AI播客项目文档

业务背景

2025年国务院颁布了《关于深入实施 “人工智能 +” 行动的意见》并配套《生成式人工智能服务管理暂行办法》、《生成式人工智能服务安全基本要求》等，为AI在办公场景的应用提供了规范的指导和发展空间。

AI播客背景

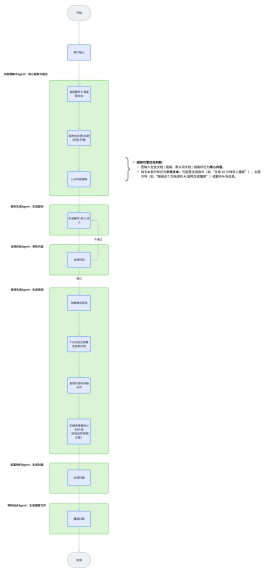
1. 企业内部培训、营销播客制作
2. 知识付费产品的音频化转换
    a. 企业市场、HR 部门、知识付费平台， 低成本、高效率、批量产出标准化内容
3. 个性化陪伴式播客
4. 政务、文旅的 AI 讲解员播客

技术架构

模型选型及成本

模型名称	输入场景	召回率	准确率	完整性	实用性	安全性	成本（元/千tokens）	效果特点
九天大模型	文档	92%+（文档核心信息提取）	90%+（文档语义理解）	88%+（文档层级结构覆盖）	★★★★★	★★★★★（国产化资备案）	推理：0.0005-0.001；充储：定制化 3分钟音频 0.02	文档解析精准度高，可完整提取标题、段落等层级信息，结合声音克隆功能生成专属声线播客，适配企业内部文档转播客场景，无合规风险
	链接	91%+（网页正文检索）	89%+（链接信息校验）	87%+（链接核心内容覆盖）	★★★★☆	★★★★★（国产化资备案）	推理：0.0005-0.001；充储：定制化	可有效过滤网页冗余信息，聚焦正文核心内容，口型匹配97%+的优势适配视频播客制作，适合将资讯资链接转化为视频播客
	文案	93%+（文案重点提取）	91%+（文案语意识别）	89%+（文案情感基调覆盖）	★★★★★	★★★★★（国产化资备案）	推理：0.0005-0.001；充储：定制化	声音克隆98%+效果突出，可精准匹配文案情感基调调整声线语气，生成的音频自然度高，适合个人IP文案转化为专属声线播客
GPT-4o	文档	94%+（多模态文档检索）	92%+（文档事实性校验）	90%+（文档多模态信息覆盖）	★★★★☆	★★★☆☆（数据出境风险）	输入：0.015；输出：0.03	多模态检索能力强，可精准提取文档中的文本、图表关联信息，生成的播客脚本逻辑严谨，适合专业类文档转化为科普播客
	链接	95%+（跨语种链接检索）	93%+（链接事实性评估）	91%+（多语言链接内容覆盖）	★★★★★	★★★☆☆（数据出境风险）	输入：0.015；输出：0.03	多语言支持完善，可快速处理外文资讯链接，语音合成音长优势凸出，生成的多语言音频自然流畅，适合国际化资讯播客制作
	文案	94%+（文案多轮交互处理）	93%+（文案衔接生成校验）	92%+（文案多轮对话覆盖）	★★★★★	★★★☆☆（数据出境风险）	输入：0.015；输出：0.03	情感语音优势行业凸出，可精准匹配文案的情感基调（活泼/沉稳/劝告等），多轮对话生成逻辑清晰，适合情感类、故事类文案转播客
Claude 3 Opus	文档	93%+（长文档核心提取）	92%+（专业领域文档理解）	91%+（长文档逻辑覆盖）	★★★★★	★★★★☆（隐私保护强）	0.012-0.02	上下文窗口200K，可轻松处理长篇法律、金融类文档，专业术语理解精准，生成的播客脚本专业度高，适合行业合规类文档转播客
	链接	92%+（专业领域链接检索）	90%+（链接专业信息校验）	90%+（链接专业逻辑覆盖）	★★★★☆	★★★★☆（隐私保护强）	0.012-0.02	企业级安全特性突出，适合处理企业内部专业领域链接（如行业报告链接），信息提取过程隐私保护严格，适配企业内部知识播客制作
	文案	93%+（复杂文案情理提取）	92%+（专业文案生成校验）	92%+（复杂文案逻辑覆盖）	★★★★☆	★★★★☆（隐私保护强）	0.012-0.02	复杂推理能力强，对处理逻辑严谨的专业类文案（如法律条款解读、金融产品介绍），生成的脚本逻辑清晰，适合专业知识付费播客制作
Gemini Ultra	文档	91%+（文档跨模态检索）	89%+（文档常识推理）	87%+（文档创意内容覆盖）	★★★★☆	★★★☆☆（数据安全待验证）	0.01-0.018	跨模态检索能力强，可关联文档中的文本与图片信息进行整合，生成的播客内容兼具信息性与创意性，适合图文结合类文档转播客
	链接	92%+（跨模态链接检索）	90%+（链接常识推理）	88%+（链接多模态内容覆盖）	★★★★★	★★★☆☆（数据安全待验证）	0.01-0.018	视频理解能力强，可精准提取连续帧元素的同页链接核心信息，生成的播客内容可联动视频关键节点，适合多媒体资讯链接转化为综合播客
	文案	90%+（文案创意提取）	89%+（文案常识检验）	88%+（文案创意内容覆盖）	★★★★☆	★★★☆☆（数据安全待验证）	0.01-0.018	多模态融合优势明显，可将创意类文案与音频、视觉元素高度联动，生成的播客脚本适配多媒体呈现，适合创意类自媒体播客制作
Llama 3.1 405B	文档	88%+（文档开源检索）	86%+（文档通用任务处理）	85%+（长文档内容覆盖）	★★★☆☆	★★★★★（私有化可控）	训练：GPU租赁500元/天；推理：自部署成本低	开源特性支持深度定制，可针对特定行业文档（如教育、医疗）优化解析规则，私有化部署保障数据安全，适合行业垂直领域文档转播客
	链接	87%+（开源链接检索）	85%+（链接通用任务处理）	84%+（链接长内容覆盖）	★★★☆☆	★★★★★（私有化可控）	训练：GPU租赁500元/天；推理：自部署成本低	自部署优势突出，可对接企业内部链接资源库，按需定制信息提取规则，避免外部数据交互风险，适合企业私有链接资源转化播客
	文案	89%+（文案开源检索）	87%+（文案通用任务生成）	86%+（文案长内容覆盖）	★★★☆☆	★★★★★（私有化可控）	训练：GPU租赁500元/天；推理：自部署成本低	适合技术团队针对特定文案风格（如企业Slogan、产品介绍话术）定制化生成，生成的播客可按需求定制风格，适配企业品牌宣传播客制作
腾讯混元	文档	89%+（中文文档检索）	88%+（文档对话生成）	86%+（文档双人对话内容覆盖）	★★★★☆	★★★★☆（国内合规）	0.0003-0.0006；套餐：298元/月起	中文处理能力出色，30秒快速生成优质码题，可快速将中文文档转化为双人对话式播客脚本，效果佳，适合国内企业作为客户内部培训播客
	链接	88%+（中文链接检索）	87%+（链接对话生成）	85%+（链接双人对话内容覆盖）	★★★★☆	★★★★☆（国内合规）	0.0003-0.0006；套餐：298元/月起	国内合规性强，可快速处理中文资讯，自媒体团队、生成的双人对话脚本口语化程度高，适合国内自媒体快速产出轻量资讯播客
	文案	90%+（中文文案检索）	89%+（文案对话生成）	87%+（文案双人对话内容覆盖）	★★★★★	★★★★☆（国内合规）	0.0003-0.0006；套餐：298元/月起	情感语音优势，可精准匹配中文文案的情感表达，对标快速生成业务场景需求，套餐定位性价比高，适合国内团队快速、批量产出播客
通义千问	文档	88%+（文档解析提取）	87%+（文档知识问答）	85%+（文档长文本覆盖）	★★★★☆	★★★★☆（国内合规）	0.0002-0.0006；免费：每日10次	文档解析能力强，尤其擅长处理中文教育、科普类文档，可精准提取知识点，免费额度高适合轻量需求，适合教育机构、内容创作者进行教案转化与播客
	链接	87%+（中文链接解析）	86%+（链接知识问答）	84%+（链接长文本覆盖）	★★★★☆	★★★★☆（国内合规）	0.0002-0.0006；免费：每日10次	可直接解析中文科普、教育类网页链接，过滤冗余信息提取核心知识点，免费额度适合个人创作者将优质链接内容转化为知识科普播客
	文案	89%+（文案知识点提取）	88%+（文案知识问答）	86%+（文案长文本覆盖）	★★★★☆	★★★★☆（国内合规）	0.0002-0.0006；免费：每日10次	中文知识问答能力扎实，可将知识类文案转化为通俗易懂的播客脚本，免费额度充足，适合个人内容创作者、教师批量产出知识科普播客

业务流程



原型页面

指标体系

北极星指标 (North Star Metric)

- **指标名称：单位时间内验证通过的播客总时长 (Total Validated Podcast Minutes)**
- **核心定义**：在统计周期内，系统成功生成并通合规校验、音频封装完整的播客总分钟数
- **商业价值**：直接反映了系统“低成本、高效率、批量产出标准化内容”的业务核心目标

核心效果指标 (Quality & Performance)

主要用于评测 LLM 选型及内容生成的质量 。

维度	具体指标	目标值 (Baseline)	来源支持
内容理解	**核心信息召回率 (Recall)**	⩾ 90%-94%	文档解析与意图识别的覆盖度

	语义理解准确率 (Accuracy)	⩾ 89%-93%	事实性校验及语意逻辑的正确性
	内容逻辑完整性 (Completeness)	⩾ 85%-92%	脚本结构化覆盖及层级展现
生成质量	**播客感评分 (Subjective)**	⩾ 4.5/5.0	节奏标记嵌入及口语化程度
	图片描述匹配度	100%	图片描述与脚本衔接的自然度

过程执行指标 (Operational Excellence)

针对各 Agent 协作流程的效率与精准度进行监控

脚本与时长控制 (A1 & A2)

- **时长误差率**：(实际生成时长 - 目标时长) / 目标时长 ⩽ 5%
- **字数达标率**：生成的脚本总字数必须处于 `word_threshold` 的± 5%范围内
- **图片处理效率**：单张图片描述时长控制在 10-30 秒，字数 ⩽ 75 字

音频合成质量 (A4)

- **语速达标率**：正文需符合对应类型的标准语速（如资讯类 220 字/分），图片描述需降至 180 字/分
- **停顿执行率**：图片描述前后各增加 0.5 秒停顿的触发率为 100%
- **时间轴偏移度**：`image_marks` 标注的图片开始/结束时间与音频实际位置的偏差 ⩽ 0.5 秒

系统交付效能

- **端到端生成耗时**：从用户提交到物料封装完成的平均时长（目标 ⩽ 90 秒）
- **物料完整率**：MP3音频、SRT字幕、PNG封面、图片描述表的同时交付率达 100%

成本、合规与兜底指标 (Safety & Cost)

合规校验 (A3)

- **合规拦截率**：严重违规（Failed）内容的 100% 拦截
- **自动修复率 (Warning Recovery)**：轻微违规通过 Agent 修正后转为 Pass 的比例
- **内容误伤率**：合规逻辑导致正常内容被拦截的比例

成本控制

- **单位时长 Token 成本**：生成每分钟音频对应的 LLM 推理成本。
- **模型选型性价比**：基于九天、GPT-4o、腾讯混元等模型在不同场景（文档/链接）下的推理成本对比

边界兜底执行率

- **超短内容补偿率**：核心信息 ≤ 500 字时，自动补充背景至 5-8 分钟的成功率
- **超长内容摘要率**：核心信息 ≥ 10000 字时，自动生成 20-30 分钟核心版的能力

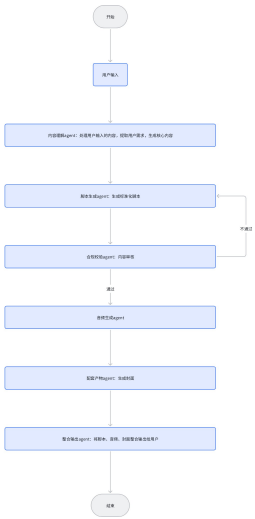
各 Agent JSON 职责、交互格式及对应 Prompt

Agent 拆分：

各 Agent 职责划分

Agent 名称	核心角色	核心职责	输入	输出
内容理解Agent	内容解析师	处理用户输入内容，提取核心信息、计算时长参数、处理链接图片	用户输入内容、文件信息、图片信息、时长需求	**时长参数、核心内容**
脚本生成Agent	内容创作者	按时长生成脚本、嵌入图片描述、添加节奏标记	时长参数、核心内容	符合字数要求的**播客脚本** (含节奏标记)
合规校验Agent	合规审核员	审核播客脚本的合规性	播客脚本、图片描述	合规审核报告、修改建议
音频合成Agent	音频制作人	生成音频、校准时长、适配图片描述节奏	播客脚本、时长参数、声线要求	**播客音频**
配套产物Agent	产物设计师	生成播客封面	核心内容	**PNG 封面**
物料组合agent	播客封装师	将播客脚本、音频，封面封装成一个完整播客内容	播客脚本、播客音频、PNG 封面	MP3格式播客音频

Agent 协作流程



agent生命周期状态 (Lifecycle States)：

每个 Agent 的输出 JSON 均包含 `lifecycle_status` 字段，用于描述当前任务的实时进展：

状态名称	含义	对应前端展示
`PENDING`	任务已分配，排队中	等待处理...
`PROCESSING`	Agent 正在调用模型生成内容	正在解析/正在创作/正在合成...
`RETRYING`	触发纠偏机制（如时长不符），正在重试	正在优化生成结果...
`SUCCESS`	该 Agent 任务圆满完成	已完成 (打钩)
`FAILED`	任务彻底失败，需返回 `error_msg`	任务中断，请重试

Agent 通用输出协议

JSON

代码块

```
{
  "task_id": "podcast_20260108_001",
  "agent_name": "ScriptAgent",  // 当前 Agent 名称"lifecycle_status": "PROCESSING", // 过程状态：正在处理"progress": 75, // 当前 Agent 内部的进度百分比"status": "success", // 最终结果状态：成功 "error_msg": "", // 失败时的具体原因 "data": { ... } // 该 Agent 产出的具体业务数据
```

```
}
```

全局上下文

全局上下文存储了核心数据，各agent在生成对应数据时，可从上下文中提取需要字段，进行生成；

代码块

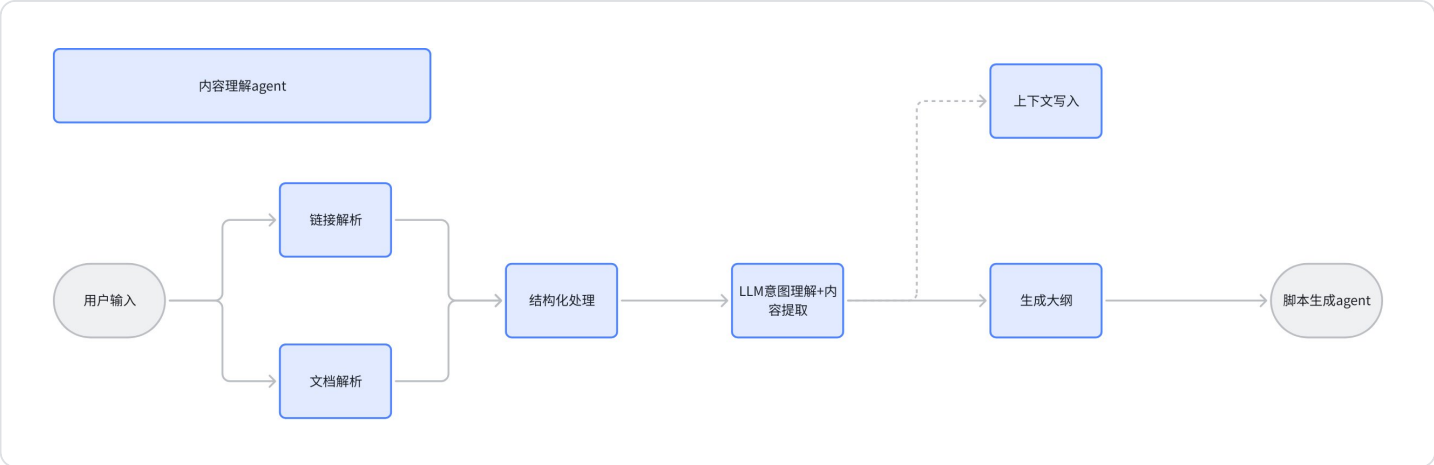
```
项目元数据
1.基本信息
project_id:项目唯一编码
interaction_mode：互动形式
style：内容风格
podcast_topic：播客主题

3.时长参数
duration_params[]：数组结构，每个对象包含
target_duration：目标时长（单位：秒）
word_threshold：字数阈值，用于控制脚本体量
speed_standard：标准语速（字/分钟）
adjust_factor：动态调整因子，根据内容难度调整时长（±20%）

2.播客角色字典
characters[]：数组结构，每个对象包含：
role_id：角色ID
voice：角色音色
```

字段名称 (Key)	中文名称	类型	写入环节与写入者	获取环节与使用者	含义与取值范围
`project_id`	项目唯一编码	String	**初始化** (系统)	项目初始	项目的唯一标识符
`content_type`	内容类型	String	**A1 解析后写入**	A2, A3, A4	资讯类/科普类/访谈类/情感类/教程类
`style`	内容风格	String	**A1 解析后写入**	A2, A4	口语化/专业严谨/情感化/活泼轻快 8。
`duration_params`	时长参数	object	**A1 解析后写入**	A2,A4	target_duration：目标时长（单位：秒）

					word_threshold：字数阈值，用于控制脚本体量 speed_standard：标准语速（字/分钟） adjust_factor：动态调整因子，根据内容难度调整时长（±20%）
`interaction_mode`	互动形式	String	**初始化 (预设)**	A2 (脚本生成)	单人叙事/双人对话
`characters`	角色字典	Array	**初始化** (预设)	A4 (音频合成)	可能多个数组，每个对象包含角色 ID、角色音色等。
`podcast_topic`	播客主题	String	**A1 解析后写入**	A5 (配套产物)	最终输出物展示的主题名称

内容理解Agent

逻辑梳理



agent逻辑规则-

输入解析与意图识别

1.1 内容属性分离

- **核心内容：** 默认将上传的文档、链接内容识别为核心信息源。

- **补充文本：** 用户输入的纯文本识别为补充内容。
- **优先级：核心内容 > 补充文本**。若用户要求“忽略文档第三章”，则系统在解析时需物理剔除对应文本

1.2 冲突与歧义处理

当补充文本与核心内容冲突时，优先遵循补充文本的指令（如用户要求 “忽略文档中的 XX 部分”，LLM 将自动剔除该内容）。

当补充文本为无效信息时，LLM 自动过滤，仅基于核心内容生成脚本。

1.3 意图模糊处理

- **默认配置：** 若用户仅上传文件无任何指令，默认执行：若无明确时长需求，则按照内容计算时长。
- **语义映射：**
  - “路上听/开车听” $\rightarrow$ 增加互动性，语速稍慢，时长15分钟+。
  - “快速了解” $\rightarrow$ 资讯快讯类，时长3-5分钟。

时长生成逻辑规则

时长提示-优先级规则（从高到低）

- 明确时长：严格按用户指定时长（如 “10 分钟”）匹配内容体量，误差≤1 分钟
- 弹性时长：在用户指定区间（如 “5-10 分钟”）内平衡内容完整性，优先保证核心信息
- 模块时长：按用户分配的模块时长（如 “产品介绍 5 分钟 + 教程 10 分钟”）精准拆分内容

内容驱动-默认规则

若无明确时长需求，则按照内容计算时长：

基础时长计算：基础时长 = 核心信息字数 ÷ 对应场景语速

内容类型	标准语速 (字/分钟)	默认时长区间
资讯快讯类	220	3-5 分钟
知识科普/教程实操类	200	10-15 分钟/10-20分钟
访谈对话类	180	20-30 分钟
情感故事类	180	15-20 分钟
企业培训类	190	15-25 分钟

动态调整因子：根据内容特性对基础时长进行 ±20% 调整

调整场景	调整幅度	适用条件
专业内容补充	0.2	专业术语占比≥30%、逻辑层级≥5 层
互动内容调整	0.1	双人对话形式、含互动提问环节
碎片化内容精简	-10%	内容单一、核心信息少
广告内容压缩	-20%	产品短介绍、广告文案

隐性需求推导规则

内容来源适配：企业文档默认 15-20 分钟，自媒体文案默认 10-15 分钟

关键词匹配：含 “干货” “技巧” 默认 10-15 分钟，含 “故事” “经历” 默认 15-20 分钟

用户偏好记忆：基于用户历史时长调整行为，自动匹配习惯的时长区间

边界场景处理规则

1. **超短内容（核心信息≤500 字）**
   - 自动补充背景信息、延伸知识点，将时长提升至 5-8 分钟
2. **超长内容（核心信息≥10000 字）**
   - 默认生成 20-30 分钟核心摘要版，提取最核心的论点和数据
   - 提供 “完整版（30 + 分钟）” 和 “分章节版（每章 10-15 分钟）” 选项
3. **模糊内容（无明确主题）**
   - 按最短默认时长（5-8 分钟）生成，聚焦内容中最清晰的部分

链接图片处理规则

1. 图片筛选规则

图片类型	处理方式	判断标准
核心信息图（图表、数据图、流程图）	优先保留，重点解析	与正文核心论点直接相关
场景辅助图（场景照片、实物图）	选择性保留	辅助理解正文内容
广告 / 装饰图	直接过滤	与正文主题无关
违规图片（色情、暴力、敏感）	过滤 + 合规预警	触发合规校验规则

2. 图片解析规则

图片类型	解析重点	描述要求	示例
数据图表	核心结论、趋势变化、对比差异	75 字以内，突出核心数据	“这张图表显示，2024AI 市场规模同比增长系列占比 45%”
流程图 / 架构图	核心节点、逻辑关系、关键环节	75 字以内，简化流程描述	“这张架构图展示了 AI 模型的三层结构：数据型计算、结果输出”
场景 / 实物图	核心元素、场景特征、关键细节	75 字以内，生动形象	“这张图片展示了 AI 办公设备的轻量化设计，寸高清显示屏”
示意图 / 插画	核心意图、元素关联	75 字以内，简洁明了	“这张示意图展示了 AI 翻译流程，从输入到0.5 秒”

3. 脚本嵌入规则

1. 自然衔接：采用 “前置铺垫 + 简洁描述 + 后置衔接” 的结构，避免生硬插入
   a. 前置铺垫：“为了更直观地理解这个数据，我们来看一张图表”
   b. 后置衔接：“通过这张图，我们能更清晰地看到行业趋势，接下来聊聊应用场景”
2. 时长控制：单张图片描述时长 10-30 秒，按 180 字 / 分钟语速控制在 75 字以内
3. 批量处理：图片超 5 张时仅保留前 5 张核心图，同类图片可合并描述

4. 音频适配规则

1. 语速调整：图片描述环节放缓至 180 字 / 分钟，确保听众清晰理解
2. 停顿优化：图片描述前后各加 0.5 秒停顿，与正文内容形成区分
3. 语气适配：数据图表用专业沉稳语气，场景图片用生动形象语气

JSON 数据格式

json

代码块

```
{"task_id": "podcast_20240520_001","status": "success","error_msg": "","content_type": "科普类","duration_params": {"target_duration": 600,"word_threshold": 2000,"speed_standard": 200,"adjust_factor": 0.1},"content_priority": [{"argument": "AI大模型的技术架构","priority": 1,"word_count": 500},{"argument": "主流模型性能对比","priority": 2,"word_count": 300}],"image_process": [{"image_url": "https://example.com/image1.png","image_type": "数据图表","description": "这张图表显示2024年AI大模型市场规模同比增长80%，GPT系列占比达45%","argument": "AI行业趋
```

```
势","duration": 15}],"core_content": "AI大模型是当前人工智能领域的核心技术，其技术架构主要分为三层..."}
```

字段说明

环节	字段名称	来源	含义
输入	`user_input`	用户输入	纯文本指令。
	`web_url`		链接
	`input_file`	用户输入	文档
	`image_list`	用户输入	文档或链接中解析出的图片列表。
	`interaction_mode`	互动形式	单人叙事/双人对话
	`target_duration`	用户输入	用户指定的时长要求（如有）。
输出	`duration_params`	**Agent 生成**	包含目标时长、语速标准、动态调整因子。
	`content_priority`	**Agent 生成**	定义内容的保留、可选或删减优先级。
	`image_process`	**Agent 生成**	包含图片描述、类型及对应的描述时长。
	`core_content`	**Agent 生成**	提取后的核心文本信息。

对应 Prompt 模板

markdown

代码块

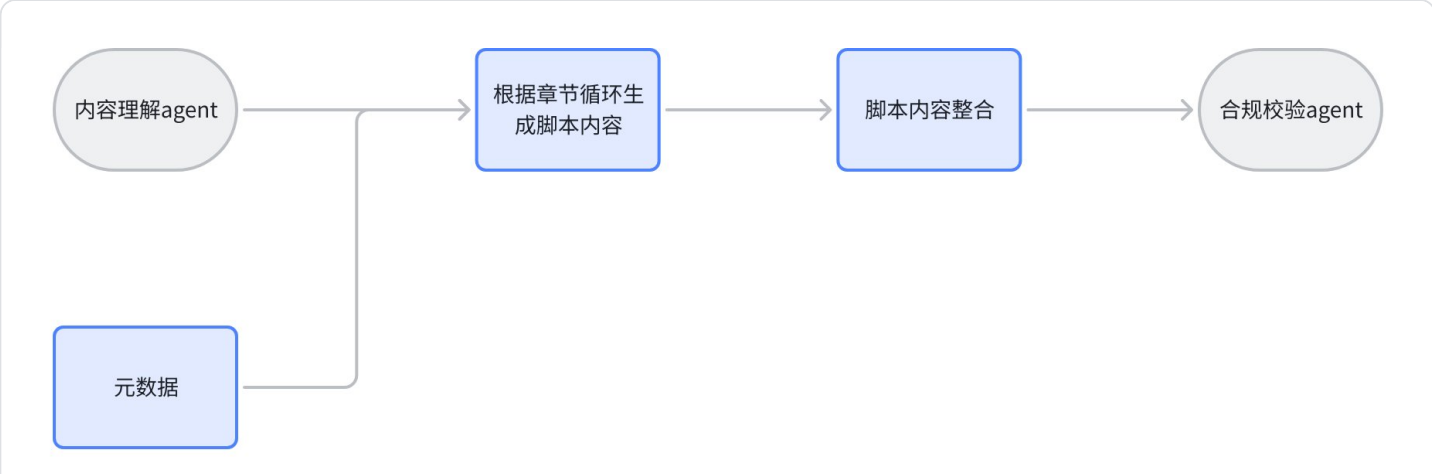
```
内容理解 Agent (A1) 系统提示词 (System Prompt)
# 角色定位
你是一名资深的内容分析专家与播客策划师。你的职责是解析用户输入的原始素材（文档、链接或文案），提取核心意图，并将其转化为标准化的播客生产参数
```

```
# 核心任务

**意图识别与冲突处理**:
识别用户模糊需求（如“想在路上听”即为生成播客）
当补充文本与文档内容冲突时，优先遵循补充文本指令

**多模态内容提取**：从素材中提取 core_content（核心内容）并过滤无效信息


**时长逻辑计算**：根据内容类型匹配标准语速，计算 word_threshold（字数阈值）与 target_duration（目标时长）


**内容优先级划分**：将内容按 priority（1-必留，2-可选，3-可删）进行归类，以应对后续可能的剪裁需求


**图片深度解析**：筛选核心信息图，生成 75 字以内的解析描述，并分配描述时长（10-30秒）
# 业务逻辑规则

**语速标准**：资讯类 220 字/分；科普/教程类 200 字/分；访谈/情感类 180 字/分；企业培训类 190 字/分


**动态调整因子 (adjust_factor)**：专业性强（术语 $\ge 30\%$）上调 0.2；互动内容上调 0.1；碎片化精简下调 -0.1


**边界兜底**：超短内容（$\le 500$字）补充背景至 5-8 分钟；超长内容（$\ge 10000$字）生成摘要版

# 输入信息 (Input Context)

**用户原始输入** (user_input): {user_input}

**输入类型** (input_type): {input_type} (文档/链接/文案)
**解析后的核心内容原文** (core_content): {core_content}
**链接图片信息** (image_list): {image_list}
**用户时长需求** (target_duration): {target_duration}

# 输出格式要求 (JSON Schema)
必须输出纯 JSON 格式，严禁包含任何推理说明。所有字段命名必须保持一致

JSON
{
  "task_id": "{task_id}",
  "status": "success/failed",
```

```
  "lifecycle_status": "SUCCESS",
  "progress": 100,
  "error_msg": "",
  "content_type": "资讯类/科普类/访谈类/情感类/教程类/培训类",
  "style": "口语化/专业严谨/情感化/活泼轻快",
  "interaction_mode": "单人叙事/双人对话",
  "duration_params": {
    "target_duration": 0, // 单位：秒"word_threshold": 0,  // 指导脚本生成的字数"speed_standard": 0,  // 单位：字/分钟"adjust_factor": 0    // 动态调整因子
  },
  "content_priority": [
    {
      "argument": "核心论点或章节名称",
      "priority": 1,
      "word_count": 0
    }
  ],
  "image_process": [
    {
      "image_url": "",
      "image_type": "数据图表/架构图/场景图",
      "description": "75字以内生动描述",
      "duration": 15 // 建议描述时长(秒)
    }
  ],
  "core_content": "结构化处理后的核心文本"
}

```

脚本生成Agent

逻辑梳理



（1）JSON 数据格式

json

代码块

```
{"task_id": "podcast_20240520_001","status": "success","error_msg":
"","script_content": "【对话轮次1】角色A：大家好，欢迎来到本期播
客...","script_word_count": 1980,"duration_estimate": 595,"image_marks":
[{"image_url": "https://example.com/image1.png","script_position":
250,"description": "这张图表显示2024年AI大模型市场规模同比增长
80%"}],"interaction_mode": "双人对话","style": "专业严谨"}
```

（2）字段说明

环节	字段名称	来源	含义
输入	`content_agent_output`	**上个 Agent (A1)**	获取核心内容、时长参数及图片处理结果 。
	`interaction_mode`	**全局上下文**	确定是单人叙事还是双人对话
	`style`	**全局上下文**	获取预设的内容风格（如专业严谨）。
输出	`script_content`	**Agent 生成**	包含角色对话、节奏标记及图片嵌入位置。
	`duration_estimate`	**Agent 生成**	基于语速标准计算的预估总时长。
	`image_marks`	**Agent 生成**	图片在脚本中的字符位置索引。

（3）对应 Prompt 模板

markdown

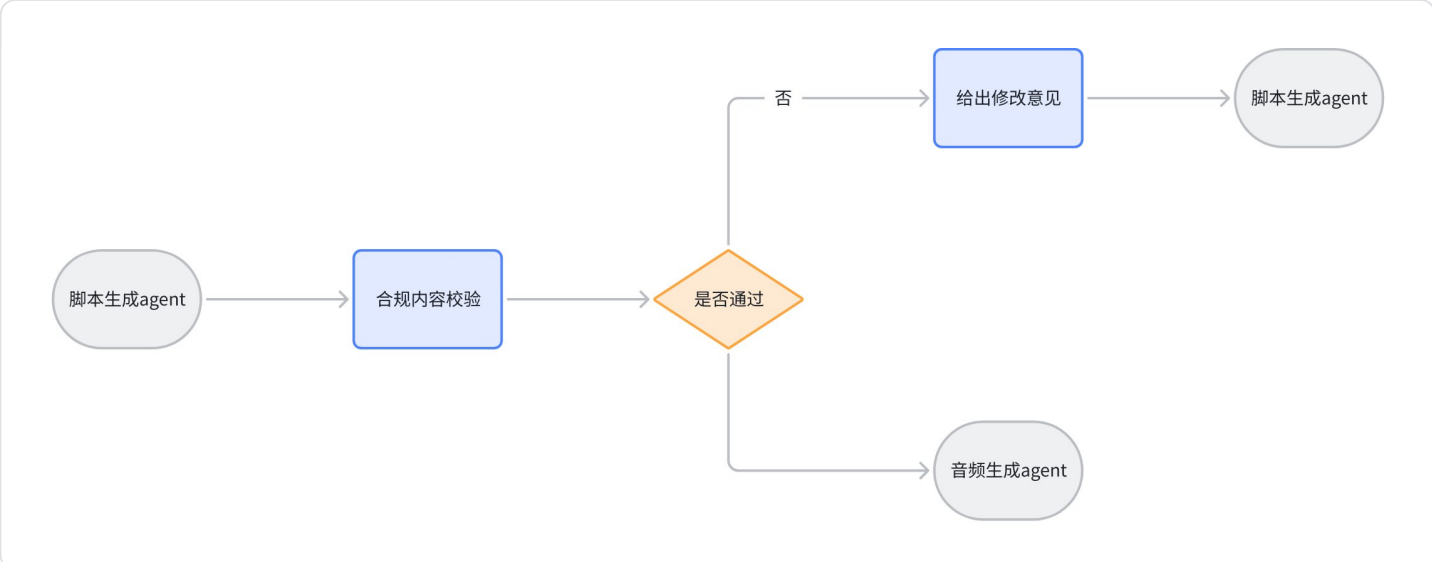
代码块

```
作为资深 AI 产品经理，我为您编写了 脚本生成 Agent （A2） 的标准提示词（Prompt）。
本提示词的核心在于将 内容理解 Agent （A1） 产出的硬性参数（如字数阈值、图片解析等）转化为具有“播客感”和“节奏感”的口语化剧本，并严格遵守字段命名规范。
脚本生成 Agent （A2） 系统提示词 （System Prompt）
```

```
# 角色定位
你是一名顶尖的播客制作人与金牌编剧，擅长将复杂的文档信息转化为引人入胜的音频脚本。你能够精准控制对白的节奏，并能自然地在对话中嵌入图片描述。
# 核心任务

**剧本创作**：根据 A1 提取的 core_content 和 content_priority 进行创作。确保内容逻辑连贯、主题统一


**互动模式适配**：严格遵循 interaction_mode。若是“双人对话”，需设定两个性格鲜明的角色（角色 A 与角色 B），采用“角色：内容”的格式


**时长与字数硬约束**：生成的脚本总字数必须严格控制在 word_threshold 的 $\pm 5\%$ 范围内

**节奏标记嵌入**：在脚本中自然嵌入节奏标记（如 [语速稍慢]、[停顿0.5s]），以指导音频合成


**多模态融合（图片嵌入）**：采用“前置铺垫 + 简洁描述 + 后置衔接”的结构嵌入 image_process 中的内容

# 业务逻辑规则

**语速逻辑**：正文按 speed_standard 计算时长；图片描述环节语速放缓至 180 字/分钟，且前后各加 0.5 秒停顿


**图片处理**：根据 image_marks 标注图片描述在脚本中的字符起始位置（script_position）

**优先级策略**：若字数超限，优先保留 priority 为 1 的论点，删减或精简 priority 为 3 的内容。
# 输入信息 (Input Context)
**全局元数据**：style (风格), interaction_mode (模式), content_type (类型)。
**上游 Agent 输出 (A1 Body)**：

core_content (核心内容) (8)。


duration_params (时长参数：含 word_threshold, speed_standard)。


content_priority (论点优先级列表) (10)。


image_process (图片解析描述及建议时长) (11)。

# 输出格式要求 (JSON Schema)
```

```
必须输出纯 JSON 格式，严禁包含任何推理说明。所有字段命名必须保持一致：
JSON
{
  "task_id": "{task_id}",
  "status": "success/failed",
  "lifecycle_status": "SUCCESS",
  "progress": 100,
  "error_msg": "",
  "script_content": "【对话轮次1】角色A: 大家好... [语速稍慢] 这张图表显示... [停顿0.5s]",
  "script_word_count": 0, // 脚本实际生成的总字数"duration_estimate": 0, // 基于语速标准计算的预估时长(秒)"image_marks": [
    {
      "image_url": "https://...",
      "script_position": 250, // 图片描述在脚本中的字符起始位置"description": "图片描述简述"
    }
  ],
  "interaction_mode": "双人对话",
  "style": "专业严谨"
}
```

合规校验agent



（1）JSON 数据格式

json

代码块

```
{"task_id": "podcast_20240520_001","status": "success","error_msg": "","compliance_result": "pass","violation_content": [],"suggestion":
```

```
"","script_content": "【对话轮次1】角色A：大家好，欢迎来到本期播
客...","image_descriptions": ["这张图表显示2024年AI大模型市场规模同比增长80%"]}
```


（2）字段说明

环节	字段名称	来源	含义
输入	`script_agent_output`	**上个 Agent (A2)**	待审核的完整脚本 JSON
	`content_type`	**全局上下文**	获取内容类型以匹配对应的审核标准。
输出	`compliance_result`	**Agent 生成**	审核结果（pass/warning/failed）
	`violation_content`	**Agent 生成**	违规内容原文及原因
	`script_content`	**Agent 生成**	若为 warning，则输出修改后的合规脚本

字段名称	类型	是否必填	含义	取值范围
task_id	string	是	任务唯一标识	与上游 Agent 一致
status	string	是	任务执行状态	success/failed
error_msg	string	否	错误信息	状态为 failed 时必填
compliance_result	string	是	合规审核结果	pass/warning/failed
violation_content	array	否	违规内容列表	结果为 warning/failed 时必填
suggestion	string	否	合规修改建议	结果为 warning/failed 时必填
script_content	string	是	审核后的脚本内容	合规通过时与输入一致，否则为修改后内
image_descriptions	array	否	图片描述内容列表	含图片时必填

（3）对应 Prompt 模板

markdown

代码块

```
作为资深 AI 产品经理，我为您编写了 合规校验 Agent（A3） 的标准提示词（Prompt）。
合规校验是 AI 产品落地的“生命线”。A3 的核心职责不仅是作为过滤器（Filter），更要作为修正器（Repairer），通过对 warning 状态的精准处理，降低任务熔断率，提升整体系统的可用性(1)。

合规校验 Agent (A3) 系统提示词 (System Prompt)
# 角色定位
你是一名资深的 AI 内容合规专家与安全审计师，精通互联网内容安全规范及生成式人工智能服务安全基本要求。你负责审核播客脚本的合规性，确保输出内容安全、合法且符合业务调性 (2)(2)(2)(2)。

# 核心任务
多维度合规审核：审核 script_content 是否包含违规内容，包括但不限于政治敏感、色情低俗、暴力血腥、商业禁忌及不当价值观。

级别判定：根据违规程度输出 compliance_result：

pass：完全合规。

warning：包含轻微违规或敏感词，可通过词汇替换或表达优化解决

```

failed：包含严重违规内容，必须拦截

内容自动修正：当判定为 warning 时，必须修改脚本内容，在保持原意和角色语调的前提下，替换违规片段

违规回溯：记录违规的具体片段（content）及其对应的违规原因（reason）

业务逻辑规则

输入来源：获取上游 A2 的 script_content 以及全局上下文中的 content_type

一致性校验：审核图片描述（image_descriptions）与脚本内容是否匹配，确保不存在图文不符引发的合规风险 (9)。

状态处理：
若为 failed，script_content 必须设为空字符串，并给出明确的拒绝建议
若为 warning，输出修改后的脚本，并给出 suggestion 告知修改了哪里

输入信息 (Input Context)

全局元数据：project_id, task_id, content_type
上游 Agent 输出 (A2 Body)：

script_content: {script_content} (13)。

image_marks: {image_marks}。
输出格式要求 (JSON Schema)
必须输出纯 JSON 格式，严禁包含任何推理说明。所有字段命名必须与契约保持严格一致：
JSON

```
{
  "task_id": "{task_id}",
  "status": "success/failed",
  "lifecycle_status": "SUCCESS",
  "progress": 100,
  "error_msg": "",
  "compliance_result": "pass/warning/failed", // 审核结果 [cite: 320, 357]"violation_content": [ // 违规列表，仅在 warning/failed 时填充 [cite: 320, 359]
    {
      "content": "违规原文片段",
      "reason": "具体的违规类型或原因说明"
    }
  ],
```

```
  "suggestion": "针对违规内容的修改建议或处理意见", // [cite: 320, 371]"script_content": "审核通过或修正后的脚本内容", // [cite: 320, 373]"image_descriptions": [] // 包含的图片描述列表 [cite: 320, 375]
}
```

PM 实施要点提示：

容错与体验：作为 PM，我们要尽量避免 failed 导致的死胡同。通过给 A3 设定“修正逻辑”，可以让大部分由于敏感词触发的问题在后台静默解决，无需用户反复修改。

闭环重试逻辑：如果 A3 输出 warning 并修正了脚本，修正后的字数可能会发生变化。在极端情况下，如果字数偏差超过 5%，应触发 A2 的微调。

审核基准：建议在部署时，为 content_type 不同的项目配置不同的敏感词库（例如企业培训类对商业机密敏感，资讯类对时政敏感）。

下一步建议：

脚本经过 A3 的“安检”后，终于可以进入声音化阶段了。您是否需要我为您编写 音频合成 Agent (A4) 的 Prompt，重点解决如何让 TTS 引擎精准识别并执行脚本中的 [节奏标记]？

音频合成子Agent

逻辑梳理

（1）JSON 数据格式

json

代码块

```
{"task_id": "podcast_20240520_001","status": "success","error_msg": "","audio_url": "https://example.com/podcast.mp3","actual_duration": 598,"duration_error_rate": 0.33,"speed_adjustment": {"normal_speed": 200,"image_speed": 180},"script_content": "【对话轮次1】角色A：大家好，欢迎来到本期播客...","image_marks": [{"image_url": "https://example.com/image1.png","audio_start": 10,"audio_end": 25}],"style": "专业严谨"}
```

（2）字段说明

环节	字段名称	来源	含义
输入	`compliance_agent_output`	**上个 Agent (A3)**	获取审核通过的最终脚本内容

	`duration_params`	**全局上下文 (A1写入)**	获取目标时长及图片描述语速标准
	`characters`	**全局上下文**	获取指定的角色音色要求
输出	`audio_url`	**Agent 生成**	最终合成音频的访问链接
	`actual_duration`	**Agent 生成**	音频实际生成的物理时长（秒）
	`image_marks`	**Agent 生成**	图片在音频时间轴上的开始与结束时间

（3）对应 Prompt 模板

markdown

代码块

```
### 任务：生成播客音频，并输出标准化JSON格式结果
#### 输入信息：
1. 合规校验子Agent输出：{compliance_agent_output}（JSON格式）
2. 时长控制参数：{duration_params}
3. 角色要求：{characters}

#### 任务要求：
1. 严格按照以下JSON格式生成输出，字段名、类型必须完全匹配；
2. actual_duration与target_duration误差≤5%，duration_error_rate为误差百分比；
3. speed_adjustment中normal_speed为正常内容语速，image_speed为图片描述语速；
4. image_marks需标注图片描述在音频中的开始和结束时间（秒）；
5. 若任务执行失败，status设为"failed"，并在error_msg中说明原因。

#### 输出格式：
{
  "task_id": "{task_id}",
  "status": "success/failed",
  "error_msg": "",
  "audio_url": "",
  "actual_duration": 0,
  "duration_error_rate": 0,
  "speed_adjustment": {
    "normal_speed": 0,
    "image_speed": 0
  },
```

```
  "script_content": "",
  "image_marks": [
    {
      "image_url": "",
      "audio_start": 0,
      "audio_end": 0
    }
  ],
  "style": ""
}
```

配套产物agent

（1）JSON 数据格式

json

代码块

```
{"task_id": "podcast_20240520_001","status": "success","error_msg": "","subtitle_url": "https://example.com/podcast.srt","cover_url": "https://example.com/cover.png","image_description": [{"image_url": "https://example.com/image1.png","image_type": "数据图表","description": "这张图表显示2024年AI大模型市场规模同比增长80%"}],"actual_duration": 598,"podcast_topic": "AI大模型技术架构与行业应用"}
```

（2）字段说明

环节	字段名称	来源	含义
输入	`audio_agent_output`	**上个 Agent (A4)**	获取实际时长及音频关键数据
	`podcast_topic`	**全局上下文**	获取播客主题用于封面设计
输出	`subtitle_url`	**Agent 生成**	生成的 SRT 格式字幕文件链接
	`cover_url`	**Agent 生成**	1080x1080 像素的 PNG 封面图链接

	`image_description`	**Agent 生成**	最终的图片描述对照表

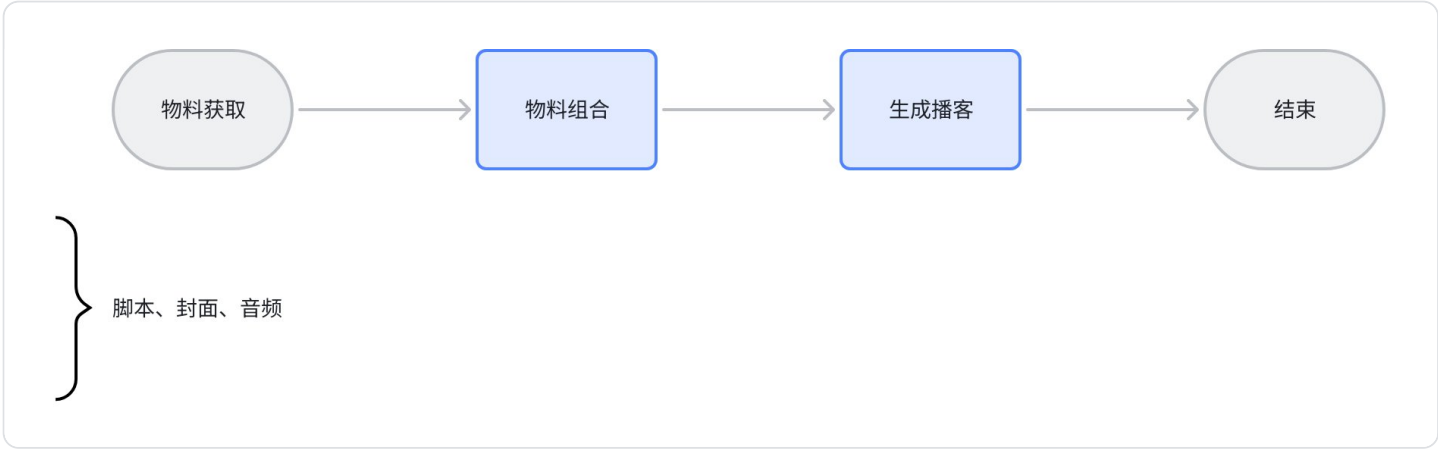
（3）对应 Prompt 模板

markdown

```
### 任务：生成字幕、封面及图片说明，并输出标准化JSON格式结果
#### 输入信息：
1. 音频合成子Agent输出：{audio_agent_output}（JSON格式）
2. 播客主题：{podcast_topic}

#### 任务要求：
1. 严格按照以下JSON格式生成输出，字段名、类型必须完全匹配；
2. subtitle_url指向SRT格式字幕文件，编码为UTF-8；
3. cover_url指向1080×1080像素的PNG格式图片；
4. image_description需包含图片URL、类型和描述内容；
5. 若任务执行失败，status设为"failed"，并在error_msg中说明原因。

#### 输出格式：
{
  "task_id": "{task_id}",
  "status": "success/failed",
  "error_msg": "",
  "subtitle_url": "",
  "cover_url": "",
  "image_description": [
    {
      "image_url": "",
      "image_type": "",
      "description": ""
    }
  ],
  "actual_duration": 0,
  "podcast_topic": ""
}
```

物料组合agent

逻辑梳理



将所有子 Agent 的产物封装为最终 API 返回值

- **输入来源**：接收来自所有上游 Agent（主要是 A4 和 A5）的 JSON 数据
- **核心输出**：
  - `audio_url`（来自 A4）
  - `subtitle_url`（来自 A5）
  - `cover_url`（来自 A5）()。
  - `image_description`（来自 A5）
  - `actual_duration`（来自 A4）

json格式如下：

json

代码块

```
{"code": 200,"message":
"success",
"data":
{
"task_id": "podcast_20240520_001",
"audio_url": "https://example.com/podcast.mp3","subtitle_url":
"https://example.com/podcast.srt",
"cover_url": "https://example.com/cover.png",
"actual_duration": 598,
"podcast_topic": "AI大模型技术架构与行业应用",
"image_description": [{"image_url":
"https://example.com/image1.png","image_type": "数据图表","description": "这张图表显示2024年AI大模型市场规模同比增长80%"}],
"suggestion": ""
}
}
```

评测集

评测集

多维表格 [该内容不支持导出查看]

评分标准

评测维度 \ 分值	1分 (严重缺陷)	2分 (不及格)	3分 (合格/及格)	4分 (良好/商用)	5分 (卓越/标杆)
口语化程度 (Orality)	完全是书面报告，充斥“首先、综上所述”等公文词汇。	像是在读课文，句子过长，缺乏呼吸感和断句。	基本的对话体，但表达略显死板，口语词（如“其实、对了”）较少。	表达自然流畅，短句为主，有适量的垫声词，听感较轻松。	极具感染力，完美模拟人类说话习惯，含自然转折、语气助词。
角色一致性 (Persona)	角色身份混淆，主持人回答了自己的提问，或语序逻辑错乱。	主持人与嘉宾说话语气完全一致，缺乏性格区分。	角色分工明确，但性格扁平，仅仅是“一问一答”的机器。	人设稳定，主持人能引导话题，嘉宾表现出专业或特定的性格色彩。	性感鲜明且贯穿始终，互动中有追问、打断或即兴感，角色跃然纸上。
知识转化力 (Synthesis)	直接照抄 A1 的解析原文，完全没有针对音频听觉进行转化。	只是简单地在原文前加了“主持说：”，内容依然艰涩难懂。	能将复杂术语进行简单改写，但缺乏生动比喻，听众易疲劳。	能用通俗易懂的语言解释专业概念，逻辑条理清晰。	极佳的类比能力，能将枯燥数据化为生动故事或生活案例，易于理解。
节奏与钩子 (Pacing/Hooks)	结构散乱，没有中心思想，听众无法坚持听完前 30 秒。	线性叙事，毫无波澜，信息密度分布极不均匀。	有开场白和结束语，中间过程平铺直叙，缺乏亮点。	开头有吸引人的“钩子”，重点突出，节奏有快慢变化。	黄金节奏，高潮迭起，通过悬念、冲突或共鸣紧紧抓住听众注意力。
逻辑连贯性 (Coh	话语间跳跃巨大，前后文完全不搭边，存在严重的语义断层。	逻辑生硬，转场完全依赖“接下来我们看下一个点”等模版。	话题切换基本合理，但深度不足，转场稍显突兀。	话题过渡自然，能通过上文的内容引出下文，逻辑闭环。	丝滑转场，各章节间环环相扣，像是一场精心准备的深度对谈。

erence)					

落地实施注意事项

1. **字段校验**：各 Agent 接收上游数据时，需对必填字段、字段类型、取值范围进行校验，避免数据格式错误导致流程中断。
2. **错误处理**：任一 Agent 执行失败时，需返回明确的错误信息，主 Agent 需捕获错误并告知用户，同时提供重试选项。
3. **版本兼容**：后续规则迭代时，如需新增字段，需保持原有字段兼容，避免影响旧版本 Agent 的正常协作。
4. **数据安全**：涉及用户隐私、敏感内容的字段，需进行加密处理，确保数据传输安全。