·d。该范围作为本案例的计算约束 品种类型参考 徐州地区主栽品种多为半冬性小麦，播期需兼顾避免过早导致冬前旺长和过晚导致积温不足 2、2025 年徐州冬小麦播期计算 基于输入气象数据，采用叶龄积温法与温度趋势法交叉计算，形成推荐窗口		
Round 3：播期计算	方法	推荐播种期	说明
	叶龄积温法	10 月 11—16 日	与本案例设定的冬前积温需求匹配，作为主要计算参考
	温度趋势法	10 月 9—27 日	给出较宽的调整范围；是否采用需结合当年温度距平和预报。本案例未设置异常天气独立判据
Round 4：结果组织	基于设定参数的推荐窗口：2025 年徐州市冬小麦 10 月 11—16 日播种。该窗口在理论上有利于冬前形成约 5 叶 1 心壮苗并改善越冬基础，但未设置同期田间对照，不表述为已验证的“最适”播期或保证性结论		

2.5　本智能体系统与基座模型性能

本智能体系统与 Qwen3 - Max 主要评价结果见表 4。在 450 组同问题配对回答中，本智能体系统相对 Qwen3 - Max 在专业核心、专业主体、专业加共识和路径要点加权召回率分别由 30.69% 和 15.43% 提高至 37.15% 和 18.84%。等级 3 回答比例由 16.67% 提高至 89.33%，平均匹配预期因果链数由 11.21 条增至 23.44 条，但因果链匹配率为 57.67%，略低于 Qwen3 - Max 的 58.79%；表明匹配链绝对数量增加的同时，未匹配扩展链也在增加。风险无支撑知识单元密度由 8.63 个/千字降至 5.49 个/千字；生育期两时期预测误差 RMSE 由

20.88 d 降至 5.79 d。与此同时,本智能体系统平均回答长度由 1 115 字增至 2 813 字,而路径要点命中密度由 7.13 条/千字降至 3.74 条/千字。

表 4 本智能体系统与同基座模型性能

Tab.4 Performance of complete agent system and the same base model

参数	本智能体	Qwen3 - Max
专业核心加权召回率/%	37.15	30.69
路径要点加权召回率/%	18.84	15.43
等级 3 可执行回答比例/%	89.33	16.67
平均匹配预期因果链数/条	23.44	11.21
因果链匹配率/%	57.67	58.79
风险无支撑知识单元密度/(个·千字$^{-1}$)	5.49	8.63
平均回答长度/字	2 813	1 115
路径要点命中密度/(条·千字$^{-1}$)	3.74	7.13
生育期预测误差 RMSE/d	5.79	20.88
生育期合并评价 R^2	0.896	0.027

3 讨论

3.1 知识证据约束与农业问答质量

冬小麦极端气候辅助决策涉及生育阶段、环境胁迫、生理响应、产量影响和调控措施等多层次知识。采用 GraphRAG 将相关实体、关系和原始证据组织为检索上下文,使语言模型能够依据图结构中关联信息形成较完整的农业灾害知识链。已有研究构建了面向作物病虫害领域的农业知识图谱,为农业领域知识的结构化组织提供了数据基础[17];知识图谱扩展的检索增强生成研究则表明,图结构信息能够用于增强问答任务中知识检索与生成[19]。

与主要依据局部文本相似度召回片段的向量检索相比,图结构检索在理论上更有利于连接跨文档、跨概念的农业知识。已有水稻病虫害专家系统通过多种 RAG 基线、知识抽取 F1 和人工评价,报告了图谱增强检索在知识组织和问答任务中的有效性[16]。在此基础上进一步表明,图谱增强系统的潜在作用不仅体现在知识片段召回,还可能体现在灾害机制链组织、风险表达约束和回答结构完整性等方面。对几个模型的具体回复进行结构化分析显示,本智能体系统能够将细胞结冰、膜系统损伤、氧化胁迫、生育阶段差异以及产量和品质后果组织为连续链条,并进一步联系环境条件与管理措施。这一现象反映了检索关系扩展、语言模型叙述组织和流程校验的协同作用,但各模块独立贡献仍需通过受控消融试验确认。

与此同时,本智能体系统平均回答长度增至原来的 2.52 倍,而路径要点命中密度下降 47.5%,说明知识覆盖的增加并未同步转换为单位文本信息效率的提升。较长回答可能容纳更多机制解释、阈值条件和调控措施,但也可能稀释关键结论、增加阅读负担,并降低决策信息的可获取性。这一结果意味着,农业决策智能体的优化目标不应仅包括实体召回率或机制链完整性,还应关注固定输出预算下关键信息密度。对于面向生产决策应用,系统还需具备关键结论前置、证据压缩、适用条件标注和高风险提示保留等能力,从而在知识覆盖与回答简洁性之间取得平衡。

主要补充在于,将“核心知识覆盖程度”与“单位长度内潜在无支撑表达的密度”分别量化,避免使用单一完整性指标同时代表信息充分性和事实可信度。与其他相关研究,如采用检索、澄清和答案生成等多模块并开展消融试验的玉米知识图谱问答研究[18]相比,进一步将极端气候灾害机制链和风险无支撑表达纳入评价框架,为农业智能体的过程性可信度评价提供了新的分析维度。不过,由于风险无支撑知识单元密度以字符数为分母,且判定资源与系统检索库部分同源,该指标更适合反映系统内部的相对证据匹配情况,不能直接等同于事实准确率。结果表明“智能体在当前任务集上提高了预设知识覆盖,并降低了单位长度风险代理指标”,但不支持“GraphRAG 必然优于所有检索方法”或“多角色编排独立产生增益”等更强结论。

3.2 作物机理模型对连续数值任务的贡献

大语言模型适合承担农业知识解释、任务理解和管理建议生成,但输出仍受到领域知识覆盖、事实可靠性和幻觉等问题影响[10,12,14]。对于生育期预测、播期计算和环境响应分析等连续数值任务,仅依赖语言生成过程并不提供显式的作物生理生态过程约束。作物生长模型能够依据结构化环境与管理输入开展连续生育过程计算[20-24],因此将其作为量化工具接入智能体,可以将自然语言任务转换为可重复执行和复核的模型计算过程。

调用 CropGrow 后开花期与成熟期合并评价的 RMSE 由 20.88 d 降低至 5.79 d,说明在当前站点、年度和模型参数条件下,作物模型能够为生育期日期任务提供较稳定的数值计算支持。与此同时,协调智能体主要承担任务识别、输入参数组织和结果解释,使自然语言交互与作物模型计算过程相衔接。该方式体现了大语言模型与作物模型在任务理解和机理计算方面的功能互补。

MCP 将自然语言交互与作物模型接口连接,任务分解、工具选择、参数校验和结构化调用过程与当前大语言模型智能体的工具学习与任务编排范式相

一致[25-29]；核心价值是把站点、作物、播期、气象序列和模拟目标转化为可校验字段。水稻精准施肥研究将知识图谱的定性方案与案例推理的定量施肥量分别评价[33]，同样说明知识检索与数值预测应采用各自适配的证据。未来应在多站点、多年度数据上，将 CropGrow 与 DSSAT[23]、APSIM[24] 及数据驱动模型进行同输入条件下比较，并进一步开展参数敏感性和预测不确定性分析。

3.3　系统集成代价、适用边界与未来工作

主要验证了 GraphRAG 知识检索、CropGrow 机理计算和任务编排协同应用于冬小麦极端气候辅助决策的可行性，研究结论仍受知识图谱质量、回答长度、自动评价方式以及单站点年度量化数据等因素限制。后续研究可从 3 条路线展开。

（1）知识与系统质量优化。知识图谱规模扩大后，关系质量控制将成为限制系统性能的重要因素。本研究中“其他未归并”关系占 17.83%，说明农业知识标准化程度仍有提升空间，未来应结合农业本体和专家审查完善关系规范。同时，目前自动评价主要依赖单一大语言模型裁判，评分结果可能受到模型偏好影响，应引入农业专家盲评、多模型裁判以及规则评价共同构建更加可靠的评价体系，并建立可公开复现的农业智能体标准测试集。此外，针对当前回答长度增加而信息密度下降的问题，需在固定输出预算下优化关键信息密度，开展等输出预算下模块消融试验，明确 GraphRAG、角色编排等各模块的独立贡献。

（2）多作物、多区域拓展。当前量化验证主要集中冬小麦开花期和成熟期预测，未覆盖产量损失预测和管理措施效果评价等更复杂任务。未来应将该框架推广至水稻、玉米、油菜等多作物，在不同生态区、不同年型下验证知识-机理协同的稳定性。随着遥感监测、物联网感知和数字孪生农业的发展，农业智能体需逐步由静态知识问答向动态环境感知和闭环决策演进，融合实时农情监测、天气预报、遥感反演及作物模型模拟，实现“环境感知-风险诊断-过程预测-管理优化”一体化智能决策[34-35]。

（3）多模型对比验证。未来应在相同语料、基座模型、提示和输出预算下，对无检索、向量 RAG、GraphRAG 以及 GraphRAG 加证据回链进行受控比较；在多站点、多年度数据上，将 CropGrow 与 DSSAT[23]、APSIM[24] 及数据驱动模型进行同输入条件下比较，并进一步报告参数敏感性、预测不确定性及跨站点泛化表现。同时引入农业专家盲评，报告事实正确率、证据蕴含率、区域适用性和专家间一致性，从而在多区域、多作物、多模型和长期连续观测试验中验证该框架的泛化能力和实际应用价值。

4　结论

（1）构建了融合 GraphRAG 知识检索、CropGrow 机理模型和流程化任务编排的冬小麦极端气候辅助决策智能体，并基于 450 组同问题回答和 111 组田间生育期观测数据开展评价。

（2）与 Qwen3-Max 相比，本智能体系统专家核心实体加权召回率由 30.69% 提至 37.15%，风险无支撑知识单元密度由 8.63 个/千字降至 5.49 个/千字，表明知识-机理协同系统能够提高既定任务集上的领域知识覆盖和内部证据一致性。

（3）调用 CropGrow 后，111 组冬小麦样本开花期与成熟期预测均方根误差由 20.88 d 降至 5.79 d，R^2 由 0.027 提至 0.896，表明作物模型接入能够增强连续生育过程数值任务的计算稳定性。

（4）在徐州播期推荐案例中，智能体实现了区域与年份识别、知识检索、气象参数组织、播期计算和结果生成之间的流程衔接，验证了知识检索与机理计算协同用于农业辅助决策任务的可行性。

参　考　文　献

[1] Hultgren A, Carleton T, Delgado M, et al. Impacts of climate change on global agriculture accounting for adaptation[J]. Nature, 2025, 642(8068): 644-652.

[2] Ortiz-Bobea A, Ault T R, Carrillo C M, et al. Anthropogenic climate change has slowed global agricultural productivity growth[J]. Nature Climate Change, 2021, 11(4): 306-312.

[3] Jägermeyr J, Müller C, Ruane A C, et al. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models[J]. Nature Food, 2021, 2(11): 873-885.

[4] Zhao C, Liu B, Piao S, et al. Temperature increase reduces global yields of major crops in four independent estimates[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(35): 9326-9331.

[5] Deutsch C A, Tewksbury J J, Tigchelaar M, et al. Increase in crop losses to insect pests in a warming climate[J]. Science, 2018, 361(6405): 916-919.

[6] Lesk C, Rowhani P, Ramankutty N. Influence of extreme weather disasters on global crop production[J]. Nature, 2016, 529(7584): 84-87.

[7] Porter J R, Gawith M. Temperatures and the growth and development of wheat: a review[J]. European Journal of Agronomy, 1999, 10(1): 23-36.

[8] Yang T, Mei Y, Xu L, et al. Application of question answering systems for intelligent agriculture production and sustainable

management: a review[J]. Resources, Conservation and Recycling, 2024, 204: 107497.
[9] Tzachor A, Devare M, Richards C, et al. Large language models and agricultural extension services[J]. Nature Food, 2023, 4(11):941 - 948.
[10] Li H, Wu H, Li Q, et al. A review on enhancing agricultural intelligence with large language models[J]. Artificial Intelligence in Agriculture, 2025, 15(4): 671 - 685.
[11] Chamara N, Pan Y, Taghvaeian S, et al. Can generative AI make farming decisions? Current status and future pathways-a case study in row crop production with ChatGPT[J]. Artificial Intelligence in Agriculture, 2026, 16(2): 837 - 853.
[12] Kuska M T, Wahabzada M, Paulus S. AI for crop production-where can large language models (LLMs) provide substantial value? [J]. Computers and Electronics in Agriculture, 2024, 221: 108924.
[13] Eldem A, Eldem H. The development and evaluation of agricultural question-answering systems based on large language models [J]. Scientific Reports, 2026, 16(1): 5357.
[14] Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation[J]. ACM Computing Surveys, 2023, 55(12): 248.
[15] Zhao P, Zhang H, Yu Q, et al. Retrieval-augmented generation for AI-generated content: a survey[J]. Data Science and Engineering, 2026, 11(1): 1 - 29.
[16] 王波, 丁俊琦, 吴奇峰, 等. 基于多源异构数据和知识图谱的作物病害关联分析[J]. 农业机械学报, 2025, 56(12): 546 - 559.
Wang Bo, Ding Junqi, Wu Qifeng, et al. Association analysis of crop diseases based on multi-source heterogeneous data and knowledge graph[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(12): 546 - 559. (in Chinese)
[17] Yan R, An P, Meng X, et al. A knowledge graph for crop diseases and pests in China[J]. Scientific Data, 2025, 12: 222.
[18] 李书琴, 杨晟沄. 基于多教师蒸馏的猕猴桃种植知识图谱多跳问答模型[J]. 农业机械学报, 2026, 57(14): 278 - 281.
Li Shuqin, Yang Shengyun. Multi-hop question answering model for kiwifruit planting knowledge graph based on multi-teacher distillation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026, 57(14): 278 - 281. (in Chinese)
[19] Linders J, Tomczak J M. Knowledge graph-extended retrieval augmented generation for question answering[J]. Applied Intelligence, 2025, 55(17): 1102.
[20] 朱艳, 汤亮, 刘蕾蕾, 等. 作物生长模型(CropGrow)研究进展[J]. 中国农业科学, 2020, 53(16): 3235 - 3256.
Zhu Yan, Tang Liang, Liu Leilei, et al. Research progress on the crop growth model CropGrow[J]. Scientia Agricultura Sinica, 2020, 53(16): 3235 - 3256. (in Chinese)
[21] Tang L, Zhu Y, Hannaway D, et al. RiceGrow: a rice growth and productivity model[J]. NJAS-Wageningen Journal of Life Sciences, 2009, 57(1): 83 - 92.
[22] 赵春江, 李静晨, 吴华瑞, 等. 基于大语言模型推理的数字孪生平台蔬菜作物生长模型研究[J]. 智慧农业(中英文), 2024, 6(6): 63 - 71.
Zhao Chunjiang, Li Jingchen, Wu Huarui, et al. Vegetable crop growth modeling in digital twin platform based on large language model inference[J]. Smart Agriculture, 2024, 6(6): 63 - 71. (in Chinese)
[23] Jones J W, Hoogenboom G, Porter C H, et al. The DSSAT cropping system model[J]. European Journal of Agronomy, 2003, 18(3 - 4): 235 - 265.
[24] Keating B A, Carberry P S, Hammer G L, et al. An overview of APSIM, a model designed for farming systems simulation[J]. European Journal of Agronomy, 2003, 18(3 - 4): 267 - 288.
[25] Qin Y, Hu S, Lin Y, et al. Tool learning with foundation models[J]. ACM Computing Surveys, 2024, 57(4): 101.
[26] Wang L, Ma C, Feng X, et al. A survey on large language model based autonomous agents[J]. Frontiers of Computer Science, 2024, 18(6): 186345.
[27] Xi Z, Chen W, Guo X, et al. The rise and potential of large language model based agents: a survey[J]. Science China Information Sciences, 2025, 68(2): 121101.
[28] Chowa S S, Alvi R, Rahman S S, et al. From language to action: a review of large language models as autonomous agents and tool users[J]. Artificial Intelligence Review, 2026, 59(2): 71.
[29] Xu W, Huang C, Gao S, et al. LLM-based agents for tool learning: a survey[J]. Data Science and Engineering, 2025, 10(4): 533 - 563.
[30] Chen F, Stogiannidis I, Wood A, et al. A conversational multi-agent AI system for automated plant phenotyping[J]. Nature Communications, 2026, 17(1): 6391.
[31] Allmendinger S, Bonenberger L, Endres K, et al. Multi-agent AI[J]. Electronic Markets, 2026, 36(1): 18.
[32] Traag V A, Waltman L, Van Eck N J. From Louvain to Leiden: guaranteeing well-connected communities[J]. Scientific Reports, 2019, 9(1): 5233.
[33] 戈为溪, 周俊, 袁立存, 等. 基于知识图谱与案例推理的水稻精准施肥推荐模型[J]. 农业工程学报, 2023, 39(2): 126 - 133.
Ge Weixi, Zhou Jun, Yuan Licun, et al. Recommendation model for rice precision fertilization using knowledge graph and case-based reasoning[J]. Transactions of the CSAE, 2023, 39(2): 126 - 133. (in Chinese)
[34] Verdouw C, Tekinerdogan B, Beulens A, et al. Digital twins in smart farming[J]. Agricultural Systems, 2021, 189: 103046.
[35] Chen D, Huang Y. Integrating reinforcement learning and large language models for crop production process management optimization and control through a new knowledge-based deep learning paradigm[J]. Computers and Electronics in Agriculture, 2025, 232: 110028.