

小红书社区搜索相关性 智能审核打标 Copilot

业务规则 · 多模态分析 · N8N workflow · 人机协作

44 组图表重排 正文完整保留 文字可搜索、可放大

原文作者: city 不 city (CSDN: m0_74750121)

原文更新: 2026-05-09 11:36 | 重排日期: 2026-10-17

[打开原文](#) | [CC BY-SA 4.0 许可](#)

本版将密集截图中的规则、定义、流程和案例重新编排为文字表格与矢量流程图。复杂表格可跨页，避免缩小字号；保留照片与清晰界面，并为含小字的图版补充文字说明。

原文 17 个章节和 53 处图组均保留对应关系。原站有 1 处图片失效；被截断、无法辨读的细节不补写。原材料存在的待确认规则和数值冲突单独标注。

版权归属保留：原文为博主原创，声明使用 CC BY-SA 4.0。本整理版注明图表重排改动，并依相同许可分享。

阅读指南与章节目录

图表规范：表题在上、图题在下，图表及标题居中。图与表分别连续编号；跨页表格重复表题和表头，并标注“续”。资料索引保留 53 组来源对应关系。

清晰度：图表主体使用 10.3-11.1 pt 原生文字，支持搜索和无损放大。每组保留原图链接，文末索引可直接跳转。

改动范围：正文叙述保留。图表按内容重排，部分图题改为实际主题；案例的冗长笔记 ID、无地址的链接标记与界面装饰不再抄入。截图中的凭据不展示。

内容边界：30% 提效、10% 自动化为原文目标。不同版本规则分别呈现，未合并为已验证标准；不清楚或被截断的内容明确说明。

一、作品概览

二、项目背景与业务痛点

三、项目目标

四、整体方案设计

五、数据输入与处理对象

六、query 拆解框架

七、规则体系设计

八、核心工作流 SOP

九、案例展示：多模态样例与规则应用

十、规则优化与细分类目展示

十一、Prompt 设计与模型输出约束

十二、模型与系统工程实现

十三、人机协作与自动化分析

十四、项目成果

十五、项目难点与解决方案

十六、个人能力体现

十七、作品集展示摘要

3

3

4

7

11

11

15

21

22

28

50

51

52

62

62

63

63

一、作品概览

作品名称

小红书社区搜索相关性智能审核打标 Copilot

作品类型

AI 业务应用落地 / 多模态审核打标 / N8N 自动化工作流 / 人机协作 Copilot

项目定位

本作品面向小红书社区搜索相关性审核打标场景，目标是通过大语言模型、多模态模型和自动化工作流，将原本依赖人工经验的搜索相关性判断流程，转化为一套可结构化分析、可批量运行、可复盘优化的 AI 辅助审核系统。

该项目不是单点调用大模型，而是围绕“query 与笔记内容是否相关、相关到什么程度、应该给几分”这一业务判断，搭建了从数据读取、query 拆解、多模态内容理解、规则匹配、分值计算、结果写入、人工复核到 bad case 迭代的完整流程。

我的角色

在项目中，我主要负责业务规则理解、流程拆解、Prompt 设计、N8N 工作流搭建、模型调用链路设计、数据清洗、评分逻辑实现、baseline 测算和 bad case 复盘。

二、项目背景与业务痛点

小红书社区搜索相关性审核的核心任务，是判断用户搜索词 query 与笔记内容之间的匹配程度，并按照业务规则输出相关性分值。笔记内容可能包含标题、正文、图片、视频、话题 tag、链接等多种信息。人工需要综合判断用户需求、内容主题、是否命中实体、是否满足限制条件，以及有效内容在整篇笔记中的占比。

原始人工流程存在以下问题：

1. 内容阅读成本高：一条笔记可能包含长正文、多张图片，甚至长视频。
2. 规则理解成本高：评分规则包含需求、实体、限制条件、主题占比、有效内容等多个维度。
3. 人工一致性不足：不同作业人员对边缘 case 的理解存在差异，容易出现打分不一致。
4. 培训周期长：新人需要理解大量规则和案例，爬坡成本高。
5. 复盘成本高：当人工与 AI 或双盲标注结果不一致时，需要人工回看完整分析链路。

根据项目过程文档，该业务具有明显的人力密集型特征：作业人员需要长时间阅读和判断，日常复盘频繁，复杂规则会导致返工和标注差异。因此，本项目希望构建一个面向作业人员的 Copilot，把“人工阅读、主观分析、经验决策”转化为“AI 自动化处理、人机协同校验、过程可追溯”的新模式。



图 1 审核打标工作模式流程

资料 01 / 53 · 矢量重排 · [原图来源](#)

三、项目目标

项目目标分为业务目标和技术目标。

业务目标

第一阶段目标是达到 Copilot 级别：系统作为作业人员的轻量化智能辅助工具，帮助作业人员快速理解 query、拆解需求、判断笔记相关性，并给出可解释的评分建议。项目过程文档中明确提出，第一阶段希望实现业务提效 30% 以上。

第二阶段目标是向 Agent 级别演进：在 Copilot 辅助分析基础上，进一步实现部分业务动作的自动化执行，目标是实现 10% 以上的业务自动化操作。



图 2 Copilot 业务目标示意图

资料 02 / 53 · 原始图像 · [原图来源](#)



豆包AI生成

图 3 Agent 业务目标示意图资料 03 / 53 · 原始图像 · [原图来源](#)

技术目标

技术目标包括：

1. 构建 query 结构化解析能力。
2. 构建正文、图片、视频等多模态内容理解能力。
3. 将业务评分规则转化为可执行的流程节点。

- 4. 将模型输出统一为 JSON 结构，方便后续计算和复核。
- 5. 通过 N8N 实现可视化、可扩展、可复用的流程编排。
- 6. 建立 baseline 测算和 bad case 归因机制。
- 7. 为后续 Model Hub、RAG、Agent 自动化作业打基础。

四、整体方案设计

项目采用 Workflow 作为技术载体，围绕 N8N 搭建 workflow 驱动模型协同体系。整体链路包括数据输入、规则拆解、模型分析、结果合并、评分计算和输出写入。

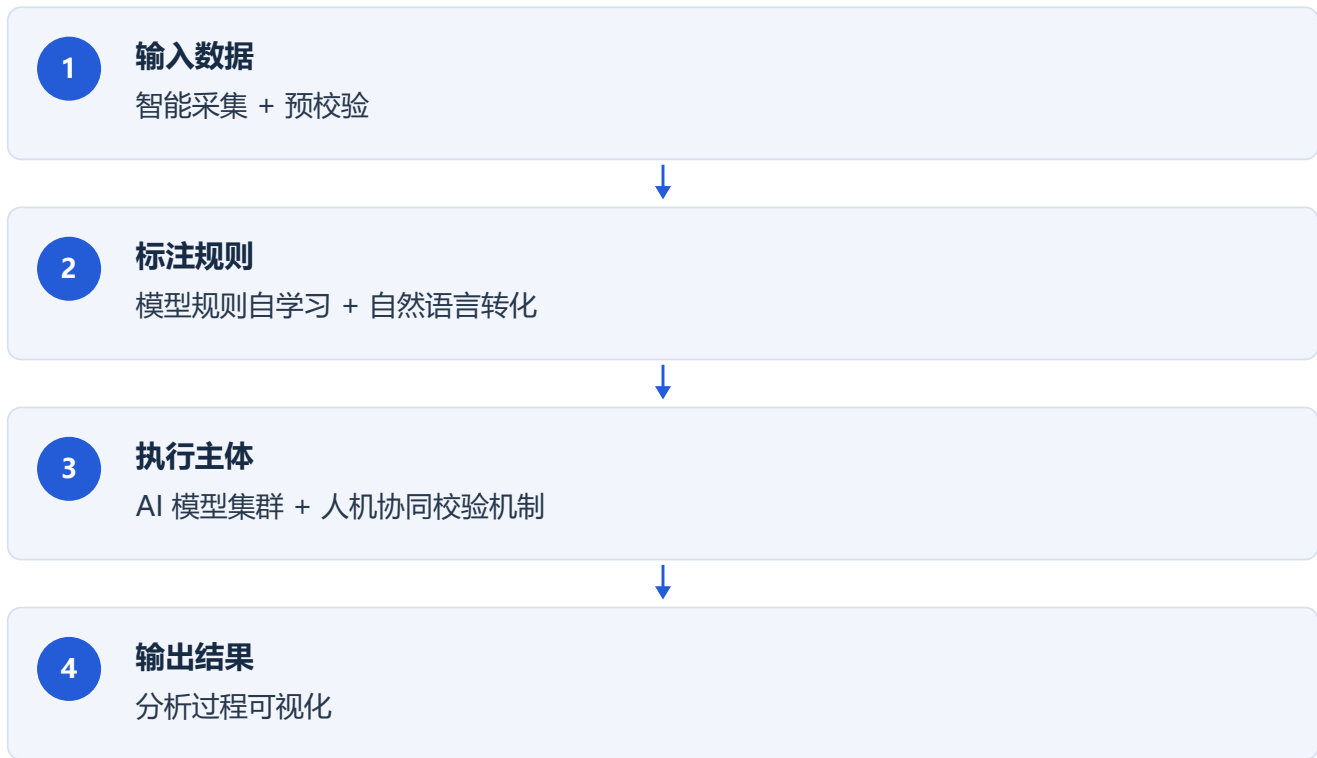


图 4 Copilot 流程价值示意

资料 04 / 53 · 矢量重排 · [原图来源](#)

主流程



图 5 Workflow 模型协同主流程

结构拆解与意图维度

表 1 语法结构与意图维度

模块	原图列出的内容
语法 / 实体	核心词、限定词、修饰词、功能词；人名、地名、时间词、数量词。
基础与深层意图	核心事务意图、信息查询意图、比较筛选意图、场景化意图、情感倾向意图、潜在需求意图等。
多轮匹配	意图 1 / 意图 2 / 意图 N → 各自提取内容 → 各自推理匹配。

此图的评分分支

表 2 workflow 评分分支

判断条件	输出
全部满足，且有效内容占比 ≥ 80%	3 分
全部满足，但有效内容占比 < 80%	2 分
未全部满足，但部分满足	1 分
不满足，但有关联	0 分
不满足，也无关联	-1 分

这是原图中的流程版本。后文“整体分档”还有 10% 边界与细类规则，分别保留，不混写成同一套规则。

资料 05 / 53 · 矢量重排 · [原图来源](#)



图 6 Model Hub 技术架构

表 3 Model Hub 配套模块与连接关系

配套模块	职责 / 连接关系
数据处理 Agent	数据交互。
流程控制 Agent	控制指令交互。

表 3 Model Hub 配套模块与连接关系（续）

配套模块	职责 / 连接关系
模型调度 Agent	模型调用交互，与 Model Hub 双向连接。
规则文档	Standard Doc. / SOP Doc. 位于核心 workflow 模块内部。
存储	向量数据库、关系数据库、非结构化数据；与核心 workflow 连接。
插件工具箱	为核心 workflow 提供工具能力。
RAG 管道	检索器、索引、评估；与核心 workflow 双向交互并反馈结果。
模型优化	PEFT 训练、评估、测试、部署运维，作用于 Model Hub。

资料 06 / 53 · 矢量重排 · [原图来源](#)



图 7 N8N 工作流

鉴权未通过时进入“失败返回”。原截图中 Ollama 节点带有警示标记；该图展示工作流结构，不能据此认定所有节点已成功运行。

资料 07 / 53 · 矢量重排 · [原图来源](#)

五、数据输入与处理对象

系统处理的数据主要包括：

- 1. query：用户搜索词。
- 2. title：笔记标题。
- 3. content：笔记正文。
- 4. file：图片或视频链接。
- 5. tag：话题标签。
- 6. 人工 score：人工评分结果。
- 7. AI score：模型计算结果。

N8N 中通过 Code 节点对数据进行清洗，自动识别图片和视频文件链接。图片链接会被转换为模型可识别的 image_url 格式，视频链接会被转换为 video_url 格式。这样可以保证后续文本模型、多模态模型在统一的数据格式下工作。

项目过程文档中还记录了数据构建问题，例如业务数据抓取不完整、部分数据缺失结果信息、同张图片重复出现等。这些问题会直接影响 baseline 准确性，因此在项目中被列为后续优化点。

六、query 拆解框架

query 是整个评分流程的入口。系统首先需要判断 query 是否可理解，再判断它属于什么需求类型，并进一步拆解出主需、次需、实体和限制条件。



图 8 Query 分解框架

资料 08 / 53 · 矢量重排 · [原图来源](#)

1. query 可理解性判断

系统会先判断 query 是否符合基本语言表达规则，是否具有明确语义，是否能在正常认知范围内理解。如果 query 不可理解，则进入异常处理或降级流程。

2. 需求类型判断

项目中将 query 分为几类：

- 1. 单主需：查询中只有一个明确、具体的需求。
- 2. 主次需：一个主要需求附带一个或多个次要需求。
- 3. 泛多需：包含多个需求，但没有明确主次。
- 4. 无意义：无法识别有效需求。

在后续规则优化文档中，这套体系进一步整理为精准需求、多需求、泛多需求等分类方式。

针对 Query 词条的需求理解，原表分为以下三类。为避免横向表格字号过小，按“分类定义 + 具体示例”重新排版。

① 精准需求

用户搜索词的意图与需求已经非常清晰，能够明确判断什么结果更合适。需求意图清晰且相对唯一。

表 4 精准需求：Query 示例与含义

Query 示例	含义
围巾编织教程	寻找围巾的编织教程。
双鱼座性格	了解双鱼座性格。
鸡脯肉怎么做好吃	寻找鸡脯肉烹饪方法。

② 多需求：有主需、次需之分

Query 表意不单一，可能有多个意图；站内、站外需求存在差异，可区分主需与次需。分为“同一主体、需求维度多样”和“主体本身多样”两类。

表 5 多需求：主需、次需与主体歧义

细分类型	Query	原图说明
同一主体，需求多样	围巾	主体为衣物；需求可能为围巾推荐、系法、编织教程等。

表 5 多需求：主需、次需与主体歧义（续）

细分类型	Query	原图说明
同一主体，需求多样	长沙	主体为地名；需求可能为百科介绍、旅游攻略、资讯热点宣传片等。
同一主体，需求多样	奔驰	主体为汽车；需求可能为价格介绍、车型展示、优势对比等。
主体多样	可乐	常识理解为饮料可乐；原图指出站内需求可能指向大 V “可乐”。
主体多样	万宁	站外（百度、竞品）理解为地点；原图指出站内需求可能指向大 V “万宁”。
主体多样	苹果	可指水果、手机品牌，也可能指电影《苹果》。

③ 泛多需求：无主需、次需之分

有多个需求，但站内外没有明确数据指向，常识理解也没有明确指向。这里不是“无法理解”，而是无法确定哪一个需求更主要，因此各需求同等重要。

表 6 泛多需求：Query 示例与说明

Query 示例	原图说明
紫色	可能寻找颜色构成、紫色发色、紫色壁纸、紫色美食等，无法确定主次。
第 12 课	可能指语文第 12 课、数学第 12 课等，各种需求之间相对无法区分主次。

资料 09 / 53 · 矢量重排 · [原图来源](#)

需求明确时

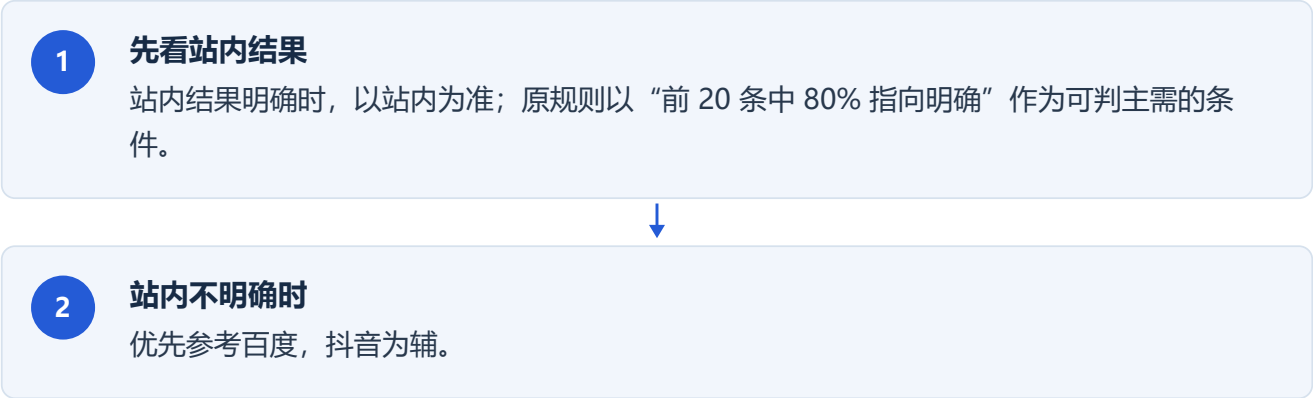


图 9 需求明确时的判断流程

需求不明时



图 10 需求不明时的判断流程

表 7 Query 需求判断案例

Query	原图结论	理解思路
积极入世不负青春	0 分	站内无好结果；百度有匹配辩论稿；笔记只给文案且无命中。
对娃	抛弃	若竞品平台均无明确结果，则抛弃。

资料 10 / 53 · 矢量重排 · [原图来源](#)

3. 实体提取

实体指真实世界中具有明确指代的人、地点、组织、品牌、产品、IP、作品名等。例如“潮汕站”“揭阳机场”“布老虎”“老虎膏”等都可能被识别为实体。

实体提取的意义在于：如果 query 明确指向某个实体，笔记内容必须真正命中该实体或其强关联内容，否则不能简单因为同类词出现就判高分。

4. 限制条件提取

限制条件包括时间、空间、数量、程度、对象属性、逻辑关系、格式、类型等。例如：

- “附近”是空间限制。
- “英文文案”包含语言和内容形式限制。
- “背后的拉链底部翘起”包含部位和问题描述限制。
- “使用方法和功效”包含信息类型限制。

限制条件是影响评分的重要先决条件。若笔记未满足核心限制条件，即使主题相关，也不能直接判高分。

七、规则体系设计

《新社区搜索相关性-规则优化整理》文档将业务规则以截图形式完整整理。该文档主要包含项目背景、分值介绍、query 分类、query 需求判断、笔记匹配规则、模型分析流程、新旧规则映射和完整判分流程。

社搜相关性项目经历 V1 → V4 版本规则变更，针对历史情况整理、整合规则，以提高实操指导性与可落地性；后续规则以此版本为准，文档持续优化更新。

项目背景

用户以 Query 表达想搜索的内容，但知识储备、认知水平、表达差异等原因可能导致搜索词无法准确表达真实需求。搜索系统需要理解真实需求，匹配并展示合适笔记；本项目对特定 Query 与笔记的搜索关联程度打分。

表 8 规则优化的配套材料

原图所列配套材料	用途
社搜 V4 更新同类型疑难 case	收集疑难案例。
新社区搜索相关性-规则同步记录文档	同步规则变更。

原截图只显示文档名称，没有可提取的实际链接。此处保留名称，不构造不存在的链接。

资料 11 / 53 · 矢量重排 · [原图来源](#)

分档依据：Query 与笔记主题的匹配程度。Q 表示 Query，N 表示笔记。

表 9 相关性分值定义

分值	定义
3	满足 Query 主要需求，且主题完全匹配。
2	满足主要需求且主题部分匹配；或满足次要需求且主题至少部分匹配。
1	满足程度低，但有参考意义。
0	不满足需求，但笔记与 Query 存在关联。
-1	不满足需求，且笔记与 Query 无任何关联。

原图案例重排

表 10 相关性分值案例

分值	Query	笔记示例
3	篮球	篮球是谁发明的；米球王高端球局、解析进攻细节。
3	苹果	吃过苹果没见过苹果树；iPhone 16 开箱机模；苹果发布会的两位常客。
3	雪纳瑞和泰迪对比	泰迪聪明听话、雪纳瑞活泼好动，到底该养谁。
2	雪纳瑞和泰迪对比	五种好养还不臭的狗狗。
2	篮球	NBA 篮球头像；CBA 篮球宝贝；巴黎奥运会篮球内场公开。
2	苹果	范爷成名后最想删除的电影；折纸小苹果。
1	昭觉寺	孕妇能去寺庙吗。
1	篮球	有两个姐姐给加油好像更容易进球。
1	八方旅人 2 灵王之杖	八方旅人 2 武器大师获取攻略。
0	八方旅人 2 灵王之杖	八方旅人 2 玩了一半。
-1	八方旅人 2 灵王之杖	骑行新手：山地车或公路车怎么选。

资料 12 / 53 · 矢量重排 · [原图来源](#)

理解笔记主题时看两个方向：整体表达的主题，以及笔记涉及的具体实物。

表 11 笔记主题与具体实物

笔记示例	主要主题	具体实物
618 必入：提升幸福感的家居卫浴好物分享	家具卫浴分享	浴室柜、智能马桶、花洒。
五种好养还不臭的狗狗	好养的狗狗介绍	雪纳瑞、泰迪、比熊犬、柴犬、博美。

表 12 笔记匹配程度与 Query 示例

匹配程度	判定	对应 Query 示例
完全匹配	Query 等于笔记主要主题。	家具卫浴推荐；哪种狗狗好养。
部分匹配	Query 命中某一具体实物；或命中部分主要主题，但占比有限。	花洒；泰迪。
不匹配	Query 与笔记主要主题、具体实物均不匹配。	装修吊顶推荐；袋鼠好养吗。

资料 13 / 53 · 矢量重排 · [原图来源](#)



图 11 标注流程

资料 14 / 53 · 矢量重排 · [原图来源](#)

原图按需求满足程度与主题占比 x 判断。原图顶部中间不等式漏写符号，下面按同图分支明确写出的“ $10\% < x < 80\%$ ”重排。

表 13 整体分档条件与判定分支

分档	条件	分支说明
3 分	满足主需； $x \geq 80\%$	精准需求命中主要主题；泛多需求中的任意需求命中主要主题；多需求中的主需命中主要主题。
2 分	主需： $10\% < x < 80\%$	主需满足，但主题仅部分匹配。
2 分	次需： $x > 10\%$	次需满足，主题可以完全匹配（ $\geq 80\%$ ）或部分匹配（ $10\% < x < 80\%$ ）。
1 分	满足程度低； $x \leq 10\%$	不分主次需。包含少量主题命中、仅提及但有帮助、对 Query 的求助等情况。
0 分	不满足，但存在关联	关键词同义命中且有展开；或无关键词命中，但属于同一大类。
-1 分	不满足，且无任何关联	需求和关联性均不满足。

求助例外：若 Query 本身是问答类，而笔记同样只是在求助提问，则不按普通求助帖处理，给 0 分。后续细则对 0 / -1 分还存在更新，保留其版本层次。

资料 15 / 53 · 矢量重排 · [原图来源](#)

前提：-1 分与 0 分都不满足用户需求。提供的类目仅供参考，明显错误时需要自行纠正。

多 term Query：先判断重要词是否命中

表 14 多 term Query 的 0 / -1 分判定

情况	0 分条件	否则
名词同义命中	命中的词同义且围绕该词展开占比 $> 10\%$ ；或者笔记与原 Query 类目相同。	-1 分
名词没有命中	笔记与原 Query 类目相同。	-1 分

单 term / 单一专有名词

表 15 单 term 与单一专有名词的处理

情况	处理
同义命中，且仅沾边	0 分。
未命中，Query 是名词	笔记与 Query 类目相同给 0 分，否则 -1 分。
未命中，Query 是其他词	泛义词情形原图认为可能不存在，列为 0 分；非泛义词为 -1 分。

原图边界案例

表 16 0 / -1 分边界案例

Query / 情形	判断依据
dorinocoffee 老板夫妻	命中“老板”且围绕老板介绍超过 10%，可进入 0 分。
陌生人给我钱背后故事	只命中“钱”，没有展开到 10%，为 -1 分。
基金申请书	出现“基金”但非同义，不能按同义命中。
探歌 / 脚垫	Query 与笔记同为汽车类时为 0；若对应美食等不同类目，则 -1。
月底入职还可以交当月社保吗	即使给定 Query 类目为商业财经 / 社科，也应识别其工作、职场属性。

本页来自早期思维导图；后文“类目判断（最新）”对类目优先级有补充。

资料 16 / 53 · 矢量重排 · [原图来源](#)

评分分档

系统最终将笔记与 query 的相关性映射为分值：

- 3 分：满足主需求且主题完全匹配。
- 2 分：满足主次要需求且主题至少部分匹配。
- 1 分：满足程度低但有参考意义。
- 0 分：不满足需求，但 note 和 query 有关联。
- -1 分：不满足需求，且 note 和 query 无任何关联。

- -2 分：无法查看或不参评。

项目早期使用“有效内容占比”进行计算，例如有效内容占比大于等于 80% 判 3 分，大于等于 10% 判 2 分。后续优化中，将有效内容占比进一步抽象为“有效内容等级”，减少模型在字数统计和比例计算上的不稳定。

原图给出的最终评分为 1 分：满足部分限定主需。主需是“获取香港三天两夜旅游的具体行程安排”；次需为美食、购物、景点、住宿推荐。分类：信息类 → 资讯类 → 旅游类 → 长期性。

表 17 案例中的 Query 条件满足情况

Query 条件	是否满足
香港	满足
三天两夜	未满足
旅游	满足
美食 / 景点 / 住宿 / 交通	满足
购物	未满足

表 18 正文内容占比与原图标记

正文内容	原图占比	原图标记
花销	16%	有效
住宿	7%	有效
景点	12%	有效
路线	7%	有效
美食	21%	有效
广告	10%	无效

表 19 图片与视频的占比明细

模态	原图明细
图片	图 1 至图 6 均为香港街道路牌，标记有效；图中汇总为 100%。
视频	1s-12s：出行目的（20%）；13s-33s：各景点（46%）；33s-57s：酒店（13%）；57s-70s：广告（17%，无效）。

原图汇总写“内容 67%、图片 100%、视频 80%”，但正文有效项相加为 63%，视频有效项相加为 79%。重排保留这些原始数值并指出差异，不将其当作校验通过的计算结果。

资料 17 / 53 · 矢量重排 · [原图来源](#)

八、核心工作流 SOP

项目过程文档中将作业 SOP 拆为五个步骤：

- 1. query 变量结构化提取。
- 2. 遍历变量，进行变量与笔记的双向匹配。
- 3. 处理匹配结果，将文本、图片、视频三维度结果合并。
- 4. 根据匹配结果计算内容得分，形成三维度加权得分模型。
- 5. 根据得分输出最终结果，进入结果分级输出机制。

Step 1：query 变量结构化提取

输入原始 query，输出结构化 JSON：

```
{
  "可理解": true,
  "限制条件": [],
  "实体": [],
  "需求": {
    "主需": [],
    "次需": []
  },
  "分析说明": ""
}
```

Step 2：正文、图片、视频分别匹配

系统分别调用不同 Prompt 判断：

- 文本是否满足意图列表。

- 文本是否命中实体。
- 文本是否满足限制条件。
- 图片是否满足意图列表。
- 图片是否命中实体。
- 图片是否满足限制条件。
- 视频是否满足意图、实体和限制条件。

每个节点都要求模型输出 JSON，包含 result、list、有效内容数量或比例、是否有关联、分析说明等字段。

Step 3: 三维度结果合并

系统合并文本、图片、视频的匹配结果，得到：

```
{
  "限制条件": true,
  "实体": true,
  "需求": "满足/部分满足/不满足",
  "是否有关联": true
}
```

Step 4: 计算内容得分

系统根据限制条件、实体、需求满足情况和有效内容等级计算最终分数。若限制条件、实体和需求都满足，则根据有效内容等级判断 3 分、2 分或 1 分；若不满足需求但有关联，则进入 0 分；若完全无关联，则进入 -1 分。

九、案例展示：多模态样例与规则应用

项目过程文档中使用多个具体 case 展示规则如何落地。

案例 1：布老虎信息可视化

该 case 用于说明 query 拆解、多图片内容理解和信息可视化需求匹配。query 会被拆成“获取布老虎相关信息、整理分类布老虎信息、选择可视化工具、设计呈现形式、完成信息可视化”等多个需求方向。系统需要判断笔记中图片和正文是否真正围绕布老虎知识、分类、展示方式展开。

原站目前只返回“图片可能违规或链接失效”提示，没有可用原图。此位置保留状态说明，不生成无法核实的替代内容。

资料 18 / 53 · 图源状态 · [原图来源](#)



图 12 布老虎案例素材 2

布老虎图版中的可读信息

表 20 布老虎图版中的可读信息

维度	图中内容
造型	头顶“王”字；招财耳；眉毛像火苗；大眼睛、大嘴巴、大尾巴。整体以夸张造型表达勇猛神态。
寓意	头顶王字、憨态可掬；尾巴硬长、造型雄壮；鱼形纹样寓意年年有余，星形纹样寓意平安。
装饰	眼睛类型、虎身纹饰、头部不同色彩搭配。头部鲜艳，眼睛大而圆，眉毛造型不同，颜色对比强烈。
颜色	红、橙、黄、绿、蓝、黑、白。
制作步骤	设计花纹 → 初步缝合 → 塞入棉花 → 缝合开口 → 缝制其余部分 → 装饰整体。
图版布局	上部为分析与做法，下部为布老虎成品欣赏。

资料 19 / 53 · 原始图像 · [原图来源](#)

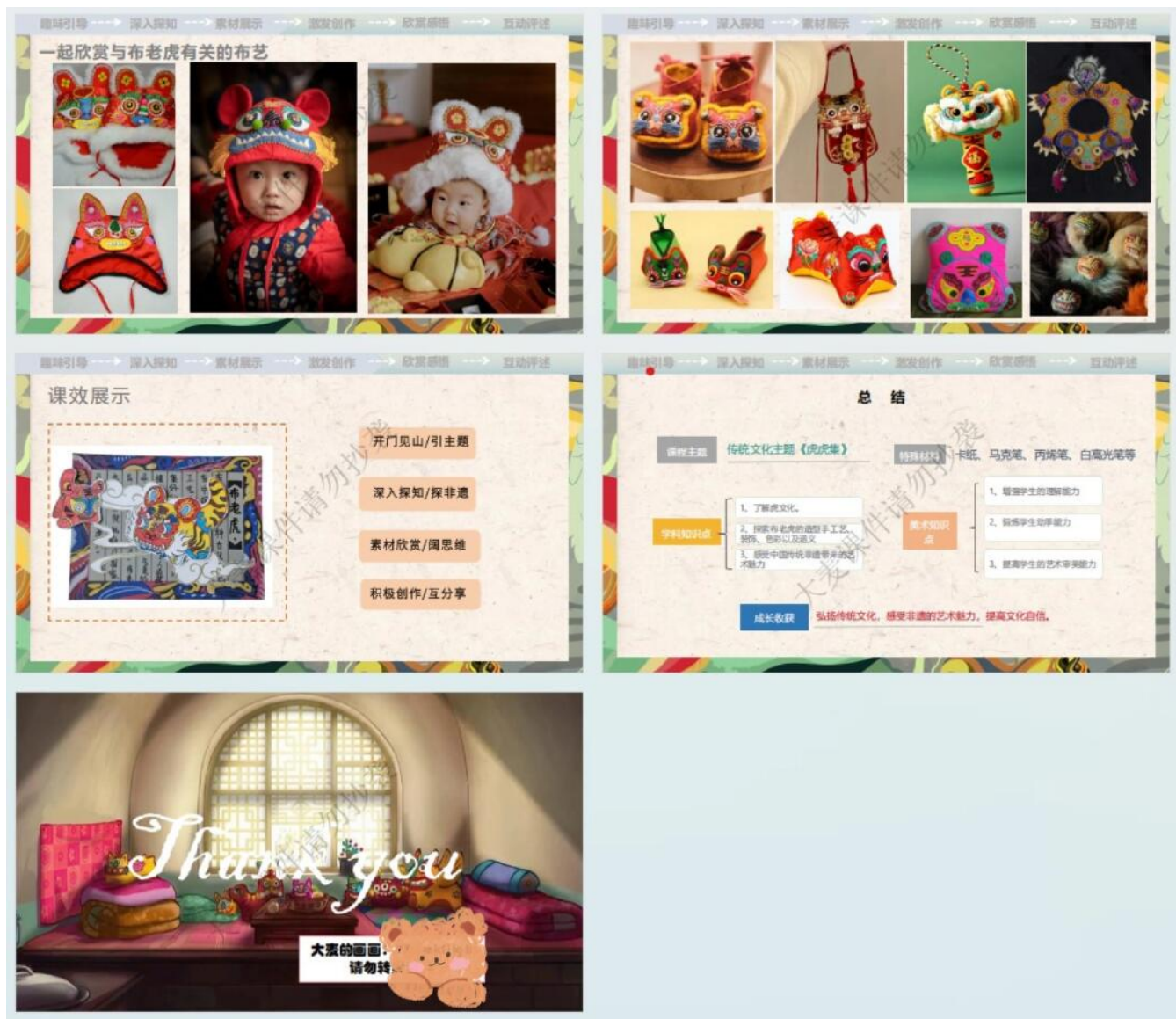


图 13 布老虎案例素材 3

布老虎教学图版：信息重排

图版展示与布老虎相关的帽饰、鞋、挂件、包袋等布艺，随后是课效展示、课程总结与结束页。



图 14 布老虎教学流程

可辨读课程总结：了解虎文化、培养学生创作与审美能力；弘扬传统文化，感受非遗的艺术魅力，提高文化自信。图版中极小、无法可靠辨认的课程细项未补写。

资料 20 / 53 · 原始图像 · [原图来源](#)

案例 2：潮汕站到揭阳机场

该 case 展示了 query 中实体和时间限制的复杂性。query 包含“潮汕站”“揭阳机场”“15:02 到站”“17:10 飞机起飞”“五一期间”等信息。系统不仅要识别地点实体，还要理解交通方式、时间是否来得及、节假日交通影响等隐含判断。



图 15 潮汕站到揭阳机场案例图片

资料 21 / 53 · 原始图像 · [原图来源](#)

十、规则优化与细分类目展示

规则优化文档中还整理了大量细分类目、判分边界和案例。由于原文档内容以截图为主，以下图片全部保留作为规则体系附录，可在作品集中按“规则文档展示”章节插入。

这张原图把 0 / -1 分判断展开为表格，与前文思维导图对应。首先纠正明显错误的类目，再看 term 的同义命中与展开程度。

表 21 0 / -1 分细则

Query 结构	条件	结论
多个重要 term / 专有词	名词同义命中，且围绕 term 展开 > 10%；或与原 Query 同类。	0 分
多个重要 term / 专有词	命中词但不满足上述条件。	-1 分
多个重要 term / 专有词	名词无命中，但与原 Query 同类。	0 分
多个重要 term / 专有词	名词无命中，且不同类。	-1 分
单 term / 单专有词	同义命中但仅沾边、只是提到 Query。	0 分
单 term / 单专有词	没有命中，但名词类目相同。	0 分
单 term / 单专有词	没有命中，且名词类目不同。	-1 分
其他词（动词 / 形容词 / 副词）	近义词情形。原图注明“可能没有这种情况”。	0 分
其他词	非近义词情形。	-1 分

可读案例

表 22 0 / -1 分案例要点

Query	图中案例要点
北京	笔记只提到北京，属于沾边，0 分。
了不起的女孩	作为专有名词，笔记若同为电视剧类，按同类判断；若是化妆教程等不同类，-1 分。
帅气	图片中的相关词若实际为“出温柔”等不同含义，按非近义词处理。

原截图右侧案例栏部分越出画面；这里只重排可辨认的规则及案例，不补写被截断的笔记内容。

资料 22 / 53 · 矢量重排 · [原图来源](#)

含限定词 Query

适用于主观、客观（含时间）、事实类限定词。限定词丢失最多给 2 分；原图取消“主观无争议直接 3 分”的旧规则。

表 23 含限定词 Query 的判定

限定词情形	规则	原图示例
不影响语义的限定词	可忽略。	最火、最新、必买、常用。
主观限定词	结合 Query 与笔记判断。	高级感、干净、个性、清冷、气质、性感、温柔、冷淡、小众、ins 风、治愈、百搭、氛围感、可爱、简约、冷门、冷淡风、简单、好吃、美味、惊喜。
常规限定词丢失 / 不完全匹配	若从需求角度仍适用，给 2 分。	订婚手捧花 → 结婚手捧花；订婚花束心形 52 朵 → 未写订婚，但可适用。
限定词转义	最多 0 分。	胖女生穿搭 → 偏瘦女生穿搭；矮个子穿搭 → 高个子穿搭；推荐 → 避雷；注册 → 注销。
事实类限定词丢失	不扣分。	杭州西湖旅游 → 西湖旅游。西湖在杭州是客观事实。

表 23 含限定词 Query 的判定（续）

限定词情形	规则	原图示例
同义词条	有依据可证明同义时不扣分。	超萌少女：原例参考百科 / 词义依据判同义，3 分。
时间限定词丢失	无法确认是否符合所需时间，扣 1 分，给 2 分；明确矛盾则 0 分。	2024 春新款 → 未明确年份，2 分；明确前年的款式，0 分。

温柔发型：四种判分

表 24 “温柔发型” 四种判分

笔记内容	评分	理由
明确为温柔发型	3	完全满足。
仅介绍发型，观感偏温柔	2	主观限定词未明确体现。
甜美发型	2	与温柔有差异，但并非完全不同。
高冷发型	0	限定词含义发生转义。

表情包需求给到可适用的头像，可给 2 分。若“可爱”与内容明显转义，则 0 分。

资料 23 / 53 · 矢量重排 · [原图来源](#)

对比类 Query

定义：比较两个或多个对象、概念、产品、服务等。

表 25 对比类 Query 的判定

笔记覆盖情况	评分	示例
提供 1 个或以上对比项，但不完整	1	雪纳瑞和哈士奇 → 只讲雪纳瑞；淡水珍珠和海水珍珠的区别 → 只讲 akoya / 澳白；阿玛尼 206、400、405 → 只讲 206、405。
提供全部对比项	2 或 3	按主题匹配 / 篇幅占比决定。原图未提供具体案例。

排行类 Query

表 26 排行类 Query 的判定

笔记内容	评分	行李箱排行示例
至少 3 个对象，且明确排序	3	三款行李箱明确第一、第二、第三。
少于 3 个对象的明确排序	2	只介绍一款“最实用”的行李箱。
隐含排序	2	“三款最实用的行李箱”，但未逐一排次。
仅推荐，没有排序	1	介绍三款行李箱，但没有排名。

资料 24 / 53 · 矢量重排 · [原图来源](#)

含颜色需求

不再区分颜色是否大众。文字完全匹配且图片符合，给 3 分；文字不完全匹配或未提及颜色时，按图片判断：相似 2 分，不相似 0 分；无法判断可从宽。

表 27 颜色需求的判定

Query / 笔记	分值	判定
粉拖鞋穿搭 / 粉色运动拖鞋，图片符合	3	文字与颜色一致。
苹果表带薰衣草紫 / 芋泥紫	2	颜色相似。
雾霾蓝针织上衣 / 水蓝色	2	图片颜色相近。
藏蓝色背带裙 / 藏青色	2	图片颜色相近。
正红色大衣 / 未写正红色	2	图片颜色相似。
藕粉色旗袍穿搭 / 未写颜色	2	从图片判断相近。
橙红色喜嫁 / 描述为橙色	2	图片相似。
蓝色 / 藏蓝色	3	藏蓝属于蓝色。
藏蓝色 / 仅写蓝色	2 或 0	图片相似 2，不相似 0。

问答类 Query

表 28 问答类 Query 的规则与示例

形式	规则	示例
直接满足	显式给出答案，至少 2 分，结合主题占比判 2 或 3。	iPhone 13 Pro 尺寸：完整围绕尺寸讨论 3 分；仅在正文提到答案、占比较低 2 分。
推理满足	没有直接回答，但能推理得到答案，至少 2 分。	“百香果和玫瑰花可以一起喝吗” → 笔记展示两者泡茶，2 分。
推理满足案例	原标注示例，保留其判分而非医学结论。	“吃母乳容易肠胀气吗” → 原例内容涉及母乳喂养妈妈饮食与婴儿胀气，图中判 2 分。
显式问答	Query 直接提问。	拼多多上的胸贴能买吗。
隐式问答	原规则当前仅限定为答案是数字的 Query。	蛋挞的热量：笔记给出热量数字，原图判 2 分，不考虑占比。

资料 25 / 53 · 矢量重排 · [原图来源](#)

商品类：三种关系

表 29 商品类 Query 的三种关系

关系	评分	规则
同类、可替代	0	A、B 为不同实体，但性质相同或可替代。
不同类、不可替代	-1	A、B 为不同实体，且不属于同性质 / 可替代商品。
配件关系	1	搜索实体 A，给到配件 B；B 需与 A 一起使用，或 A、B 为搭配使用关系。

同类可替代：原图均判 0 分

唇釉 / 唇泥 / 固体口红；粉底液 / 气垫 / 粉饼；眼线液笔 / 眼线膏；牙膏 / 洁牙粉；洗衣液 / 洗衣凝珠 / 洗衣粉；华为手机 / 苹果手机；棉服 / 羽绒服；针织衫 / 毛衣；不同品牌辣条；牛仔裤 / 卫裤；长裤 / 短裤；帆布鞋 / 运动鞋；红宝石项链 / 珍珠钻石项链。

不同类商品：原图均判 -1 分

洗衣液 / 柔顺剂；补水面膜 / 补水面霜；一次性胶片机 / 微单相机；棉服 / 毛衣；棉服 / 大衣；上衣 / 裤子或鞋子；面包 / 蛋糕；口红 / 润唇膏；水 / 乳液；华为手机 / 华为平板。

配件案例

表 30 商品主体与配件案例

主体	配件或结果	评分
吸尘器	地刷、滤芯	1
手机 / 平板 / 耳机	对应壳、膜、挂绳	1
散粉	粉扑	1
眼影	眼影刷	1
隐形眼镜	隐形眼镜护理液	1
小白鞋	小白鞋清洁剂	1
沙发	沙发套	1
沙发	毛毯，即使笔记称当作沙发套使用	-1

重要例外：当 Query 不是搜索实体 A 本身，而是询问 A 的测评、推荐、对比等需求，笔记仅给配件 B，不适用配件 1 分规则，应按类目判断。

表 31 商品类 Query 补充案例

Query	笔记	评分
苹果 15	苹果 15 手机壳	1
苹果手机测评 / 苹果手机种草推荐 / 苹果与华为手机对比	苹果手机壳	-1

账号类规则的生效条件

Query 与笔记内容无关，仅与发布账号有关时，才应用账号规则。主需不指向唯一、明确账号时，账号匹配按次需处理，最多 2 分。

表 32 账号类 Query 的处理规则

情况	原图规则
主需不是唯一明确账号	账号完全匹配 / 近似完全匹配：2 分；部分匹配：1 分。
主需就是该账号	给到完全匹配的账号；或给到该账号博主的相关内容，按原需求判断。
同一人物的其他账号	次需，2 分。
非同一人物的其他账号	同类目 0 分；不同类目 -1 分。
中英文差异	可视为等价关系。
账号内特殊符号	一般可忽略；Query 本身的字符需另行判断。
Query 字符是否能忽略	站内有匹配网红账号（原图阈值：粉丝 500 以上）时，字面符号不能忽略；粉丝 500 以下则考虑非强账号需求。
Query 非常精确	例如符号很多且确有对应结果，可不考虑粉丝量，视为强账号需求。

案例要点

“古月老师”等可能首先指向人物而非同名账号，需要先辨别主需；“小田装修设计”若明确找账号，给到另一个装修账号为同类 0 分，给到美食账号则 -1 分；“南宁大聪明”示例中，原图通过视频角标归属辨认相关博主内容。

原截图右侧多处文字截断，部分案例的完整句子、分值不可辨认。上面仅列可确认的规则，不补造缺失结论。

美食类 Query

适用于寻找某道菜、食谱、做法教程、食材等需求。

表 33 美食类 Query 的规则与案例

情形	规则 / 分值	原图案例
食谱是一个整体，只给单道菜	1 分	一岁宝宝的一日三餐食谱 → 苹果山药饼做法；生酮食谱 → 酱油渍做法。
“大全” 数量规则	原图标为待确认：1 个 1 分；2-3 个 2 分；4 个以上 3 分。	家常菜做法大全 → 辣椒炒肉 1 分；13 种家常菜做法 3 分。
多种食材，只命中部分主食材	0 分	韭菜豆腐盒子 → 韭菜盒子；淮山党参鹌鹑汤 → 淮山鹌鹑汤；黄瓜炒虾仁 → 黄瓜炒蛋；白菜包子 → 猪肉包子。
食材和做法相似，有帮助	可给 1 分；原图强调不绝对、需要更多案例验证。	盐肉烧笋 → 春笋炒肉片；仔姜鸭 → 老姜 + 母鸭；大白菜炒肉 → 娃娃菜类似做法。
笔记是 Query 的子需求，主需完全满足	3 分	胡萝卜 → 胡萝卜炖牛肉；排骨汤 → 山药排骨汤。

资料 28 / 53 · 矢量重排 · [原图来源](#)

旅游类 Query

表 34 旅游类 Query 的判定

需求 / 情形	评分	原图案例
旅游攻略提供 1-2 个维度	2	宜昌下午半日游 → 三峡大瀑布；北京一日游 → 环球影城。
旅游攻略提供 3 个及以上维度	3	青岛游玩攻略 → 多景点介绍。
行程天数不满足，但有帮助	1	三天两夜 → 两天一夜；香港旅游攻略 3 天 2 夜 → 实际天数不满足。

文案类 Query

侧重点是文本，包括正文、图片和视频中的文字。图片画面内容本身不是主要判断对象。

表 35 文案类 Query 的判定

情形	评分	原图案例
文本内容符合文案需求	3	风景文案 → 正文为风景文案、配风景照片；油菜花文案 → 正文和图片 OCR 均为油菜花文案。
未明确标注文案，但可作为文案使用	2，次需	牛仔裤文案 → “窄版直筒裤我的神！！”；“真的巨遮肉，巨显瘦”。

资料 29 / 53 · 矢量重排 · [原图来源](#)

明星 IP

表 36 明星 IP 的判定案例

Query	笔记	原图评分
龙卷风周杰伦	邓紫棋演唱《龙卷风》；相关但未满足指定歌手	0
羽绒服肯豆	展示肯豆风格羽绒服，满足需求	3

品牌、影视、游戏、动漫 IP

专有名称通常不拆分；若涉及简称，可算次需。简称需通过站内简单搜索验证，指向唯一才认为是简称。中英文组合名称的中文部分也可能属于简称。

表 37 品牌、影视、游戏与动漫 IP 的判断

Query / 情形	原图判断
冬眠 → 国王在冬眠	IP 不可拆分，“冬眠”不是该剧简称：-1 分。
atm → 漫步者 Atom Max 小金标耳机	圈内简称且可验证命中，次需 2 分。
保罗 → 保罗彼得	品牌指向不唯一，不能认定简称：-1 分。

表 37 品牌、影视、游戏与动漫 IP 的判断（续）

Query / 情形	原图判断
喵呜 → MiaouBistro 喵呜	中文部分为店名简称，非随意拆词，次需 2 分。
头七	原例站内前排约 80% 指向头七怪谈游戏，判 3 分。
巨蟒	主需为蟒蛇；游戏中以“巨蟒”称某武器，圈内简称作为次需 2 分。
好东西	主需明确为电影。命中字面“好东西”的好物分享可作次需；无命中且类目不一致为 -1。
加油妈妈	主需为电视剧；小孩给妈妈加油且命中词条，次需 2 分。
探险家学院	主需为书籍；同词课程计划，次需 2 分。

资料 30 / 53 · 矢量重排 · [原图来源](#)

人物类 Query

适用于一个或多个姓名。先区分人物本身、作品、翻版、同名人物与社会关系。

表 38 人物类 Query 的评分规则

情形	原图评分	示例 / 解释
未围绕姓名展开，但主题与人物有关	1	卜凡形象在视频出现；董洁推荐过的饼干；杰克隽逸分享过的代餐；与汪苏泷建立关联的 DIMOO 手办。
未围绕姓名展开，主题也与人物无关	0	搜索汪小菲，笔记重点为大 S 身故及生平。
多个人名只出现其一	0	凌忆唐书欣 → 凌忆；倪妮刘亦菲 → 倪妮。
多人同时出现、无法分割	3	综艺双人画面；讲某人的事情且人物不可分割等原例。
名人的作品	2，次需	杨丽萍导演的作品；村上春树的书；苏西书签会。
作品的翻版	1，提及	汪苏泷歌曲翻唱；临摹 Nikita Busyak 的画。
同名人物较多	任意匹配者可 3	Ben、黄烦烦等泛多需案例，需围绕对应对象展开。

表 38 人物类 Query 的评分规则（续）

情形	原图评分	示例 / 解释
姓名拆分	姓与名可分，名字不再拆	“美子”与“樋口愉美子”中的“愉美子”不能再次拆出美子，原例 -1。
社会关系 / 同垂类人物	0	周杰伦 → 昆凌（夫妻）；周杰伦 → 林俊杰（同为歌手）。
不同垂类人物	-1	周杰伦 → 金庸（作家）。
搜索人物，给到本人账号	3	即使人名和账号名不一致，只要是该人物本人账号。

资料 31 / 53 · 矢量重排 · [原图来源](#)

视频 / 图片 / 音频需求

表 39 视频、图片与音频需求的判定

需求	规则	原图例子
明确要视频	必须有视频才满足；只有图片给 0 分。	xxx 视频 → xxx 图片。
明确要图片	视频可通过截图满足部分需求，按主题 / 篇幅给分，最多 2 分。	xxx 图片 → xxx 视频。
强音频需求	音乐、有声读物等主要看音频，可忽略画面。	爸爸去哪儿主题曲：原例给 3 分。

POI：明确地点需求

表 40 POI 明确地点需求的判定

情形	规则 / 评分	原图案例
旅游周边	地点客观属于所需周边可满足。	北京周边游 → 河北承德吃喝攻略，虽未直说北京周边，原例给 3 分。
POI 不满足，但属同类	0 分	香港青衣酒店 → 香港非青衣地区酒店； 某商场美食 → 同城其他广场美食； 老满客粥铺 → 三米粥铺。
连锁店	不满足指定 POI 的不跨国连锁店，原图列 0 分。	按图中条件保留。
POI + 旅游	同省不同市旅游不满足，0 分。	不能只因同省判满足。
具体小范围景点	主需为旅游攻略，局部关联内容可作次需。	沙面岛 → 岛内某个美食点，2 分。
景点名命中但需求不符	0 分	徐州云龙湖 → 湖边楼盘介绍。
学校地点	主需为学校攻略 / 介绍；租房可作次需。	建大 → 周边租房，2 分。

资料 32 / 53 · 矢量重排 · [原图来源](#)

歌曲类：默认需求为原版

表 41 歌曲类 Query 的判定

情形	评分	原图案例
原曲无官方 MV，创作视频符合风格、氛围	3	《星月落》自创剪辑。
原曲有官方 MV，笔记是自创 MV	2，次需	《后来的我们》。
二创：变速、降调、换词	2，次需	《怪咖》降调；《群青》变速。原图《有何不可》换词例的分值漏字，按本行通则列示。
词 / 谱 / 曲 / 伴奏	2	《星月落》琵琶伴奏；Love Story 曲谱；《青花瓷》歌词。
未指定歌名，默认仍要原版	翻唱 2	盗墓笔记歌曲 → 《十年人间》的翻唱。
需求为翻版，给到原版	1	“我的梦” 对应翻版需求，笔记给英文原版。

壁纸类

需求明确为某类壁纸时，笔记应体现壁纸用途。未明说“壁纸”，但图片可当壁纸使用，原图给 2 分。案例包括“崔然竣壁纸”“刺客信条奥德赛壁纸全屏”。

资料 33 / 53 · 矢量重排 · [原图来源](#)

纯数字类 Query

表 42 纯数字类 Query 的判定

情形	评分
无命中	正常判断类目是否一致。
命中，且数字单独存在	3 分。
只是更大数字或编号的一部分	2 分。

原图以“475”为例：属于编号一部分为 2 分，独立数字为 3 分。

穿搭类 Query

表 43 穿搭类 Query 的规则与案例

情形	规则 / 分值	案例或说明
春秋与春 / 秋	互相匹配为 3 分。	原图按春秋可通用处理。
春夏与春 / 夏	互相匹配为 2 分。	春夏飘逸半裙 → 仅夏季穿搭。
秋冬与秋 / 冬	互相匹配为 2 分。	外套秋冬 → 仅秋天穿搭。
季节不匹配	看实际是否可适用。	原图仍有待补充案例。
体重 + 穿搭	差 0-3 给 2 分；3-6 给 1 分；大于 6 给 0 分。	120 斤 → 118 斤穿搭，2 分。原图在边界 3 处有区间重叠，未另给消歧规则。
要穿搭，笔记是单品	有穿搭展示 1 分；无展示 0 分。	草原穿搭 → 拍照主题但有穿搭图，1 分；鞋子搭配 → 仅鞋子介绍，0 分。
穿搭占比	满足指定商品的穿搭时，不单独扣其他搭配单品占比。	白色衬衫 + 棕色格子百褶裙穿搭，加背心仍可满足；若只要两件单品但给三件则另按单品占比。
少量无关文案	仅一两句、篇幅很小可不扣分。	穿搭笔记附一句天气或成长感悟。

原截图右侧部分案例句尾截断；这里只保留可确认规则，不推测未显示的分值。

资料 34 / 53 · 矢量重排 · [原图来源](#)

医疗类 Query

原规则从严处理：梗图形式不能算满足医疗需求或提供参考。案例“中医把脉能看出什么”：即使梗图有相关词，原例仍给 0 分。本节是历史相关性标注规则，不是医疗建议。

翻唱 / 跳 / 拍类内容

表 44 翻唱、翻跳与翻拍内容的判定

情形	评分	原图案例
某形象的 cos / 仿妆	2，次需	唐僧 cos；宝钗仿妆；苏在在角色造型。
网络热梗需求给翻拍	2，次需	潘周聃梗翻拍。
歌曲 + 舞蹈，需求为原版	翻跳 2	Queencard 舞蹈。
人物 + 舞蹈，人物版本是原版	他人翻跳 2；本人可 3	一冷编舞：翻跳 2，本人跳 3；张艺兴舞蹈：翻跳 2。
明确要某人跳舞	他人跳只能弱满足	张艺兴跳舞 → 其他人跳张艺兴舞蹈，1 分。
歌曲带人物 IP	按 IP 标准	SEVENTEEN 孙悟空 → 视频只有组合一人，原例 1 分。
歌曲没有人物 IP 限定	按需求标准	Cruel Summer → 粉丝翻唱，次需 2 分。

资料 35 / 53 · 矢量重排 · [原图来源](#)

图片、正文参评与占比

表 45 图文参评与内容占比规则

条件	规则	案例
笔记明确多个主题	图片可分开计算占比；主题有争议时按原 V3 方式。	杨幂耳钉：笔记明确有耳钉与红裙两个主题，2 分。
Query 无具体实体	图片不参评。	八卦帖：原例 3 分。
非强图片 / 视频需求	正文通常都要参评。	视频相关但正文无关，原例按部分主题匹配给 2 分。
主题主要靠文本表达	文本更重要，图片权重相对弱。	郁金香几月份种植合适。
主题主要靠图片表达	图片更重要，文本权重相对弱。	北京旅游攻略。
图片与文本共同表达	综合两者判断主题。	一周年蛋糕。

满足需求与主题占比

表 46 需求满足与主题占比案例

案例	原图结论
迪丽热巴直播	视频大部分时间该人物发言，正文相关；其他人物类似背景，3 分。
无尽的冒险 / 熊跳舞	笔记主要强调自己的翻跳，按翻跳逻辑 2 分。
华为 P60 Pro	多款手机对比，目标主题占比在 10%-80%，2 分。
发票红冲操作步骤	主题占比 10%-80%，2 分。
术后注意事项（原图手术名称不清）	原图可确认的判分说明为主题占比 10%-80%，2 分。
适合拍宝宝的相机	主题完全匹配，结尾相机专业术语亦与宝宝拍照相关，3 分。

资料 36 / 53 · 矢量重排 · [原图来源](#)

Tag 参考规则

表 47 Tag 与笔记关系的处理

Tag 与笔记关系	处理	原图案例
完全一致	正常评估。	按实际内容判断。
一致性存疑	尊重笔记内容。	温柔文案：判断有分歧，原例 3 分；带黄字谐音网名：“小黄小黄，前途辉煌”，原例 2 分。
完全不匹配	忽略 Tag。	赵丽颖撞衫孙悟空：相关 Tag 与实际内容完全无关，原例 -1 分。

求助类笔记

表 48 求助类笔记的判定

情况	评分	原图案例
正文和图片都只求助，无可用信息	1	福州出发 → 找拼车；薄荷变紫色 → 求助原因；小天才 Z8 → 求助手表占线。
求助笔记中也有有用信息	2	计量经济学网课推荐：包含课程有用信息。
Query 为问句，笔记同样仅提问	0	苹果手机拍人怎么调参数 → 笔记也在求助同一问题。
产品需求，展示两种以上产品求助选择	扣 1 分，原例 2	MLB 帽子推荐：展示两款颜色；迪士尼发箍：展示三款问选哪个。

资料 37 / 53 · 矢量重排 · [原图来源](#)

出物 / 收物

Query 并非买卖需求时，原图说按求助类规则处理；其中“通则”与右栏后来补充的案例分值有不一致，以下分别保留。

表 49 出物、收物通则与更新案例

场景	原图通则	右栏案例 / 更新
常规商品，含推荐 / 测评需求	出现销售意图按次需给 2；重点若在销售，则按销售理解。	种草 / 测评 3；“种草 / 测评 + 售卖”更新案例也给 3；纯售卖示例给 1。
常规商品本身	满足内容需求 3；出现售卖 2；不满足最多 1。	种草测评 3；粉扑有强销售但伴随测评，更新案例 3；纯售卖黑色喇叭裤、闲置吉他 2。
非常规商品	出现出售 / 求购，按冷需给 1，不区分是否有用信息。	无畏契约账号出售、宠物蛇、拉布拉多幼犬：原例均 1。
Query 本身明确出物 / 收物	满足需求可 3。	诚心收 CCD：原例 3。

本页特意保留原材料的更新冲突，不能把所有“种草 + 售卖”机械统一为 2 分或 3 分。实际使用时应核对原业务规则的有效版本。

资料 38 / 53 · 矢量重排 · [原图来源](#)

“有帮助”指内容不能完全满足需求，但能提供有价值、指导性的建议。核心限定词缺失时，1 分与 0 分的区别在于是否仍对 Query 有用。

有价值：1 分

表 50 有帮助且有价值的笔记：1 分

Query	原图中的帮助点
舞台睫毛怎么贴	普通睫毛与舞台睫毛有差异，但粘贴教程仍有帮助。
金华银行卡异常怎么解除	没有金华限定，但给通用银行卡异常解除方法。
哑女怎么画全身	给出半身画法，对全身绘制仍有参考价值。
说我像一个演员，很喜欢怎么回复	笔记的回复内容确有价值。
迪庆拍藏族写真文案	均为藏族写真，可有价值；还需判断地点限定的重要性。

无价值：0 分

表 51 缺乏帮助价值的笔记：0 分

Query	不满足原因
湖北美食描写	仅泛谈口腹之欲，没有湖北美食描述。
假日文案高级	仅八月文案，没有假日内容。
考研座右铭高级小众	只有爱自己的“大女主”文案，没有考研内容。

资料 39 / 53 · 矢量重排 · [原图来源](#)

主需不明确时的类目判断

主需无法确定，但站内结果有较明确的 2-3 个类目指向，可按泛需求处理；笔记与其中类目一致给 0 分。原例包括显眼包（包包）、韩东的身世（影视）、夏津周昱（教育）。

0 / -1 分的类目细化



图 16 0 / -1 分的类目判断流程

表 52 类目判断的领域细则

领域	原图规则
医疗 / 学术	区分人与宠物；即使命中 term，也需看是否跨专业知识领域；常规药品与药食同源产品需区分。
美妆	原图此版本将美瞳与美妆、护肤品与彩妆 / 药妆视作同类；普通美妆与医美项目属不同类。
时尚	服装穿搭与眼镜、珠宝、包包等配饰属同类。
家居 / 家电	区分带电（3C）与不带电；二者不一致给 -1，一致按同类给 0。

原图右侧多处截断；次页“类目判断（最新）”又更新了实体优先规则。本文件按原顺序展示版本变化。

类目补充

表 53 类目补充案例

比较对象	原图结论
明确的三农与美食	不同类；带有美食属性的三农与美食可同类，如玉米与窝窝头，0 分。
人体护理洗液与隐形眼镜护理液	作用对象不同，非同类，-1 分。
美甲、美睫、美发	互相不算同类；美甲与美妆也不同类。
洗洁精与家居家电	不算同类。洗洁精、去污剂、洁厕剂、厨房抹布、垃圾袋按生活日化类。

原图标注“最新”的类目判断

表 54 实体优先与最新类目判断

规则	案例
双方都能提炼实体：按实体类目	显白腮红 → YSL 口红：-1；火锅店 → 拉面店：-1；米汤功效 → 蜂蜜：-1；草莓送礼 → 兔子玩偶：-1。
无法提炼实体：按核心领域	鱼缸造景 → 室内装修：-1；阿迪达斯 → 耐克 / LV：0；睫毛膏推荐 → 护肤流程：-1；查血常规 → 中药蒸制：-1；新西兰壁纸 → 伦敦壁纸：0。
影视综互通	同类 0：卡布达 → 蜡笔小新；好看的电影 → 《奔跑吧兄弟》片段。
需求含多个实体类目，命中其一	可按同类给 0，但该类内容占比需 > 10%。LV 外套和鞋子 → 耐克外套、项链、裤子；外套占比不足 10% 则 -1。
类目与拆词同时命中	优先类目：苹果手机 → 苹果平板，虽命中“苹果”，实体类别不同，-1。
不影响沾边 / 提及规则	狮王祛痘膏 → 主要是狮王洗面奶，但末尾提及配合祛痘膏的使用体验，可按提及给 1。

医疗领域特殊划分

以下仅转录原文的业务标注分类，不新增医学判断。药品、保健品、保健用品、医疗器械之间互相给 -1 分；同领域内不再细分。不同症状之间原图允许给 0 分。

表 55 医疗领域的特殊分类

分类	原图例子
药品	阿司匹林、布洛芬、青霉素、氯雷他定等。
药食同源例子	枸杞子、山药、红枣、莲子、百合、薏苡仁、菊花、桂圆、黄精、桑葚、橘皮、黑芝麻、生姜、当归、山楂；原图列为不同于普通药品的例子。
保健品	钙片、维生素、鱼油、褪黑素等。
保健用品	保健枕、按摩垫、足浴盆、艾灸贴、护腰带等。
医疗器械	血压计、血糖仪、体温计、心电图机、呼吸机、输液泵等。

评分示例：阿莫西林 → 布洛芬，0 分；999 感冒灵 → 足浴盆，-1 分。原图关于两类腰痛的案例右端截断，此处不补写缺失结论。

拆词判断

搜索结果多样时可拆词；需求明确、搜索结果大多一致时，作为专有名词，不能拆。

表 56 Query 拆词判断案例

Query	原图结论
火鸡工厂	结果既有游戏也有餐饮；可拆，命中“工厂”同义且展开 10%，0 分。
江南春糕点推荐	可拆，但两个“江南春”不同义，且类目不一致，-1 分。
大公鸡小酒馆	结果大多一致，专名不可拆；原例类目不一致，-1 分。
郑子新营养大观园	“营养大观园”属专名不可拆，未命中且类目不一致，-1 分。
那个不为人知的故事原型	《那个不为人知的故事》是影视剧名，不可拆；原例未命中且类目不一致，-1 分。

沾边与提及

共同前提：Query 在笔记中基本出现、主题占比 < 10%，且 Query 的需求可以明确。Query_N 指 Query 在笔记中出现的内容。

表 57 沾边与提及的判断标准

类型	判断标准	分值
提及	Query_N 与笔记主题一致，确实在讲 Query 本身；去掉它会影响表达含义。	1
沾边	Query_N 与主题不匹配，或出现时重点不在 Query；去掉它不影响主要表达。	0

表 58 沾边与提及的判分案例

Query	笔记情形	分值
上海交通大学	主题一致，其中一所学校是上海交大。	1
小鱼	笔记原料是小鱼，主题一致。	1
赫尔格达	只是车辆接送范围描述，重点为车辆接送。	0
迈巴赫	提到词但并非真正在讲迈巴赫。	0
北京动物园冬天	并不是讲该地点冬季主题。	0
喝酒吃肉	Query_N 与笔记主题不匹配。	0
杉杉来了	主要介绍赵丽颖相关内容，重点不在作品。	0
护手霜	讨论护手霜沾到手镯如何清理，与护手霜有关。	1
大东沟	主题为靖远舰，大东沟是其历史的一部分。	1
领地	主题是儿童读物，删去“领地”不影响读物介绍。	0
申根签	主要求推荐旅游路线，签证词缺失不影响表达。	0

整体需求与局部供给

Query 需求为有具体数量的整体，笔记只给部分，原规则为 1 分。例：The Album 专辑 → 其中一首歌；BLACKPINK 团体 → 其中一个成员。

地点 + 搭子

表 59 地点与搭子需求的判定

情形	评分	案例
符合地点，且在本地范围内	3	德昌搭子 → 本地约拍 / 创业搭子。
跨地区	2	德昌搭子 → 德昌到西昌的顺风车搭子。

时间 / 地点无法明确

表 60 时间或地点不明确的处理

Query	结果	原图处理
莎莎新加坡大满贯今日	给到 2 月 9 日赛事，但无法确认搜索当天	抛弃并备注无法确认搜索时间。
莎莎新加坡大满贯 2 月 8 日	给到 2 月 9 日赛事	时间不满足，0 分。
刨冰附近	给珠海刨冰，但无法确认用户地点	抛弃并备注无法确认搜索地点。

品牌 / 店铺推荐

需求明确要某商品品牌 / 店铺推荐，只给商品而无品牌或店铺信息，按 0 分。原例：高质量袜子店铺、大同烧烤店推荐、小众短袖 T 恤店铺、EVA 拖鞋店铺。

资料 44 / 53 · 矢量重排 · [原图来源](#)

十一、Prompt 设计与模型输出约束

项目中 Prompt 设计遵循两个原则：

- 1. 把业务规则明确写入模型指令中，减少模型自由发挥。
- 2. 要求模型稳定输出 JSON，便于下游节点解析和打分。

典型 Prompt 包括：

- query 是否可理解判断。
- query 主需、次需识别。
- query 限制性条件提取。
- query 实体提取。
- 文本意图匹配。
- 图片意图匹配。
- 视频意图匹配。
- 文本/图片/视频实体匹配。
- 文本/图片/视频限制条件匹配。

为解决大模型输出不稳定问题，项目采用了 JSON 样例嵌入、结果校验、自动重试、AI 验证节点等方案。过程文档中将“大模型指定格式稳定输出”列为已完成事项，处理方式是增加重试节点和 AI 验证节点，对 JSON 格式异常数据进行修复。

十二、模型与系统工程实现

项目涉及多个模型和工具：

- 文本模型：qwen-flash、qwen3-coder-plus 等。
- 多模态模型：qwen-vl-max、GLM-4.1V-9B。
- 本地化模型：qwen7b、qwen32b、智谱 GLM 9B、TARS。
- 自动化能力：TARS、UI-TARS、Browser Use、Midscene、Magnetic-UI。
- 工作流：N8N。
- 数据库：MySQL。
- 容器部署：Docker Compose。

技术实现的重点不只是模型调用，而是把多个模型节点的输出组织成一个可计算的数据结构。项目通过 N8N 将多个模型调用拆成子流程，再由主流程合并结果。

系统实现中遇到的技术卡点包括：

1. 多个模型请求合并后，如何不降低返回精度。
2. 模型结果不一致、不稳定。
3. 非常规限制词识别不足。
4. 同义词、近义词、相似实体容易混淆。
5. 图片和视频信息识别耗时长。
6. 模型输出 JSON 格式异常。
7. 业务数据抓取不完整或重复。

对应优化方向包括：

- 输入数据预处理。
- Prompt 细化限制。
- 建立限制词词典。
- 建立业务类目库。
- 引入重试和格式校验节点。
- 对异常数据进行人工二次复核。

十三、人机协作与自动化分析

《智能化-人机协作&自动化分析》文档对人机协作平台进行了调研。该部分为作品提供了后续 Agent 化演进方向。

当前主流人机协作平台基于 LLM，将用户自然语言指令解析为结构化操作序列，再通过浏览器控制、DOM 结构分析、截图理解、视觉定位等方式执行任务。

文档中拆解了多智能体协作角色：

- orchestrator：任务协调器，负责整体流程控制。
- web_surfer：网页内容获取模块，支持网页抓取。
- coder：代码生成与执行模块。
- file_surfer：本地文件检索模块。
- action_guard：动作防护模块，监控并过滤风险操作。



图 17 人机协作平台流程架构

原图右下署名为“掘金技术社区 @ 顺凡”。此处依据原文所载图重新绘制结构并保留该署名信息。

资料 45 / 53 · 矢量重排 · [原图来源](#)

项目文档提出的路线是“本地 agent 与云端专家 agent 协同交互作业”。本地 agent 负责浏览器、文件系统、任务解析和基础执行，云端专家 agent 负责复杂策略判断、模型分析和知识增强。

三层协同架构



图 18 本地 Agent 与云端专家的三层协同架构

表 61 本地 Agent 与云端专家的交互方向

方向	原图标注
云端 → 本地	作业指引
本地 → 云端	数据采集
本地 → 终端	执行调度
终端 → 本地	交互采集

资料 46 / 53 · 矢量重排 · [原图来源](#)

原图左侧为 N8N “Query 大模型分析” 画布，右侧为 Vende AIFlowOps 的业务专家选择界面。原文图注写 “Midscene 运行界面截图” ；这里保留图注并按实际可见内容重排。

画面中的两个区域

表 62 N8N 与业务专家选择界面的可确认内容

区域	可确认内容
N8N 工作流区	可见多分支模型分析流程及其他子工作流；节点名称过小，无法逐一可靠识别。
业务专家选择区	小红书审核打标专家、华为代码标注专家、电商数据分析专家、金融风控专家。

正文所述工作链路



图 19 N8N 与业务专家的工作链路

此流程依据本文章节文字重绘，用于解释截图含义，并非不可辨读节点的逐一复原。

资料 47 / 53 · 矢量重排 · [原图来源](#)



图 20 VendeAI 平台登录页

界面文字：文德 AI 作业平台（Vende AIFlowOps）；用户名、密码两个输入框；“登录系统”按钮。

资料 48 / 53 · 原始图像 · [原图来源](#)



图 21 VendeAI 作业列表页

表 63 VendeAI 业务专家与界面说明

业务专家	界面说明
小红书审核打标专家	专业内容审核与标签标注，支持多维度内容分析。
华为代码标注专家	代码质量分析与智能标注，提供代码优化建议。
电商数据分析专家	商品数据挖掘与趋势分析，支持市场预测。
金融风控专家	风险评估与智能决策支持，提供投资建议。

以上是原界面功能文案，不表示这些能力已在本次整理中测试验证。

资料 49 / 53 · 原始图像 · [原图来源](#)



图 22 VendeAI 任务执行页



图 23 VendeAI 任务执行流程

资料 50 / 53 · 原始图像 · [原图来源](#)

自动化操作步骤重排



图 24 自动化操作步骤

原图 URL 与部分元素 ID 显示不完整，重排保留步骤含义，不补写地址、密码或缺失 ID。

资料 51 / 53 · 矢量重排 · [原图来源](#)

自动化平台调研结论

文档中对多个平台进行了比较：

- 1. Browser Use：适合浏览器自动化，结合 LLM 和 Playwright，可以通过自然语言解析网页任务。
- 2. UI-TARS-desktop：适合界面视觉定位和桌面级操作，但调试复杂，依赖模型输出。
- 3. Magnetic-UI：基于多智能体协同，适合复杂网页任务规划和执行。
- 4. 实在智能：面向 RPA 和数字员工场景，适合封闭系统操作。
- 5. UFO² AgentOS：偏系统级深度集成，适用于 Windows 生态。

Magnetic-UI 运行界面：信息重排

原截图界面名为 Magnetic-UI，原文图注使用 “Magnetic-UI”。左侧为会话列表，中部为计划与执行状态，右侧为 Live View 浏览器预览。

表 64 Magnetic-UI 界面信息

界面区域	原图可见内容
左侧导航	Current Session; Saved Plans; New Session; 历史会话。
任务计划	5 步：打开目标网址 → 输入用户名 → 输入密码 → 处理验证码 → 点击登录。
状态	Plan Accepted; Step 1 处于 Processing。
浏览器预览	Screenshots / Live View 标签；目标系统的登录页。



图 25 Magnetic-UI 计划与执行过程

本图按可见信息重新排版，省略截图中的具体登录凭据与服务器地址。截图只显示处理中，不能据此认定任务已登录成功。

Browser Use：任务与结果

原任务：“进入百度，搜索今天深圳天气信息，返回龙岗区天气。” 这是原截图中的历史任务，不代表当前天气。

表 65 Browser Use 截图中的任务指标

截图指标	显示值
模型	GPT-4.1
耗时	1 分 55 秒
费用	\$ 0.31
步骤	10 completed steps

截图可见的执行过程



图 26 Browser Use 截图中的执行过程

截图只展开了其中部分步骤；下方内容未显示，重排未补写其余步骤或最终天气结论。

资料 53 / 53 · 矢量重排 · [原图来源](#)

该调研说明，本项目不仅停留在 Copilot 辅助分析阶段，还考虑了后续向 Agent 自动化作业演进的可能性。

十四、项目成果

本项目形成了以下成果：

1. 完成小红书社区搜索相关性审核打标业务流程梳理。
2. 建立 query 结构化拆解框架。
3. 建立需求、实体、限制条件、有效内容等级等评分维度。
4. 搭建基于 N8N 的审核打标主流程和子流程。
5. 实现文本、图片、视频的多模态匹配。
6. 实现模型输出 JSON 化和自动化解析。
7. 实现匹配结果合并和分值映射。
8. 完成规则优化文档整理。
9. 完成 Copilot 到 Agent 演进路径分析。
10. 沉淀可复用的 AI 审核打标 SOP。

结合前序 baseline 测算材料，该项目通过规则细化、Prompt 优化、流程重构和人工复核，将早期模型评分一致性从较低水平提升到可验证的 MVP 阶段，为后续业务试点和模型专项调优提供了基础。

十五、项目难点与解决方案

难点 1：业务规则复杂，模型容易直接语义判断

搜索相关性不是简单的语义相似度任务。很多 case 中，query 和正文可能出现相同关键词，但不代表满足主需求。例如问答类 query 需要真正回答问题，对比类 query 需要覆盖对比对象和对比维度。

解决方式：将规则拆解为 query 类型、主需、次需、实体、限制条件、有效内容等级，并在 Prompt 中强制模型按步骤输出。

难点 2：多模态信息之间存在冲突

有些笔记正文满足 query，但图片无关；有些图片满足 query，但正文没有有效信息；有些 tag 与正文不一致。若简单合并，很容易高估或低估相关性。

解决方式：文本、图片、视频、tag 分开判断，再按规则合并。对于无实体 query，图片可不参与主题占比；对于 tag 与笔记内容完全不匹配的情况，忽略 tag。

难点 3：模型输出格式不稳定

大模型有时会返回自然语言，有时 JSON 格式错误，有时字段名不一致。

解决方式：Prompt 中嵌入 JSON 样例，增加结果校验和重试节点，并在 N8N 中使用代码节点对格式异常进行修复。

难点 4：外部知识缺失

例如地理位置、品牌别名、实体别称、药品概念等，如果模型没有外部知识，就可能误判。

解决方式：提出建立业务类目库、限制词词典、同义词/实体映射库，并在后续引入 RAG 或搜索增强能力。

难点 5：自动化作业不稳定

人机协作和浏览器自动化工具在复杂页面、跨域 iframe、验证码、动态 DOM、文件上传等场景下存在限制。

解决方式：对 Browser Use、UI-TARS、Midscene、Magnetic-UI 等方案进行调研，明确各自适用边界，为后续 Agent 级别自动化提供技术路线。

十六、个人能力体现

这个项目体现了以下能力：

1. 业务理解能力：能够把复杂审核规则拆成模型可执行的结构化步骤。
2. Prompt 工程能力：能够针对 query 分析、多模态匹配、实体识别、限制条件识别设计专门 Prompt。
3. workflow 搭建能力：能够使用 N8N 搭建主流程、子流程、循环节点、模型调用节点和数据库写入节点。
4. 数据处理能力：能够对原始业务数据进行清洗、格式化、图片视频识别和结构化存储。
5. 多模态应用能力：能够将文本、图片、视频分析结果合并为统一评分依据。
6. 工程落地能力：能够处理模型 JSON 输出不稳定、异常数据、重复图片、路径挂载等实际问题。
7. 评测复盘能力：能够通过 baseline 测算和 bad case 归因持续优化规则。
8. 自动化调研能力：能够分析 Browser Use、UI-TARS、Midscene、Magnetic-UI 等人机协作工具的适用场景。

十七、作品集展示摘要

本项目围绕小红书社区搜索相关性审核打标场景，构建了一套基于大模型、多模态理解和 N8N 工作流的智能审核打标 Copilot。系统将人工打标过程拆解为 query 结构化解析、正文/图片/视频匹配、实体和限制条件判断、有效内容等级计算、最终分值映射等环节，实现了从原始数据到 AI 评分建议的自动化闭环。

项目重点不在于简单调用大模型，而在于把复杂业务规则工程化、流程化、可解释化。通过规则文档整理、Prompt 设计、N8N 流程编排、MySQL 数据存储和 bad case 复盘，项目沉淀出一套可复用的 AI 审核打标方法论。同时，通过人机协作与自动化工具调研，为后续从 Copilot 辅助分析升级到 Agent 半自主执行提供了技术路线。

图表资料索引

以下按原材料的 53 组资料顺序索引，序号对应正文“资料”编号。图、表已分别编号；点击名称或页码可跳转至对应资料。

01. 审核打标工作模式流程框架图	4
02. Copilot 业务目标示意图	5
03. Agent 业务目标示意图	6
04. Copilot 流程价值示意图	7
05. Workflow 模型协同体系图	8
06. Model Hub 技术架构图	9
07. N8N 工作流示意图	10
08. Query 分解框架	11
09. 规则优化整理中的 Query 需求分类	12
10. Query 需求判断规则	13
11. 规则优化整理项目背景	15
12. 分值介绍	15
13. 笔记匹配规则	16
14. 标注流程	17
15. 整体分档及判定逻辑	17
16. 0 / -1 分判定流程	18
17. 评分标准样例，1 分规则展示	20
18. 布老虎案例素材 1（原图失效）	22
19. 布老虎案例素材 2	23
20. 布老虎案例素材 3	25
21. 潮汕站到揭阳机场案例图片	27
22. 0 / -1 分细则与案例	28
23. 含限定词 Query	29
24. 对比类与排行类 Query	30
25. 颜色需求与问答类 Query	31
26. 商品类 Query 与配件边界	32
27. 账号类 Query	34
28. 美食类 Query	35
29. 旅游类与文案类 Query	35

30. 明星、品牌与 IP 类 Query 36

31. 人物类 Query 37

32. 媒介需求与 POI Query 38

33. 歌曲类与壁纸类 Query 40

34. 纯数字与穿搭类 Query 40

35. 医疗类与翻创内容 41

36. 图文参评与主题占比 42

37. Tag 与求助类笔记 43

38. 出物 / 收物类规则 44

39. 有帮助类笔记 45

40. 类目判断细则更新 46

41. 类目补充与实体优先规则 47

42. 医疗类目与拆词判断 48

43. 沾边与提及的判断 49

44. 整体、搭子、时地与店铺案例 50

45. 人机协作平台流程架构图 53

46. 本地 Agent 与云端专家 Agent 协作图 54

47. N8N 与业务专家选择界面 54

48. VendeAI 平台登录页 56

49. VendeAI 作业列表页 57

50. VendeAI 任务执行页 58

51. 自动化操作步骤截图 59

52. Magentic-UI 计划与执行界面 60

53. Browser Use 运行界面截图 61

来源与重排说明

原文：【作品集】小红书社区搜索相关性智能审核打标。作者 city 不 city (m0_74750121)。

许可：CC BY-SA 4.0。重排范围包括表格、流程图、界面信息、标题、图注与版式；原文章节文字保留，原图链接保留供核对。少量案例文字做等义压缩，未新增实测成绩、产品能力或原材料缺失的判分。

原材料图源：52 处可用，1 处失效。资料 45（原文图 46）另署名“掘金技术社区 @ 顺凡”，已在对应图下保留。截图中的个人或系统登录凭据未重新展示。