Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2026, 15(6), 1469-1480
Published Online November 2026 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2026.156133

Hans汉斯

基于大语言模型的花卉电商选花咨询系统设计与实现

袁傲祥，余　洪，罗　乐

重庆三峡科技大学计算机科学与工程学院，重庆

收稿日期：2026年10月10日；录用日期：2026年11月6日；发布日期：2026年11月16日

摘　要

针对花卉选购需求难以通过关键词准确表达、新用户缺少历史行为记录的问题，本文设计并实现了基于大语言模型的花卉电商选花咨询系统。系统采用前后端分离架构，由智能体引擎调度商品推荐、用户画像查询等5个工具，要求模型依据工具结果回答，并通过结构化数据展示商品。对需求不明确的新用户，系统先通过多轮追问把需求问清楚再给推荐，关键词匹配只做兜底；花语与养护问答则采用触发式检索增强生成(RAG)。测试表明，场景化推荐、知识问答、画像推荐和商品查询功能符合预期，核验样本中的商品信息与数据库一致。系统实现了商品查询与花卉知识问答的集成。该系统为用户提供了自然语言选花咨询方式，也为大语言模型与电商业务数据、领域知识的结合提供了参考。

关键词

大语言模型，智能体，函数调用，检索增强生成，花卉电商

Design and Implementation of a Flower Selection Consultation System for E-Commerce Based on a Large Language Model

Aoxiang Yuan, Hong Yu, Le Luo

School of Computer Science and Engineering, Chongqing Sanxia University of Science and Technology, Chongqing

Received: October 10, 2026; accepted: November 6, 2026; published: November 16, 2026

Abstract

This study presents the design and implementation of a flower selection consultation system based

文章引用: 袁傲祥, 余洪, 罗乐. 基于大语言模型的花卉电商选花咨询系统设计与实现[J]. 人工智能与机器人研究, 2026, 15(6): 1469-1480. DOI: 10.12677/airr.2026.156133

on a large language model to address the difficulty of accurately expressing flower shopping needs through keywords and the lack of historical interaction data for new users. The system adopts a decoupled frontend-backend architecture, with an agent engine coordinating five tools, including product recommendation and user profile retrieval. The model is instructed to answer based on tool outputs, while products are displayed using structured data. For new users whose needs are unclear, the system clarifies the request over several rounds before recommending, and falls back to keyword matching only when clarification fails; questions about flower meanings and care are handled by triggered retrieval-augmented generation (RAG). Tests showed that scenario-based recommendation, knowledge question answering, profile-based recommendation, and product queries met the expected requirements. Product information in the checked samples was consistent with database records. The system integrates product queries with flower knowledge question answering, providing users with a natural-language interface for flower selection and a practical reference for integrating large language models with e-commerce business data and domain knowledge.

Keywords

Large Language Model, Agent, Function Calling, Retrieval-Augmented Generation, Flower E-Commerce

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

“母亲节送什么花”“表白送什么花”是花卉选购中常见的场景化表达，但这类表达通常不含具体的花卉名称，直接检索很难选到满意的花卉商品；新注册用户没有浏览和购买记录，基于历史行为的推荐方法也缺少数据支撑。大语言模型能够理解自然语言需求、处理多轮对话，恰好可以将用户提问与商品查询、知识问答衔接起来，相关方法已有研究积累。例如，Yao 等[1]提出 ReAct 框架，将推理与行动结合，使模型能够借助外部工具完成任务；Carraro 等[2]将大语言模型用于推荐重排序，验证了推荐多样性的改善；Lewis 等[3]提出检索增强生成(RAG)，将外部知识检索与文本生成结合。查龙宝等[4]研究了大语言模型增强与协同信息融合的序列推荐方法，王道平等[5]讨论了大语言模型在银行业的应用与风险。在农业问答方面，吴桓宇等[6]将大语言模型与知识图谱用于水稻病虫害专家系统，王婷等[7]开展了果蔬农技知识智能问答研究。Li 等[8]关注知识图谱上的生成式推理，邹佳贤等[9]则通过澄清式问答处理用户需求不明确的情况。

花卉电商咨询需要同时处理两类信息：一类是平台实际销售的商品，另一类是花语、送花场合和养护知识。商品类回答须对应到具体的商品记录，知识类回答也要有可靠依据。围绕这两类需求，本文构建选花咨询系统：模型理解用户问题并选择工具，后端查询商品或检索知识，结果再交由模型组织回复；花语、养护类专业问题由触发式 RAG 机制作答，工具调用以 5 轮为上限，避免模型反复查询。系统设计的核心是把需求理解与数据获取衔接起来，使回复关联到平台商品数据与知识库内容。

2. 系统总体设计

2.1. 总体架构

系统采用前后端分离架构，由应用层、业务服务层、算法服务层和数据层组成。业务服务与算法服务分别处理常规业务和智能咨询，二者在架构中并列设置，主要模块及其关系见图 1。

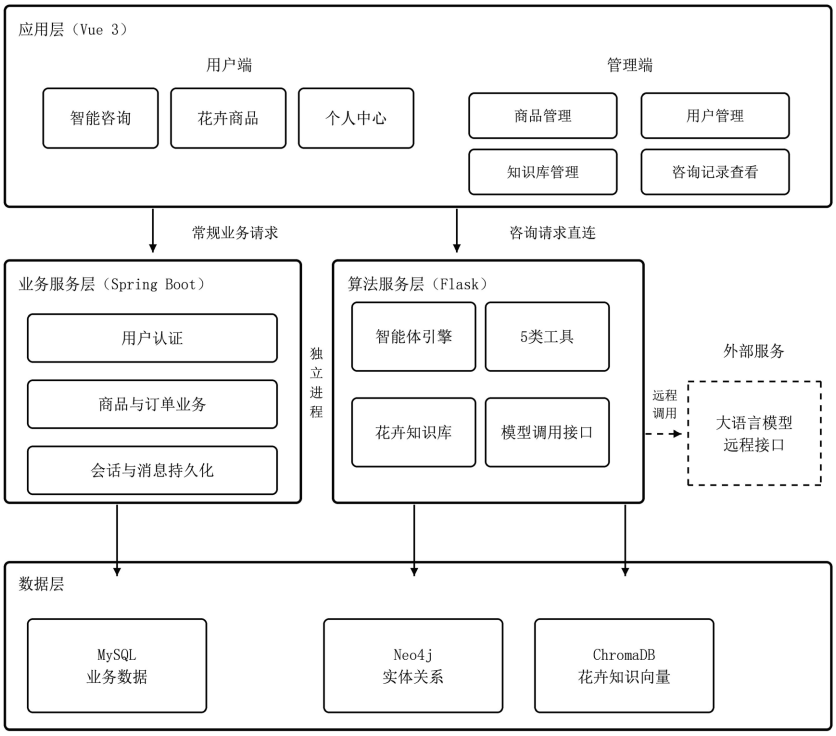


Figure 1. Overall architecture of the intelligent flower selection consultation system
图 1. 智能选花咨询系统总体架构

2.2. 多源数据存储设计

三类数据的查询方式不同，因而采用不同的存储方式。用户、商品和会话等业务记录保存在 MySQL 中，用户与商品之间的关系由 Neo4j 存储，花卉知识的嵌入向量则写入 ChromaDB，具体分工见表 1。

Table 1. Data management division of three databases
表 1. 三类数据库的存储分工

数据库	存储内容	访问方式
MySQL	用户、商品、订单、购物车、会话与消息等业务数据	结构化查询，支持事务
Neo4j	用户、商品、品类实体及浏览、购买、收藏、归属等关系	Cypher 多跳关系查询
ChromaDB	花语、养护等花卉知识文本及其嵌入向量	向量相似度检索

MySQL 保存系统的主要业务记录，包括用户、商品、品类、购物车、用户行为、会话和会话消息等数据。会话消息表记录消息角色、内容及工具名称，为历史消息读取提供依据，其字段见表 2。

Table 2. Field composition of chat message table
表 2. 会话消息表字段组成

序号	字段名	数据类型	长度	主键	备注
1	id	bigint	20	主键	消息编号
2	session_id	bigint	20		所属会话编号
3	user_id	bigint	20		用户编号

续表

4	role	varchar	32	消息角色
5	content	text	—	消息内容
6	tool_name	varchar	64	调用的工具名称
7	create_time	datetime	—	创建时间

3. 基于大语言模型的智能体引擎设计与实现

大模型不掌握平台实时的商品信息，已有的花卉知识也可能不完整。为此，咨询过程中的数据获取交由后端工具完成，模型负责理解问题并选择工具，智能体引擎执行工具调用，结果返回模型后组织回复。本章依次说明大模型选型与接入、咨询会话处理、工具定义与调度、Prompt 约束、Prompt 工程设计思路、调用轮次控制、商品推荐与冷启动、触发式 RAG 知识问答以及知识库构建的实现。

3.1. 大语言模型选型与接入

本系统接入阿里云通义千问系列模型，默认使用 qwen-plus，配置文件中同时保留 qwen-flash 和 qwen-max 两个选项。模型通过 DashScope 的兼容接口远程调用，接口形式便于调整所用模型，服务采用按量计费方式。每次请求由 Prompt、历史消息和工具结果组成；模型返回函数调用请求后，由智能体引擎解析并执行。商品主数据保存在系统数据库中，相关字段按需传入模型。

3.2. 咨询会话处理流程

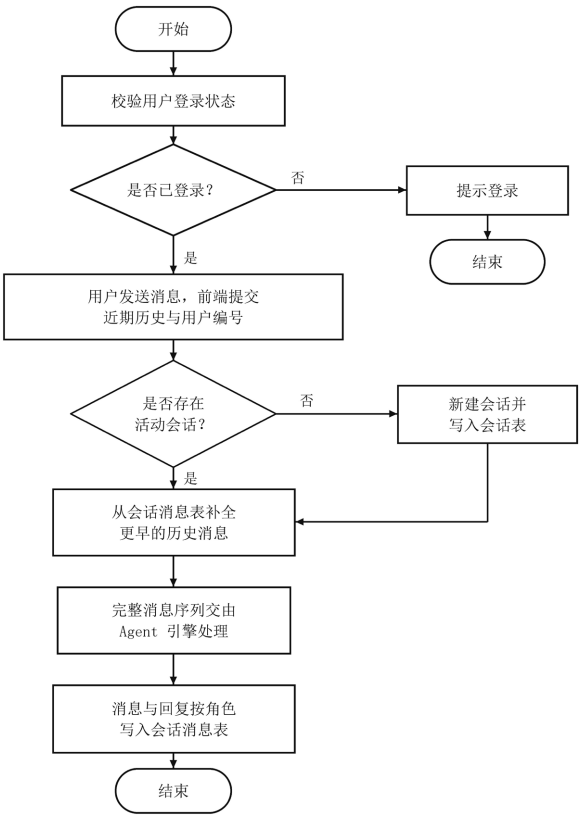


Figure 2. Processing flow of consultation session
图 2. 咨询会话处理流程

咨询会话的处理流程见图 2。当用户建立会话后，后端读取所需历史消息，再将消息序列交给智能体引擎；在本轮结束时保存用户的输入信息和系统的回复。

(1) 会话建立。用户进入智能选花页面时首先先检查用户的登录状态，如果用户未登录则提示登录。在前端提交近期历史消息和用户编号后，服务端开始查找活动会话；若不存在，就创建会话并写入会话表。后续消息通过会话编号与该会话关联。

(2) 历史消息补全。前端只提交最近几轮消息，更早的记录由服务端从会话消息表读取，再交给智能体引擎。这样可以减少对前端缓存的依赖，但可传入的历史长度仍受模型上下文窗口限制。

(3) 上下文持久化。每轮用户输入和最终回复分别以 user、assistant 角色写入会话消息表。下一次追问时，后端可读取这些记录，为模型理解省略表达和指代关系提供上下文。

3.3. 工具定义与函数调用调度

是否调用工具、选择哪个工具，由模型根据用户消息判断；具体执行由智能体引擎完成。图 3 给出了两者之间的调度过程。

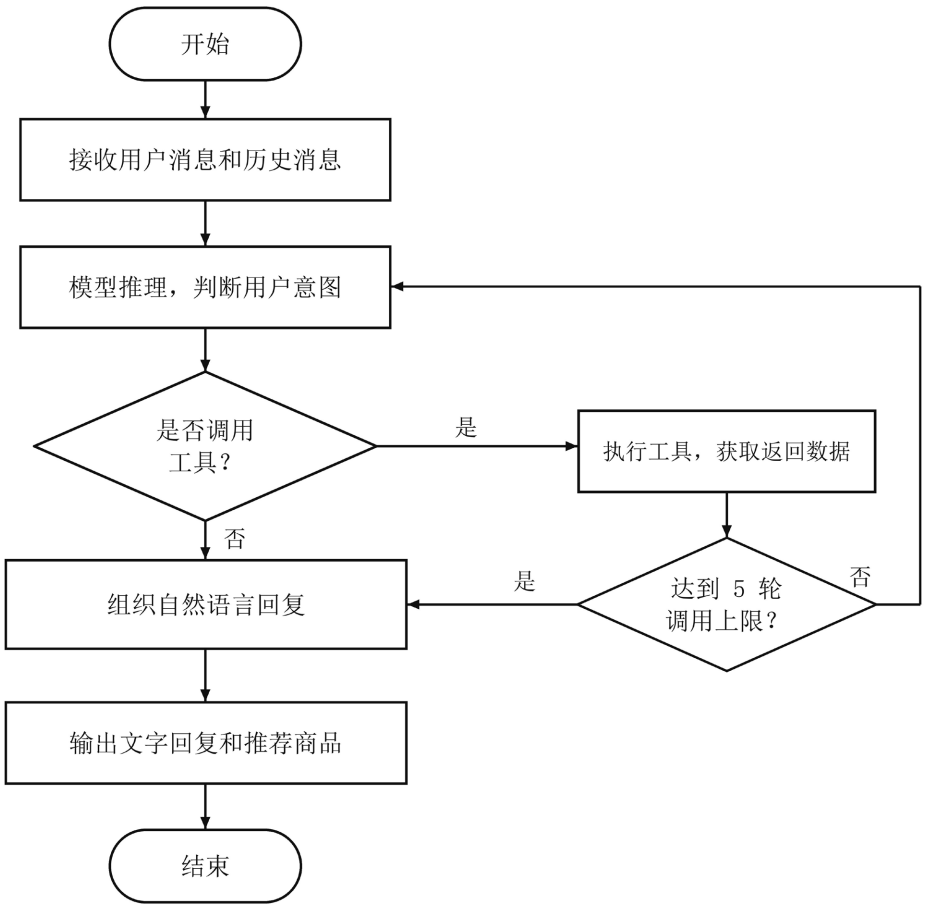


Figure 3. Flow of tool scheduling
图 3. 工具调度流程

系统向模型注册 5 个工具，通过 JSON Schema 描述工具名称、用途和参数。模型据此选择工具，系统不再单独设置意图分类模块。工具的主要功能、参数和数据来源列于表 3。

Table 3. Definitions, parameters and data sources of tools
表 3. 工具定义、参数与数据来源

工具名称	主要功能	关键参数	数据来源
花卉知识检索	检索花语、养护、搭配和节日送花等知识	query	ChromaDB
商品推荐	按协同过滤、基于内容或关键词策略返回商品	user_id、strategy、keywords	Neo4j
用户画像查询	统计偏好品类、偏好标签与行为分布	user_id	Neo4j
商品详情查询	返回商品描述、标签、所属品类和热度	item_id	Neo4j
关键词商品搜索	在商品标题、描述和标签中匹配商品	keyword	Neo4j

商品推荐和用户画像查询都需要 user_id。该参数由服务端从登录态获取并注入，不由模型填写。引擎执行函数调用后，将返回结果作为工具角色消息加入消息序列，供模型继续处理。若当前问题无需调用工具，模型可直接生成回复并结束本轮咨询。

多工具协作的调度过程较复杂，这里选取一次实际咨询为例。用户的问题是“朋友刚做完手术，想订一束适合送医院的花，预算二百元以内，再讲讲寓意”，其中既有送花场景的知识诉求，也有预算约束下的商品筛选，当前单个工具覆盖不足。

模型首先调用花卉知识检索工具，query 取“探病送花”。返回结果建议选用向日葵、百合等寓意康复的花材，并提示避开香味浓烈或带刺的品种。之所以先查知识而不直接推荐，是因为“探病”只是场景词，推荐工具的 keywords 参数需要具体花材名称，中间需做一次转换。

接下来调用商品推荐工具，keywords 取“向日葵、百合”，价格上限设为 200 元。返回的候选列表只有名称、价格和简要描述，香味、带刺这些标签并不在返回字段里，是否违反上一轮的禁忌无从判断。

于是模型又对候选商品逐个调用详情查询工具核对标签，剔除带浓香标签的条目。信息齐备之后才开始组织回复，先交代寓意依据，再逐条介绍剩下的商品。整轮咨询共三次工具调用，没有触及 5 轮上限。

3.4. Prompt 约束与幻觉抑制

取得数据后，还需要限制模型如何使用这些数据。系统 Prompt 要求推荐问题先调用商品推荐或关键词搜索工具，只介绍实际返回的商品，不编造名称、价格和编号；知识回答应以检索片段为依据。对于未获得有效结果的调用，Prompt 要求模型改用其他工具或说明未查到相关信息，避免仅改写查询词后反复调用。获得所需信息后，应转入回答过程。

商品卡片采用结构化数据渲染。算法服务端汇总各轮工具返回的商品，按编号去重，最多保留 10 件，与文字回复一同返回前端。前端直接读取该列表生成卡片，图片、名称和详情页关联信息均来自数据库记录，不从模型生成的文字中提取。这样可减少卡片字段被编造的风险。

3.5. Prompt 工程设计思路

系统 Prompt 工程设计分五步，依次为角色设定、任务流程、内容约束、输出格式、结尾重复关键约束和完整系统提示词，具体设计思路见图 4。

Prompt 并不是一次写完的。调试中出过两类问题。工具查不到结果时，模型会反复换关键词调用同一个工具；介绍商品时它偶尔自己添加促销话术。为解决上述问题，每改一处都用固定的一组测试问题重新测试，确保没有出现新的问题。Prompt 模板主体见附录。

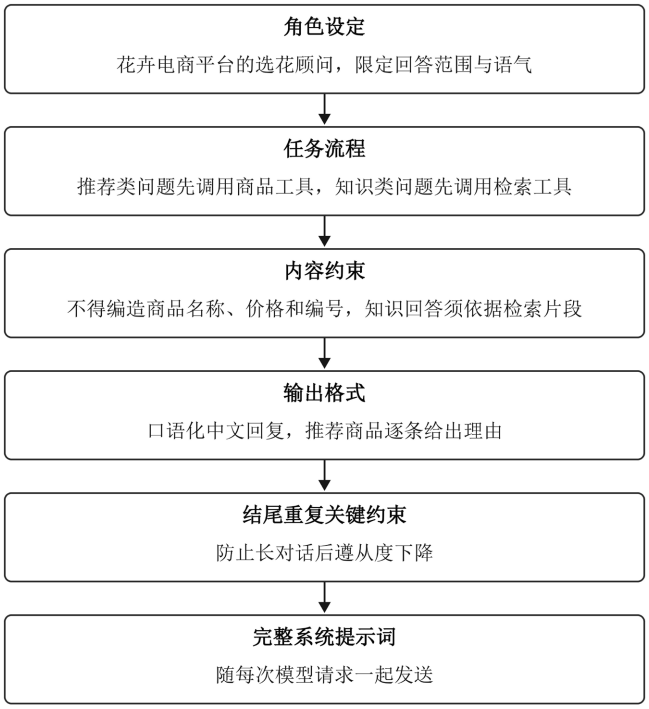


Figure 4. Prompt engineering design
图 4. Prompt 工程设计图

3.6. 工具调用轮次控制

工具调用可能连续进行多轮，也可能因重复查询而迟迟不能结束。系统为此设置调用计数器，将单次咨询的工具调用限制在 5 轮以内。达到上限后，引擎不再提供工具定义，并要求模型依据已有结果回答；信息不足时，改为给出通用性建议。该限制能够减少循环调用，但总响应时间还取决于模型、工具和网络的耗时，不能仅由轮次上限确定。

3.7. 商品推荐咨询与新用户冷启动

新用户没有浏览和购买记录，协同过滤与基于内容两种策略缺少可用数据，只能依赖关键词匹配，而用户表述较为笼统时，检索结果的针对性往往不足。系统因此在匹配之外增加追问环节，当前消息中缺少花卉名称、颜色、用途等要素时暂不推荐，先反问“请问花是送给长辈还是朋友？”“您对价位有要求吗？”，回答并入需求描述后重新判断，一到两轮通常即可形成可用的检索词。追问计入轮次，受 5 轮上限约束，避免反复提问。

执行顺序上，新用户请求先经协同过滤和基于内容策略，没有候选才转入追问，得到的要素再驱动关键词检索。用户仅回答“随便”或直接跳过时不再追问，改用当轮消息中的词语检索，仍无结果再调用关键词商品搜索工具。以“我是新手，想在卧室放一盆好养的植物”为例，画像查询与协同过滤均未返回，最终通过关键词匹配得到绿萝等耐阴易养的盆栽。这条路径不依赖历史数据，提供的是基于当前需求的初始推荐。

3.8. 触发式 RAG 知识问答

花语和养护回答需要具体的知识依据，不能只依靠模型已有的知识。系统在这类问题中引入 RAG，处理流程是先检索花卉知识库，再把相关片段提供给模型，要求其据此作答。处理流程见图 5。

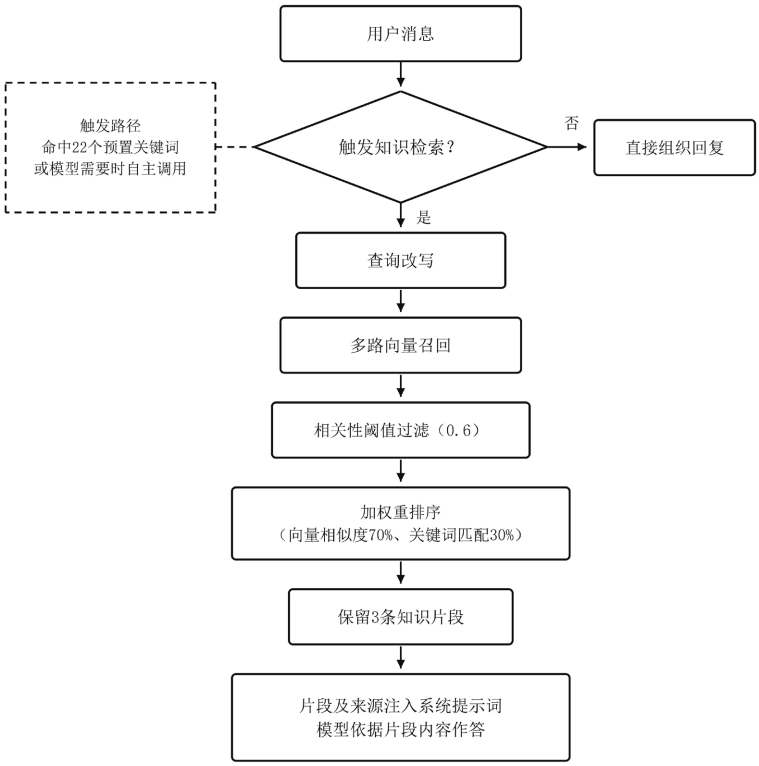


Figure 5. Flow of triggered RAG knowledge question answering
图 5. 触发式 RAG 知识问答流程

(1) 触发判断。用户消息命中预置的 22 个关键词时，系统触发知识检索；模型也可在处理问题时自主调用知识检索工具。未触发检索的消息进入常规回复流程，但未命中触发条件并不等同于消息一定属于寒暄。

(2) 检索召回。系统先将问题改写为语义相关的查询，在 flower_knowledge 集合中进行多路向量召回。候选片段经 0.6 的相关性阈值过滤后，按向量相似度与意图词匹配得分重排序，两项权重分别为 70% 和 30%，最终保留前 3 条片段。

(3) 注入生成。系统将选中的片段及来源标注加入系统 Prompt，要求模型根据片段回答。这里的约束是生成要求，其实际遵循情况仍需通过回复核验判断。

3.9. 知识库构建与管理

知识库收录花语寓意、养护方法、搭配技巧、节日送花、品类介绍和购买指南 6 类内容，共 3000 余条，每条记录类别和来源。管理员可新增、修改、删除条目，也可按类别查询。文本更新后，系统重新调用 text-embedding-v3 生成向量并写入 ChromaDB，使检索索引反映条目变化。咨询会话也可供管理员检索，用于发现高频问题和现有知识的缺口。

3.9.1. 知识库构建流程

知识库的构建分资料收集、文本清洗、主题切分、向量化入库四步，流程见图 6。原始资料先按 6 个类别归档，清洗后切成单主题条目，再由 text-embedding-v3 生成向量并写入 ChromaDB。条目增删改之后索引要同步，新增的直接写入向量，修改的先按 metadata 中的条目标识删掉旧向量再生成新的，删除的同步移除对应向量并做一次一致性校验。

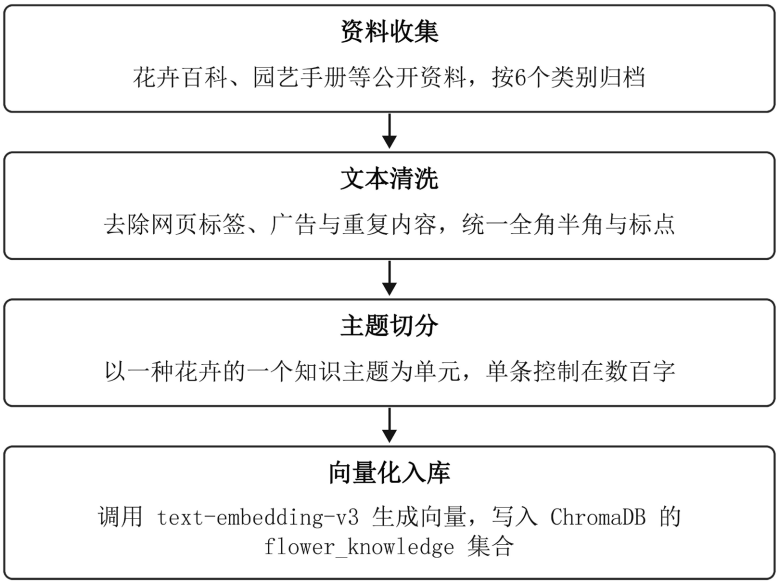


Figure 6. Flow of knowledge base construction
图 6. 知识库构建流程

3.9.2. 文本处理策略

清洗共分为以下三步，去掉网页来源的多余标签，按模式匹配删除版权声明和广告片段，再统一全角半角符号与中英文标点。最后一项容易被忽略，同一花卉名称若全角点与半角点混用，嵌入得到的向量并不相同，检索时就会互相漏掉。

切分以一种花卉的一个知识主题为单元，单条长度控制在 300~500 字，对齐 text-embedding-v3 的有效嵌入窗口。边界取句号或分号这类自然分隔，条目开头统一写明花卉名称，这样片段被单独检出来时仍然能明确是哪种花。

4. 系统测试

测试主要检查咨询功能是否按预期执行、完成一次回复需要多长时间，以及回复中的信息能否对应到数据源。

4.1. 测试环境与方法

测试环境见表 4，模型通过阿里云 DashScope 远程调用 qwen-plus。功能测试和响应时间测试分别设置 4 类用例，每类执行 50 次。

Table 4. Configuration of test environment
表 4. 测试环境配置

类别	配置项	具体配置
硬件环境	处理器	Intel Core i5-11400H (2.70 GHz，6 核 12 线程)
硬件环境	内存	16 GB
硬件环境	网络环境	家庭宽带
软件环境	操作系统	Windows 11
软件环境	开发语言与框架	Java 17、Spring Boot 3.2.1；Python 3.12、Flask 3.1.1；Vue 3 + Vite

续表

软件环境	数据库	MySQL 8.0、Neo4j 5.15、ChromaDB 0.5.23
软件环境	大模型服务	阿里云 DashScope (qwen-plus，远程调用)
软件环境	测试浏览器	Microsoft Edge

4.2. 功能测试

功能测试选取场景化推荐、花卉知识问答、画像推荐和商品查询 4 项功能。表 5 列出各项操作及其预期结果，所列用例均符合预期。

Table 5. Test results of intelligent consultation functions
表 5. 智能咨询功能测试结果

系统功能	操作内容	预期结果	测试结果
场景化推荐	发送“母亲生日送什么花”	返回康乃馨等商品及花语解释	符合预期
花卉知识问答	发送“红玫瑰的花语是什么”	回答与知识库内容一致	符合预期
画像推荐	发送“根据我的喜好推荐几束花”	推荐品类与浏览历史相符	符合预期
商品查询	发送“康乃馨属于什么品类”	返回的描述、标签与数据库记录一致	符合预期

4.3. 响应时间测试

响应时间测试设置普通对话、花卉知识问答、商品推荐和多轮连续咨询 4 组请求，每组执行 50 次，计算平均耗时，结果见表 6。本文以 15 s 作为响应时间的参考值，各组均值均低于该值。

Table 6. Test results of system response time
表 6. 系统响应时间测试结果

请求类型	功能描述	平均响应时间/s
普通对话咨询	无需调用工具的问答	6.35
花卉知识问答	触发知识检索后生成回答	10.86
商品推荐咨询	调用推荐工具后生成回答	12.47
多轮连续咨询	携带历史消息的追问	13.28

4.4. 回复信息一致性核验

一致性核验使用功能测试中的 200 次请求。对其中涉及商品推荐和查询的回复，提取商品编号、名称，并与 MySQL 商品表和 Neo4j 商品节点比对，同时检查品类与标签。知识问答则检查回答要点是否能在命中的知识片段中找到依据，结果见表 7。

Table 7. Verification results of reply information consistency
表 7. 回复信息一致性核验结果

核验内容	核验方法	核验结果
商品编号与名称	与 MySQL 商品表、Neo4j 商品节点逐条比对	全部一致，无虚构商品
品类与标签属性	与数据库中的商品记录比对	全部一致
知识类表述	与命中的知识库片段比对	要点均有出处

4.5. 新用户冷启动专项测试

为检验追问环节是否真的起作用，本实验注册了 20 个没有任何历史行为的新账号，用“想买一束花，不知道怎么选”“第一次买花，帮我推荐一下”等 10 类模糊需求分别测试。每类需求在两种模式下分别测试，模式 A 含追问，模式 B 直接按关键词匹配，共计 400 次咨询。评价指标按照以下四项，需求要素获取成功率、推荐商品与用户需求的匹配率、平均澄清轮数以及平均完成时间；匹配率由人工判断商品品类或用途是否与用户描述相符，结果见表 8。

Table 8. Special test results of the cold-start strategy for new users
表 8. 新用户冷启动专项测试结果

对比项	模式 A：澄清式问答	模式 B：关键词保底
需求要素获取成功率	95%	58%
推荐匹配率	86%	63%
平均澄清/重试轮数	1.7	—
平均完成时间/s	21.4	12.9

5. 结论

本文实现了一个面向花卉电商的选花咨询系统。用户以自然语言提出需求，模型通过智能体引擎调用商品查询、推荐和知识检索等工具，再根据返回结果组织回复。对缺少历史记录的新用户，系统会先问清需求再据此检索，关键词匹配只在追问无效时兜底；对花语和养护问题，采用触发式 RAG 提供回答依据。商品卡片直接使用结构化数据，工具调用限制在 5 轮以内。所测功能符合预期，4 类请求的平均响应时间为 6.35~13.28 s，核验样本中的商品信息与数据库记录一致。现有测试说明该流程能够完成所列咨询任务。新用户专项测试中，加入追问后需求要素获取成功率为 95%，推荐匹配率为 86%。该系统将选花需求、商品信息与养护知识整合到同一咨询过程中，为不熟悉花卉名称或缺少购买经验的用户提供了表达需求和获取信息的途径。同时，通过工具调用连接平台数据、通过知识检索补充专业内容，为大语言模型在垂直电商场景中的应用提供了可借鉴的实现方式。

参考文献

[1] Yao, S., Zhao, J., Yu, D., *et al*. (2023) ReAct: Synergizing Reasoning and Acting in Language Models. 2023 *International Conference on Learning Representations*, Kigali, 1-5 May 2023, 1-33.

[2] Carraro, D. and Bridge, D. (2025) Enhancing Recommendation Diversity by Re-Ranking with Large Language Models. *ACM Transactions on Recommender Systems*, **4**, 1-40. https://doi.org/10.1145/3700604

[3] Lewis, P., Perez, E., Piktus, A., *et al*. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, Online, 6-12 December 2020, 9459-9474.

[4] 查龙宝, 黄琪, 王明文, 等. 基于大语言模型增强和协同信息融合的序列推荐方法[J]. 模式识别与人工智能, 2025, 38(10): 925-937.

[5] 王道平, 梁雅夕. 人工智能大语言模型的技术基础、应用场景与风险防范——以银行业为例[J]. 东北财经大学学报, 2025(4): 17-30.

[6] 吴桓宇, 陈明, 郭钦鸿. 基于大语言模型与知识图谱的水稻病虫害专家系统[J]. 农业工程学报, 2025, 41(22): 244-255.

[7] 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.

[8] Li, K., Zhang, T., Wu, X., Luo, H., Glass, J.R. and Meng, H.M. (2025) Decoding on Graphs: Faithful and Sound

Reasoning on Knowledge Graphs through Generation of Well-Formed Chains. *Proceedings of the* 63*rd Annual Meeting of the Association for Computational Linguistics* (*Volume* 1: *Long Papers*), Vienna, July 2025, 24349-24364. https://doi.org/10.18653/v1/2025.acl-long.1186

[9] 邹佳贤, 陈雷, 何浩楠, 等. 基于大语言模型的澄清式玉米知识图谱问答方法[J]. 农业工程学报, 2026, 42(5): 223-234.

附　录：核心 Prompt 模板

你是花卉电商平台的选花顾问，负责解答用户的选花问题。请严格遵守以下规则：

(1) 用户提出商品推荐或查询需求时，先调用商品推荐或关键词商品搜索工具，只介绍工具实际返回的商品，商品名称、价格、编号和促销信息一律不得编造；

(2) 花语、养护、搭配这类知识问题，请先调用花卉知识检索工具，回答限定在返回的知识片段之内，不要脱离片段自行发挥；

(3) 工具没有返回有效结果时，可以换个工具再试一次；还是查不到，就如实告知用户并给出一般性建议，不能虚构；

(4) 一次推荐的商品不超过 5 件，每件都要说清推荐理由。回复用口语化的中文，夸张的营销话术不要出现；

(5) 工具参数根据用户消息里的送花对象、场合、价位等要素确定，user_id 这类参数由系统自动注入，不用你填。