

Dumpling智能Excel公式解析与 结构化写回系统V1.0 设计说明书

版本	V1.0
负责人	杨致一
创建日期	2026-04-25

修订历史

时间	版本	修改者	描述
2026-04-25	V1.0	杨致一	按V1.0发版口径补充架构、模块、数据模型和安全边界设计

目录

1 简介	4
1.1 编写目的	4
1.2 使用对象	4
2 产品介绍	5
2.1 产品愿景	5
2.1 目标行业与用户	5
2.3 模块说明	6
2.4 总体架构	7
2.4.1 前端设计	8
2.4.2 后端设计	8
2.4.3 数据模型设计	9
2.4.4 Agent Runtime 设计	9
2.4.5 工具 Schema 设计	10
2.4.6 SQL 解析设计	10
2.4.7 写回安全设计	10
2.4.8 存储设计	11
2.4.9 异常与恢复设计	11
2.4.10 扩展设计	11

1 简介

Dumpling智能Excel公式解析与结构化写回系统的设计目标是构建一个通用Excel公式与结构化写回Agent基座。系统通过模型、工具、程序校验和人工确认之间的协作，实现从自然语言任务说明和Excel文件到写回计划、公式明细、SQL解析和结果文件的完整处理链路。

系统遵循模型负责开放目标理解和候选解释、程序负责结构化工具和可证明校验、人工负责低置信和不可证明场景最终确认的原则。Excel写回必须由后端程序执行，模型不能直接写文件。

1.1 编写目的

本文档用于说明Dumpling智能Excel公式解析与结构化写回系统V1.0的总体架构、前后端模块、数据模型、Agent Runtime、工具Schema、SQL解析、写回安全、存储与异常恢复设计。

1.2 使用对象

本文档主要面向系统架构人员、开发人员、测试人员和运维维护人员，用于理解系统设计边界、模块职责和后续扩展原则。

2 产品介绍

2.1 产品愿景

Dumpling旨在为复杂Excel公式、跨文件结构映射和结果写回任务提供可核对、可追踪、可恢复的智能处理工作区。系统将自然语言任务说明、Excel文件、SQL数据来源、知识库资产和人工确认流程组织为统一任务链路，帮助用户减少人工查找公式、核对SQL字段和手工回填结果的工作量。

2.1 目标行业与用户

设计范围包括前端工作区、后端任务服务、Excel解析、高级解析、SQL解析、公式知识库、用户资产、Agent Runtime、模型接入、人工确认、受控写回、结果版本、管理后台和应用内Wiki。

V1.0以预算模块作为首发验证路径，同时保留通用Excel公式与结构化写回能力，后续报表、单据等模块可复用同一Agent Runtime、工具Schema、知识库和写回裁决机制。

系统V1.0以预算模块作为首发验证路径，但总体设计面向通用Excel公式解析和结构化写回场景，不把预算字段、目标列或样例答案固化为产品硬规则。

设计边界上，系统不承诺模型直接写Excel，不承诺任意表样全自动落格；所有低置信、不可证明或缺少上下文的候选都进入人工确认或只读说明流程。

系统核心技术特点包括：模型与程序职责分离、SQL文本可证明校验、写回候选裁决、人工确认闭环、结果版本管理、用户资产按模块检索以及可观测Agent循环事件。模型不能直接写Excel，所有写回动作均由后端受控工具在用户确认和程序校验后执行。

系统采用前后端分离架构。前端基于React、TypeScript、Vite、Zustand、TanStack Query、Axios、Tailwind CSS和shadcn/ui实现；后端基于FastAPI、Pydantic v2、Openpyxl、SQLite、JSON文件存储和OpenAI-compatible模型接口实现。

数据流为：用户在前端创建任务并上传文件；后端保存文件和任务状态；后端解析Excel、SQL和文件关系；模型根据任务说明和工具观察生成结构化意图或候选；程序校验候选并生成写回计划；不确定项进入人工确认；确认后后端执行受控写回；前端展示结果预览、执行日志和版本信息。

系统通过HTTP API通信，当前主链路以任务详情刷新和轮询为主，不以WebSocket实时事件流作为V1.0发布阻塞项。

前端通过HTTP API与后端通信，任务状态以服务端TaskRecord为准。后端负责文件保存、任务状态机、公式解析、候选生成、模型调用、人工确认恢复、受控写回和结果版本记录。

2.3 模块说明

模块	模块描述
前端工作区模块	提供用户登录后的任务工作台，包含任务列表、新建任务、任务详情、文件信息、执行进度、结果摘要和结果预览入口。用户可在同一工作区完成文件上传、任务说明填写、任务状态查看、人工确认处理、结果下载和后续追问。
Agent 对话与执行日志模块	提供右侧 Agent 面板，用于展示用户对话、系统回复、人工确认问题、模型操作、工具执行记录和任务流程进度。该模块负责把模型、工具、程序校验和用户确认过程以可读方式呈现给用户。
文件上传与文件角色模块	支持用户上传公式列表、表样文件、基础数据、SQL 包、输出样例和参考文件。系统为每个上传文件生成稳定文件 ID，并记录文件角色、文件大小、上传路径和任务归属，避免同名文件互相覆盖。
文件关系配置模块	支持用户描述来源文件、目标文件、参考文件和依赖文件之间的关系。对于复杂映射场景，用户可补充来源 Sheet、来源值、目标 Sheet、目标值等结构化线索，供后续公式解析和写回候选生成使用。
任务服务与状态机模块	负责创建任务、保存任务记录、维护任务状态、调度后台执行、写入任务消息、恢复人工确认后的执行流程，并管理任务结果路径和版本信息。任务状态包括待处理、运行中、等待用户、成功、失败和取消。
Excel 解析模块	基于 Openpyxl 读取 Excel 工作簿结构，识别 Sheet、单元格、公式表达式、表头、行列锚点和可写回区域。该模块为公式解释、表样定位、候选生成和结果预览提供结构化数据基础。
公式解析与知识库模块	解析 Excel 公式中的函数、参数、变量、代码和业务含义，并结合公式库、语法库和模块知识生成公式明细及中文说明。公式库支持按模块隔离、Excel 导入导出、管理员覆盖和历史版本维护。
SQL 解析模块	读取 SQL 包中的 SQL 标识和 SQL 文本，解析字段引用、SQL 片段、表名、字段名和外部取数来源。系统采用“开放目标 + 可证明校验”机制，只认可可在 SQL 文本中证明的候选，无法证明的候选进入人工确认或只读说明。
写回候选生成模块	根据目标表样、行列锚点、公式明细、SQL 证据、基础数据、指标库和文件关系生成写回候选。每条候选包含目标文件、目标 Sheet、目标行列、写回值、证据摘要、置信度、状态和下一步建议。
人工确认模块	当系统无法唯一确定主文件、目标文件、目标列、候选位置、资产选择或写回策略时，生成结构化人工确认问题。用户可选择候选项、补充说明或上传补充文件，确认结果会写入任务上下文并驱动后续恢复执行。

Dumpling智能Excel公式解析与结构化写回系统

模块	模块描述
受控写回模块	负责将经过程序校验和用户授权的写回候选写入结果工作簿。模型不能直接写 Excel，写文件类工具必须具备明确授权和结构化写入参数。低置信、不可证明或未确认候选不会自动写入。
结果文件与预览模块	生成最终输出工作簿，并保留写回计划、公式明细、SQL 解析等辅助 Sheet。前端以页签形式展示结果文件内容，支持用户核对每条写回结果的来源、目标位置、证据和状态。
结果版本管理模块	任务完成后，系统记录结果版本、结果文件路径、版本号、修订线索和差异摘要。用户继续追问或要求修订时，系统基于当前结果生成新版本，保留版本链路，便于复核和回溯。
Agent Runtime 模块	负责组织模型、工具、任务状态、人工确认和结果版本之间的执行流程。该模块记录用户输入、意图解析、模型 JSON 校验、工具调用、工具观察、下一步动作、人工确认、写回裁决和结果版本等事件。
工具 Schema 与工具调度模块	将系统工具分为只读工具、非文件副作用工具和文件写入工具，并为每类工具定义输入输出结构和副作用边界。未知工具会记录失败但不破坏主任务，文件写入工具必须经过授权和程序校验。
模型接入模块	支持 OpenAI-compatible 模型配置，包括 provider、model、base_url、api_key 和默认模型设置。系统仅在开放目标理解、候选补充、解释组织或不确定判断时调用模型，并记录模型调用用量。
用户资产模块	管理基础数据、指标库和表样库等用户业务资产。资产按用户和模块范围隔离，未选择或未确认的资产不进入运行时明细上下文，避免无关资产影响模型判断。
智能导入模块	将用户上传的 Excel 文件解析为可维护资产，生成导入预览，用户确认后写入基础数据、指标库或表样库。该模块用于将业务资料沉淀为后续任务可复用的结构化知识。
管理后台模块	提供模块注册表、公式库、语法库、基础数据、指标库、表样库、模型配置、用户管理、邀请码、用量日志和系统设置等维护入口。管理员可通过该模块维护系统运行所需的知识和配置。
应用内 Wiki 模块	提供产品定位、架构说明、任务模式、文件角色、Agent 状态制品、API 速查、前后端模块地图和验证矩阵等说明，作为项目知识入口和使用参考。
数据存储模块	采用 SQLite、JSON 文件和文件系统混合存储。SQLite 用于公式知识库、语法库和部分用户资产；JSON 用于任务、会话、模型配置和用量日志等轻量状态数据；文件系统用于保存上传文件、结果文件和任务临时文件。
安全与边界控制模块	负责控制模型权限、写回授权、低置信候选处理、用户资产范围和任务结果证据留存。系统明确禁止模型直接写 Excel，所有写回结果必须经过程序校验和授权，确保结果可核对、可追踪、可恢复。

2.4 总体架构

系统采用前后端分离架构。

前端：React 19。TypeScript。Vite。Zustand 状态管理。TanStack Query 数据请求。Axios API 客户端。Tailwind CSS 和 shadcn/ui 组件库。

后端：FastAPI Web 服务。Pydantic v2 数据模型。Openpyxl Excel 处理。SQLite 知识库存储。JSON 文件任务状态存储。OpenAI-compatible 模型调用接口。

数据流：用户在前端创建任务并上传文件。后端保存文件和任务状态。后端解析 Excel、SQL 和文件关系。模型根据任务说明和工具观察生成结构化意图或候选。程序校验候选并生成写回计划。不确定项进入人工确认。确认后后端执行受控写回。前端展示结果预览、执行日志和版本信息。

2.4.1 前端设计

前端由路由、布局、工作区、Agent 面板、结果预览、管理后台和 Wiki 组成。

主要页面：登录页。工作区首页。会话详情页。智能导入页。管理后台页面。Wiki 页面。

会话详情页采用三栏布局：

左栏：模块、最近任务、文件和关系摘要。

中间：任务创建、执行状态、结果预览和摘要信息。

右栏：Agent 对话、人工确认、执行日志和模型操作。

前端通过轮询或刷新任务接口获取任务状态。人工确认、补充文件、取消任务、删除会话和清空对话均通过 API 与后端交互。

2.4.2 后端设计

后端围绕任务服务、Excel 转换服务、高级解析服务、SQL 解析服务、知识库服务、用户资产服务和模型服务构建。

核心模块：

task_service：任务创建、状态机、消息、人工确认恢复、结果版本。

excel_converter：Excel 公式读取、转换、结果工作簿生成。

advanced_resolution：高级转换、文件关系、目标定位、写回候选。

sql_bundle：SQL 包解析、字段引用和模型候选校验。

budget_knowledge: 公式库、语法库、内置同步和 Excel 导入导出。

user_assets: 基础数据、指标库、表样库和智能导入。

model_service: OpenAI-compatible 模型调用和用量日志。

2.4.3 数据模型设计

主要数据对象:

Session: 用户会话和任务入口。

Task: 任务状态、文件、消息、结果和人工确认。

TaskFile: 上传文件及其角色、路径和文件 ID。

TaskRelation: 文件之间的关系和映射说明。

HITLQuestion: 人工确认问题、选项和上下文。

WritebackCandidate: 写回候选、置信度、证据和状态。

ResultVersion: 结果版本、文件路径、修订摘要和差异说明。

ModelConfig: 用户模型配置。

UsageLog: 模型调用用量日志。

FormulaKnowledge: 公式库条目。

SyntaxKnowledge: 语法库条目。

BaseDataset / Metric / Template: 用户业务资产。

2.4.4 Agent Runtime 设计

Agent Runtime 是系统的调度层。它不直接替代业务工具，而是负责组织模型、工具、任务状态和人工确认。

Agent 事件包括:

- ① 用户输入。
- ② 意图解析。
- ③ 模型 JSON 校验。

- ④ 工具调用和观察。
- ⑤ 下一步动作。
- ⑥ 人工确认请求。
- ⑦ 人工确认结果。
- ⑧ 写回裁决。
- ⑨ 结果版本。

每轮 Agent 循环会生成聚合状态，包括阶段、状态、事件计数、模型校验次数、工具观察次数、人工确认次数和结果版本数。

2.4.5 工具 Schema 设计

系统将工具分为：

- ① 只读工具：读取结果、解析候选、生成诊断。
- ② 非文件副作用工具：生成确认问题或回复上下文。
- ③ 文件写入工具：受控写回 Excel 文件。

④ 工具调用必须符合已登记 Schema。未知工具会记录失败但不破坏主任务。写文件类工具必须经过授权和程序校验。

2.4.6 SQL 解析设计

SQL 解析采用“开放目标 + 可证明校验”的设计：

- ① 模型可以返回字段引用、SQL 片段、SQL 解释、改写、审查或 custom 候选。
- ② 程序只验证 SQL 文本中能证明的表名、字段、表达式和片段。
- ③ 无法验证的候选不自动写回。
- ④ 低置信或不可证明候选进入人工确认或只读说明。

2.4.7 写回安全设计

写回安全边界：

- ① 模型不能直接写 Excel。
- ② 低置信候选不能自动写。

- ③ 写回工具必须带有 `confirmed_candidates` 或 `user_confirmed` 授权。
- ④ 程序从写回计划中抽取高置信候选生成结构化 `writes`。
- ⑤ 执行写回后记录 `controlled_writeback` 和结果版本。

2.4.8 存储设计

当前系统采用混合存储：

SQLite：公式知识库、语法库和部分用户资产。

JSON：任务、会话、模型配置、用量日志等轻量状态数据。

文件系统：上传文件、结果文件和任务临时文件。

该设计便于 V1.0 版本在本地或内网环境中快速部署和排查，同时为后续迁移到更完整数据库保留边界。

2.4.9 异常与恢复设计

系统对异常采取以下策略：

- ① 文件缺失或角色不明：进入人工确认或提示补充文件。
- ② 主文件或目标文件不唯一：请求用户确认。
- ③ SQL 候选不可证明：保留说明或请求确认。
- ④ 写回候选低置信：进入候选分组确认。
- ⑤ 模型不可用：回退本地规则或进入人工确认。
- ⑥ 任务失败：保留错误消息和任务状态。

2.4.10 扩展设计

系统支持通过模块注册表扩展业务模块。新增模块可以配置：

模块 ID。模块名称。是否启用公式库。是否启用语法库。是否支持标准 Excel 能力。是否仅管理员可见。