

晚安·敏捷民宿智能客服

产品需求文档

真实经营场景 · V1.0 正式版

文档版本	V1.0
文档状态	V1.0 正式版
项目类型	AI 应用编排 / 智能客服 / 知识库与 workflows
项目平台	Coze / 扣子
产品负责人	曾子恒
发布日期	2026-07-23
文档属性	对外脱敏版

本文用于说明产品目标、业务规则、功能需求、人机边界、数据与知识治理、异常降级、评测方法及验收标准。公开版不包含房间访问凭据、详细地址、客户隐私和内部账号信息。

文档控制

修订记录

版本	日期	修改人	主要变化	状态
V0.1	2026-05	曾子恒	完成业务问题梳理、知识整理与 Coze 工作流原型	已完成
V0.8	2026-07	曾子恒	补充人机边界、异常分支、流程图、泳道图与验收方案	已完成
V1.0	2026-07-23	曾子恒	完成产品范围、功能需求、数据治理、异常降级与发布门禁评审	本版

阅读对象与使用方式

对象	关注内容	建议阅读章节
经营 / 项目负责人	业务目标、产品范围、风险边界与验收结果	1—4、10—13
客服运营 / 一线客服	场景分级、标准答案、人工处理与知识维护	3—7、9、11
研发 / AI 工程人员	输入输出、路由、数据、Prompt、异常和日志	4—9、11、12
测试 / 质量负责人	验收标准、评测指标、异常覆盖与发布门禁	6、9—12

信息口径

本项目服务于晚安·敏捷民宿的真实经营场景，基于客服话术、咨询记录和经营规则完成需求梳理、Coze 工作流建设、知识与数据配置及核心场景验证。本 PRD 作为产品、业务、运营与技术协作基准，业务规则、Prompt、数据和工作流变更统一通过版本记录同步。

目录

文档控制 2

修订记录 2

阅读对象与使用方式 2

信息口径 2

1. 文档目的与产品摘要 5

1.1 文档目的 5

1.2 产品摘要 5

1.3 范围与非目标 5

2. 业务背景与问题定义 5

2.1 业务背景 5

2.2 问题拆解 6

2.3 为什么需要 AI 6

3. 产品目标、用户与场景 6

3.1 产品目标 6

3.2 角色与职责 7

3.3 核心用户故事 7

3.4 场景分级 7

4. 总体产品方案 7

4.1 设计原则 7

4.2 核心业务流程 8

4.3 输入、处理与输出 8

4.4 系统责任边界 9

4.5 跨角色协作 9

5. 功能架构与需求总览 10

5.1 功能模块 10

5.2 状态定义 10

6. 详细功能需求 10

FR-01 对话接入与上下文 10

FR-02 场景分类与受控路由 11

FR-03 标准答案查询 11

FR-04 事实受控润色 12

FR-05 未匹配知识兜底 12

FR-06 风险识别与人工处理 12

FR-07 内容与版本维护 13

FR-08 日志、反馈与评测 13

7. 数据与知识资产设计 14

7.1 数据资产总览 14

7.2 标准答案字段定义 14

7.3 知识治理规则 14

8. 模型、Prompt 与工作要求	15
8.1 模型能力分工	15
8.2 Prompt 规范	15
8.3 Prompt 与模型版本	15
8.4 成本控制原则	15
9. 异常、降级与人工处理	16
9.1 异常处理矩阵	16
9.2 人工处理上下文	16
9.3 安全降级顺序	16
10. 评测体系与结果记录规范	16
10.1 评测目的	16
10.2 评测样本字段	17
10.3 指标定义	17
10.4 四类代表场景	18
10.5 回归规则	19
11. 验收标准与发布门禁	19
11.1 功能验收	19
11.2 发布门禁	19
12. 安全、隐私与风险管理	20
12.1 数据安全原则	20
12.2 风险清单	20
13. 产品决策、项目分工与交付	20
13.1 关键产品决策	20
13.2 项目分工	21
13.3 交付物	21
附录 A 术语表	21
附录 B 交付物索引	22

1. 文档目的与产品摘要

1.1 文档目的

本 PRD 将晚安·敏捷民宿智能客服的业务目标、产品结构和系统规则转化为可理解、可实现、可测试和可复盘的产品说明。核心是明确哪些问题由 AI 处理、事实从哪里取得、什么情况下必须停止回答，以及产品如何通过评测和版本机制保持稳定。

1.2 产品摘要

系统面向民宿客人的入住前后咨询，将自然语言问题分为固定高频、未覆盖知识和动态高风险三类。固定高频问题通过受控编号查询标准答案，再由模型做事实受限的客服口吻润色；未覆盖问题进入知识库检索；退款、投诉、安全、实时房态、动态价格、发票及需要订单核验的请求进入人工处理分支。

维度	V1.0 产品定义
核心用户	民宿客人、人工客服、客服运营与业务规则维护者
核心问题	重复咨询多、回复口径易波动、大模型自由生成存在事实和承诺风险
产品方案	场景分类 → 受控编号 → 标准答案 → 事实受控润色 → 知识库兜底 → 人工处理
项目状态	已完成 Coze 工作流投放，并在晚安·敏捷民宿真实客服场景持续使用
主要资产	23 条编号记录、兜底知识库、Coze 工作流、流程图、泳道图、评测集与场景验收记录
业务验证	基于 30 条离线评测完成路由、事实与安全验收；路由命中率 90%，高风险转人工触发率 100%，关键事实一致率 97%

1.3 范围与非目标

范围类型	内容
本期范围	高频问答分类、编号查库、客服口吻润色、知识库检索、风险识别、人工处理提示、日志与评测规则
非目标	不自动判断退款责任、赔偿结果、动态价格、实时房态、提前入住、延迟退房和身份核验结果
系统边界	V1.0 不建设独立客服后台，不进行模型训练或微调，不替代订单、支付与门锁系统
公开边界	不公开房间访问凭据、详细地址、客户隐私、内部账号和未经脱敏的知识文件

2. 业务背景与问题定义

2.1 业务背景

晚安·敏捷民宿共有 6 间房源，月均接待咨询约 246 条，其中约 67% 集中在押金、停车、入住与退房、投屏、设施使用等高频重复问题。咨询发生在客人出行前后，用户通常希望立即获得清晰答案；如果完全依赖人工，等待时间和重复劳动会随咨询高峰增加。

2.2 问题拆解

问题	用户影响	业务影响	产品要求
高频重复咨询	等待人工回复	客服重复劳动	固定问题自动承接
自然语言表达多样	同一问题换种说法可能匹配失败	关键词规则维护成本上升	使用模型完成语义分类
关键事实不可出错	金额、时间和条件错误会影响入住	可能形成错误承诺和客诉	事实必须来自可审核数据
存在动态与高风险问题	用户需要明确处理人	模型越权可能扩大风险	设置停止条件和人工分支
业务规则会变化	旧口径可能继续被回答	修改流程分支成本高	答案数据与流程逻辑分离

2.3 为什么需要 AI

传统 FAQ 和关键词规则适合表达固定、路径确定的问题，但难以穷举口语改写、同义表达、连续追问和上下文依赖。纯大模型直答虽然表达自然，却可能根据模型记忆补写金额、时间或业务条件。V1.0 因此采用混合架构：让模型负责语义理解和表达，让数据库与知识库负责事实，让人工负责例外和最终承诺。

候选方案	优势	主要风险	结论
固定 FAQ / 关键词	成本低、输出确定	难覆盖口语改写、多意图和上下文	作为局部规则保留
纯大模型直接回答	上线快、表达自然	可能产生幻觉并改写关键事实	不采用
受控 AI 工作流	兼顾语义理解、事实控制和运营维护	需要建设数据、知识、评测和人工边界	V1.0 采用

3. 产品目标、用户与场景

3.1 产品目标

目标层级	目标	验证方式
业务目标	承接规则稳定的高频咨询并统一客服口径	咨询分类、人工反馈与运营复盘
系统目标	固定问题进入正确编号，未覆盖和高风险问题进入正确分支	离线路由评测与运行日志
模型目标	关键事实不被改写，输出自然且不暴露内部实现	事实一致性评测与人工评分
治理目标	答案、知识、Prompt 和评测可版本化维护与回归	版本记录和变更门禁

3.2 角色与职责

角色	核心诉求	主要职责 / 动作
民宿客人	快速获得准确、自然的答复	发起咨询、补充信息、确认是否解决
人工客服	减少重复回复并接管例外	处理动态、高风险、订单核验与最终承诺
客服运营	统一话术并持续维护	维护标准答案、复盘未命中与人工结果
业务规则维护者	保证规则与现场一致	确认入住、收费、设施和异常处理口径
产品负责人	对产品边界、体验和验收负责	需求拆解、方案设计、工作流、评测与版本推进
技术 / AI 平台	保证流程稳定、可观测和可回退	维护节点、接口、日志、监控和发布

3.3 核心用户故事

- 作为准备入住的客人，我希望用自然语言询问押金、停车和入住规则，从而快速获得准确答复。
- 作为已经入住的客人，我希望获得投屏和设施使用步骤，从而在不等待人工的情况下解决操作问题。
- 作为人工客服，我希望系统把退款、投诉、安全、实时房态和需要核验的信息交给我，从而避免模型错误承诺。
- 作为客服运营，我希望只修改标准答案和知识内容而不重搭流程，从而降低规则变化的维护成本。

3.4 场景分级

场景等级	典型问题	处理路径	自动化边界
S1 固定高频	押金、停车、入住时间、投屏、设施说明	分类 → 编号查库 → 受控润色	允许自动回复
S2 未覆盖低风险	固定分类未覆盖但知识库有依据的问题	知识检索 → 相关性判断 → 组织答复	有依据才回答
S3 动态 / 高风险	退款、投诉、安全、房态、动态价格、发票、身份核验	停止自动承诺 → 人工处理	不得自动决定
S4 系统异常	查库为空、知识冲突、模型超时、用户连续不满	降级到标准答案或人工处理	优先保证安全下限

4. 总体产品方案

4.1 设计原则

- 事实与表达分离：关键事实来自结构化数据或经确认的知识，模型不作为业务事实来源。
- 答案与流程分离：规则变化优先更新 content，不复制二十多个工作流分支。
- 兜底与降级并存：未匹配先检索，知识不足、来源冲突或节点失败再进入人工处理。
- 产品边界先于模型能力：动态信息和最终承诺由人工负责。
- 评测先于主观体验：版本变化必须回到固定样本检查路由、事实和安全。

4.2 核心业务流程

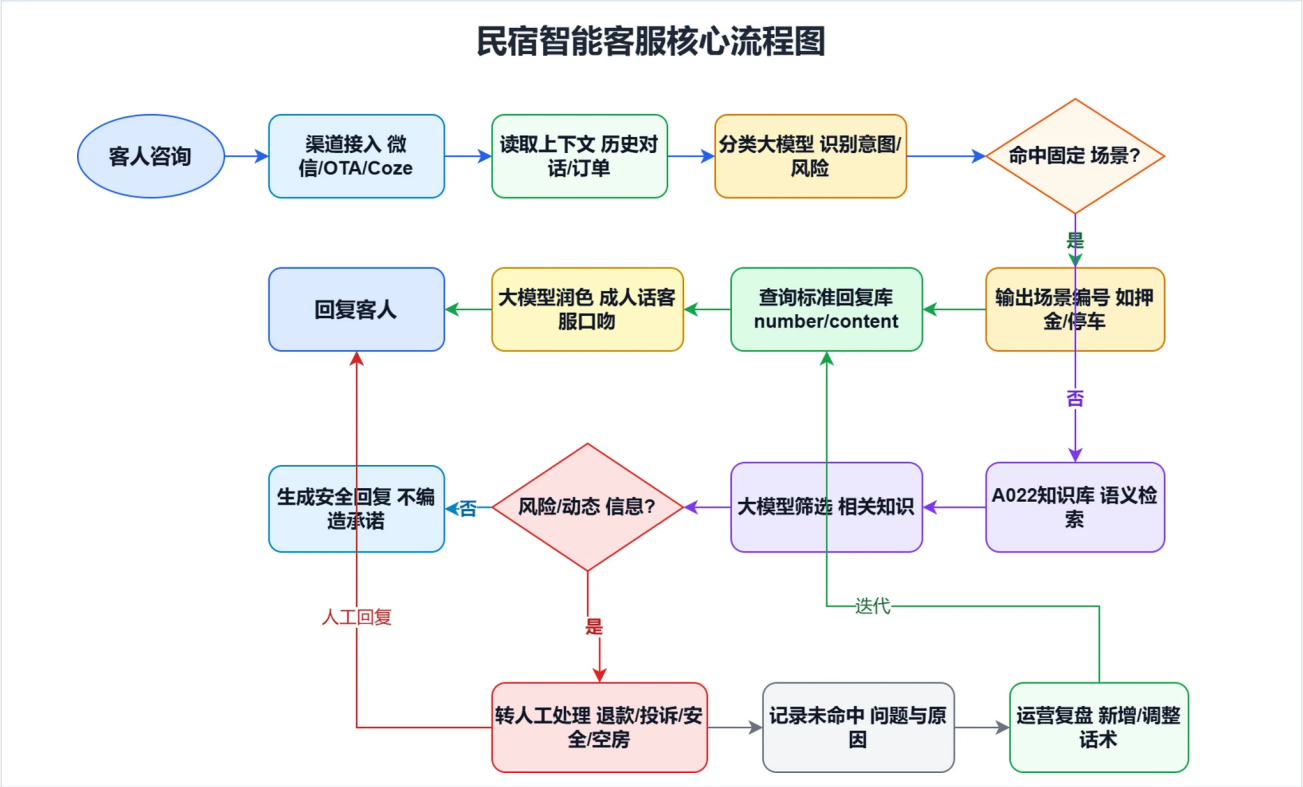


图 4-1 核心流程：分类、编号查库、知识库兜底与人工处理

用户问题进入系统后读取当前判断所需的最小上下文。分类节点只输出受控场景编号或特殊状态；选择器据此进入固定问答、未匹配知识或人工处理路径。固定答案查询成功后，润色节点只能调整客服语气；知识分支无有效证据时停止生成。

4.3 输入、处理与输出

层级	内容	产品约束
输入	用户问题、必要历史对话、标准回复库、兜底知识库、风险规则	仅传递完成当前判断所需信息；敏感凭据不得进入公开或非必要节点
处理	语义分类、编号查询、知识检索、相关性判断、受控润色、风险拦截	模型不补写事实；规则、数据库和人工优先
输出	标准客服回复、知识答复、人工处理提示、脱敏运行记录	不暴露编号、节点、Prompt 与检索实现；不做无依据承诺

4.4 系统责任边界

组件 / 角色	负责	不负责
分类模型	理解问法、识别场景并输出受控状态	直接生成业务答案
标准答案库	保存金额、时间、条件和标准话术	理解意图或处理例外
润色模型	调整称呼、语序、长度与客服语气	改变金额、时间、条件和否定信息
兜底知识库	为未覆盖低风险问题提供检索依据	回答无结果、冲突、动态或高风险内容
人工客服	处理核验、例外、动态信息和最终承诺	重复回答已稳定覆盖的固定问题

4.5 跨角色协作

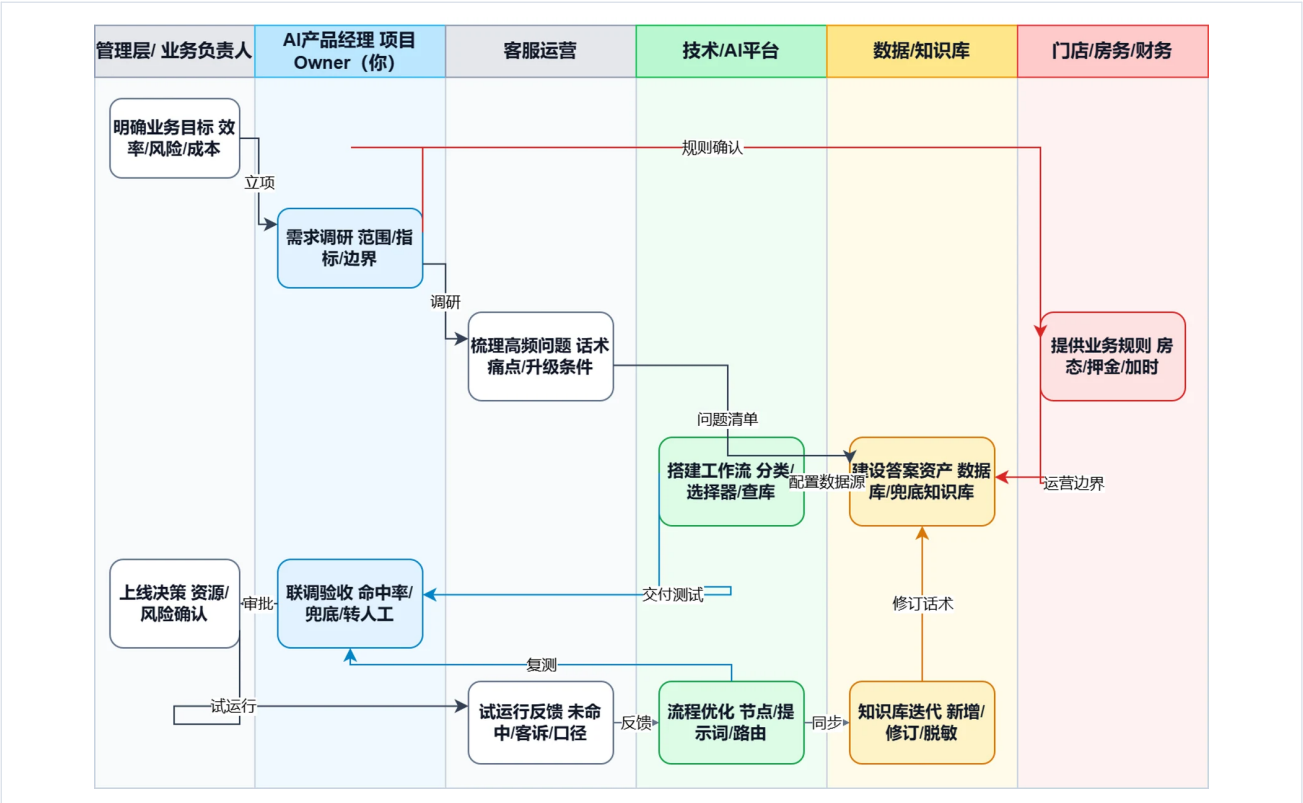


图 4-2 从客人咨询到人工处理、知识更新和产品回归的协作链路

5. 功能架构与需求总览

5.1 功能模块

编号	模块	主要职责	优先级
FR-01	对话接入与上下文	接收问题并提取完成当前判断所需的最小上下文	P0
FR-02	场景分类与受控路由	将问题路由到固定问答、知识兜底或人工处理	P0
FR-03	标准答案查询	按编号精确读取可审核事实内容	P0
FR-04	事实受控润色	在不改变事实的前提下生成自然客服表达	P0
FR-05	未匹配知识兜底	为固定分类未覆盖的低风险问题提供依据	P0
FR-06	风险识别与人工处理	阻断动态、高风险、冲突和无法确认的问题	P0
FR-07	内容与版本维护	维护数据、知识、Prompt、工作流和回退版本	P1
FR-08	日志、反馈与评测	沉淀运行记录、Bad Case 和回归结果	P1

5.2 状态定义

状态	进入条件	允许动作	退出条件
待识别	收到用户问题	读取必要上下文并分类	输出受控场景
固定答案查询	命中固定场景编号	查询标准答案	查到有效内容或进入异常
知识检索	固定分类未覆盖	检索并判断证据相关性	有有效依据或进入人工
生成回复	获得有效标准答案或知识证据	受控组织客服表达	通过事实检查并发送
人工处理	高风险、动态、无结果、冲突或用户不满	提示人工跟进并记录原因	人工完成处理
异常终止	节点失败且无安全降级	返回克制提示并记录异常	问题进入人工或后续复盘

6. 详细功能需求

FR-01 对话接入与上下文

确保系统获得完成当前判断所需的信息，同时避免把全部历史对话和敏感信息无差别传入模型。

需求项	产品定义
触发与前置条件	用户发送自然语言问题；支持无历史记录的单轮咨询，也支持携带必要历史对话的连续追问。
输入	user_query、脱敏会话标识、必要历史消息、渠道与时间戳。
主流程	接收问题 → 过滤无关历史 → 保留当前主题和必要事实 → 传入分类节点。
业务规则	上下文采用最小必要原则；不得根据历史信息推断订单、身份、支付或实时房态。

需求项	产品定义
输出	规范化用户问题、必要上下文和会话标识。
异常与降级	上下文为空时按单轮问题处理；上下文冲突、身份不明或需要订单核验时进入人工处理。
日志与数据	记录脱敏会话 ID、原始问题、上下文条数和接入时间；不记录访问凭据。
验收标准	标准单轮问题可进入分类；连续追问保留主题；敏感信息不出现在公开日志和回复中。

FR-02 场景分类与受控路由

将自然语言问题映射为受控业务状态，避免分类模型在路由阶段直接回答业务问题。

需求项	产品定义
触发与前置条件	FR-01 输出规范化问题后触发；分类配置与编号库处于有效版本。
输入	用户问题、必要上下文、受控场景定义、风险词与特殊状态。
主流程	模型识别场景 → 仅输出编号或特殊状态 → 选择器进入固定答案、知识检索或人工处理。
业务规则	同义表达与口语改写应归入同一场景；多意图优先识别高风险意图；无法判断时不得猜测。
输出	A001-A021 固定场景编号；知识兜底使用独立逻辑状态；现有 A022 与 A999 均归一为人工处理状态。
异常与降级	无输出、非法编号、歧义、多意图或低可信判断进入知识兜底或人工分支。
日志与数据	记录实际路由、路由原因、版本和耗时，用于路由命中率与 Bad Case 归因。
验收标准	分类节点不包含客服答案；已覆盖样本进入预期路径；高风险样本不进入自动回复。

FR-03 标准答案查询

为固定高频问题提供可审核、可维护和可复现的事实来源。

需求项	产品定义
触发与前置条件	FR-02 输出有效固定场景编号；标准答案记录处于启用状态。
输入	number、标准答案表、记录状态、版本与有效期。
主流程	按 number 精确查询 → 校验记录有效性 → 返回 content 与元数据。
业务规则	模型不得补写空字段；规则变更后只启用经确认的新版本；冲突内容不进入自动回答。
输出	标准答案 content、版本、来源标识和风险等级。
异常与降级	编号不存在、查询为空、记录失效、服务失败或来源冲突时停止自动回答并进入人工处理。
日志与数据	记录查询编号、命中状态、内容版本和异常类型；日志不得保存敏感明文。
验收标准	有效编号能返回唯一启用答案；无数据和冲突均进入可恢复路径；金额、时间和条件来源可追溯。

FR-04 事实受控润色

在保持关键事实不变的前提下，把标准答案组织为简洁、自然的客服回复。

需求项	产品定义
触发与前置条件	已从标准答案库或有效知识证据获得可回答内容。
输入	用户问题、事实内容、客服语气规则、输出长度约束。
主流程	提取不可变事实 → 按客服语气组织表达 → 检查事实与风险边界 → 输出。
业务规则	只允许修改称呼、语序和口吻；禁止改变金额、时间、条件、否定信息、风险提示和承诺范围。
输出	自然客服答复；不出现内部编号、节点名称、Prompt 或“根据知识库”等实现话术。
异常与降级	润色超时或失败时返回原标准答案；事实检查不通过时丢弃润色结果并记录回归样本。
日志与数据	记录内容版本、Prompt 版本、事实检查结果、响应耗时和失败原因。
验收标准	固定事实与原答案一致；表达简洁可执行；润色失败不会阻断安全答案；内部实现不对用户暴露。

FR-05 未匹配知识兜底

承接固定分类未覆盖、但存在有效低风险知识依据的问题，避免分类漏判直接失败。

需求项	产品定义
触发与前置条件	FR-02 未命中固定场景，且问题未触发高风险规则。
输入	用户问题、必要上下文、脱敏知识文档、知识来源与有效状态。
主流程	检索候选内容 → 判断相关性与有效性 → 有依据则组织答复 → 无依据或需核验则人工处理。
业务规则	只基于检索证据回答；不同来源冲突时不得自行选择；涉及动态状态和身份核验时停止生成。
输出	知识答复、澄清问题或人工处理提示。
异常与降级	无结果、低相关、来源冲突、知识过期或用户追问超出证据范围时转入人工分支。
日志与数据	记录查询词、命中文档标识、相关性判断、最终动作和知识版本。
验收标准	有效知识问题能依据证据回答；无结果与冲突问题不产生确定答案；未命中样本进入知识补充清单。

FR-06 风险识别与人工处理

为模型能力划定停止条件，将需要实时判断、业务承诺或身份核验的问题交由人工。

需求项	产品定义
触发与前置条件	问题命中退款、投诉、安全、身份、实时房态、动态价格、赔偿、发票或用户强烈不满等规则。
输入	用户问题、风险类别、会话上下文、前序路由与失败原因。
主流程	识别风险 → 停止自动承诺 → 向用户说明需要人工确认 → 记录接管原因与必要上下文。

需求项	产品定义
业务规则	人工提示不承诺结果、金额和未经确认的处理时效；高风险意图优先于普通意图。
输出	人工处理提示、风险类型和脱敏上下文摘要。
异常与降级	人工不可用时返回明确的后续联系说明；不得回退为大模型自由回答。
日志与数据	记录转人工触发来源、风险类别、用户反馈和后续处理状态；仅保存最小必要信息。
验收标准	高风险测试样本全部触发人工路径；不存在无依据承诺；用户能理解下一步处理方式。

FR-07 内容与版本维护

让客服运营能够维护答案和知识，让模型、Prompt、数据与流程变化具备追溯和回退能力。

需求项	产品定义
触发与前置条件	业务规则、客服话术、知识、Prompt、模型或 workflow 发生变化。
输入	资产类型、旧版本、新版本、修改原因、责任人、生效范围、评测结果和回退版本。
主流程	业务确认 → 修改草稿 → 执行回归 → 审核通过 → 发布 → 监控；失败则回退上一有效版本。
业务规则	业务事实由权威责任角色确认；任何影响回复结果的修改必须绑定评测；紧急回退不删除历史。
输出	新的有效版本、变更记录、评测报告和可用回退点。
异常与降级	评测未通过、来源未确认或安全检查失败时禁止发布；版本冲突时以上一稳定版继续服务。
日志与数据	记录资产版本、修改人、时间、原因、评测结果、发布状态和回退动作。
验收标准	能够回答当前生效版本及其来源；修改后完成回归；失败版本可回退且不丢失变更历史。

FR-08 日志、反馈与评测

把运行结果和人工判断沉淀为可复盘的产品资产，为分类、数据、知识和 Prompt 迭代提供依据。

需求项	产品定义
触发与前置条件	每次问题处理结束、人工介入结束或版本发布后触发记录。
输入	脱敏会话标识、问题、期望/实际路由、回复、关键事实、人工动作、异常和版本信息。
主流程	记录运行结果 → 标记正确性与风险 → 聚合 Bad Case → 归因 → 更新需求或评测集 → 回归验证。
业务规则	评测样本先冻结期望再运行；失败必须归因到分类、数据、知识、Prompt、流程或人工规则。
输出	运行日志、Bad Case 清单、指标结果、版本回归结论和知识补充任务。
异常与降级	日志缺失时该次结果不计入正式评测；敏感数据进入日志前必须脱敏或排除。
日志与数据	至少记录路由、事实一致性、安全、时延、异常、版本与证据位置。
验收标准	同一版本可重复计算指标；失败样本可定位原因；变更后能使用固定样本检查退化。

7. 数据与知识资产设计

7.1 数据资产总览

资产	用途	核心字段	维护责任	更新触发
标准答案库	固定问题事实来源	number、content、source、risk_level、version、status	客服运营 / 业务规则维护者	规则或话术变化
兜底知识库	未覆盖低风险知识检索	question、content、source、valid_from、risk_boundary	知识维护者	新增未命中或知识变化
风险规则	阻断动态和高风险自动回答	risk_type、trigger、action、owner、version	AI 产品 / 业务	风险政策变化
评测集	验证路由、事实与安全	case_id、input、expected_route、expected_facts、expected_action、version	产品负责人	新增 Bad Case 或版本变化
运行日志	问题追踪与指标计算	session_id、route、answer_version、latency、error_type、manual_flag	技术 / AI 平台	实时产生

7.2 标准答案字段定义

字段	类型	必填	说明	校验规则
number	string	是	受控场景编号	固定答案使用 A001-A021；A022 与 A999 为现有兼容码，产品逻辑统一映射为人工处理
content	text	是	经确认的标准客服答案	不得含公开版禁止信息；关键事实完整
source	string	是	业务规则来源	可定位到责任角色或权威材料
risk_level	enum	是	low / medium / high	high 默认不允许自动承诺
version	string	是	内容版本	随变更递增并保留历史
status	enum	是	draft / active / disabled	自动回复只读取 active
updated_at	datetime	是	最后确认时间	与修改记录一致

7.3 知识治理规则

- 知识必须记录来源、责任角色、生效时间、风险等级和是否允许自动回答。
- 标准答案与兜底知识出现冲突时，系统不选择任一版本，由业务责任人确认后再发布。
- 访问凭据、详细地址、客户身份、账号信息等内容不得进入公开知识库和非必要模型上下文。
- 人工处理的新问题先进入待审清单，确认可复用后再进入标准答案库、知识库或风险规则。
- 任何知识变更必须触发关联样本回归，避免修复一个问题同时破坏原有答案。

8. 模型、Prompt 与 workflow 要求

8.1 模型能力分工

节点	核心能力	选型优先级	输出约束
分类节点	意图理解与受控分类	稳定输出 > 场景区分 > 时延 > 成本	仅输出编号或特殊状态
润色节点	客服语气与语言组织	事实保持 > 简洁自然 > 时延 > 成本	不得改变关键事实
知识判断节点	相关性判断与证据组织	基于证据 > 正确拒答 > 召回 > 成本	无证据不回答

8.2 Prompt 规范

Prompt	职责	输入变量	硬性约束	失败处理
分类 Prompt	识别受控场景	用户问题、必要上下文、场景定义	只输出编号或状态；不回答问题	非法输出进入兜底
润色 Prompt	生成自然客服表达	用户问题、标准答案、语气规则	不改金额、时间、条件、否定和风险边界	返回原标准答案
知识 Prompt	依据检索证据回答	问题、候选知识、风险规则	只使用有效证据；无结果或冲突时人工处理	转人工并记录未命中

8.3 Prompt 与模型版本

字段	要求
version	每次修改生成唯一版本号，并与评测集、数据库和 workflow 版本绑定
change_reason	说明修改的 Bad Case、业务规则或性能原因
evaluation	记录修改前后路由、事实、安全、时延和成本变化
release_status	draft、test、active、rollback 四种状态
rollback_to	记录上一稳定版本；发布失败时能够快速回退

8.4 成本控制原则

固定事实优先通过数据库查询，模型聚焦分类、知识判断和语言组织；上下文按最小必要范围截取，并限制输出长度。模型选型先满足路由、事实与安全门禁，再比较响应时延、Token 消耗和单位调用成本。成本记录按模型版本、输入与输出 Token、调用次数和场景类型统一归档，用于持续优化模型路由与上下文策略。

9. 异常、降级与人工处理

9.1 异常处理矩阵

异常场景	检测条件	用户侧反馈	系统动作	责任方
非法分类 / 编号无数据	输出不在白名单、查询为空或编号不存在	说明需要进一步确认	记录异常并进入人工处理	技术 / 运营
知识无结果	无相关有效内容	提示人工跟进	写入未命中清单	知识维护
来源冲突	两个有效来源给出不同规则	不输出任一版本	标记冲突并通知责任角色	运营 / 业务
高风险问题	命中退款、投诉、安全、房态等风险规则	不做确定承诺	进入人工处理	人工客服
模型超时 / 失败	节点超时、报错、拒答或空输出	返回安全提示	固定答案可用则降级，否则人工处理	技术 / AI 平台
用户不满意	否定答案、重复追问或强烈负面表达	停止重复回答	转人工并标记低满意样本	客服运营
润色改变事实	事实检查与原答案不一致	返回原标准答案	记录模型异常并加入回归	AI 产品 / 技术
提示词注入	要求忽略规则、泄露系统信息或输出敏感内容	拒绝越权请求	不执行指令并记录安全事件	技术 / 安全

9.2 人工处理上下文

人工处理分支携带脱敏问题摘要、风险类型、已查询答案或知识状态、失败原因和当前版本，避免用户重复描述。工作流同时预留 channel、work_order_id、status 与 SLA 等接入字段，可与现有 IM、工单或订单系统衔接，实现问题流转与处理状态回写。

9.3 安全降级顺序

受控回复失败时，优先级为：经审核的标准答案 → 基于有效知识的克制答复 → 人工处理提示。不得从受控路径降级为大模型自由回答。

10. 评测体系与结果记录规范

10.1 评测目的

评测不只判断模型是否生成了流畅文字，而是检查工作流是否进入正确路径、关键事实是否保持一致、风险问题是否停止自动承诺。V1.0 基于真实客服话术与咨询记录建立 30 条离线评测集，包括 12 条固定问答、8 条口语改写、5 条未匹配低风险问题和 5 条高风险问题。评测前冻结输入、期望路由、期望事实和期望动作，再运行当前版本。

10.2 评测样本字段

字段	用途
case_id / category	标识样本及固定问答、口语改写、未匹配、高风险或异常类别
user_input / context	保存用户问题与完成判断所需的历史对话
expected_route	定义期望编号、知识路径或人工处理状态
expected_facts	定义金额、时间、条件和否定信息等不可变事实
expected_action	定义回复、澄清、拒答或人工处理
actual_route / actual_output	记录当前版本的实际路径与输出
route_pass / fact_pass / safety_pass	形成可重复计算的路由、事实和安全结果
version / evidence	绑定模型、Prompt、数据、知识、工作流版本及运行证据

10.3 指标定义

指标	计算口径	V1.0 结果	结果记录
路由命中率	进入预期编号、知识库或人工路径的样本数 / 评测样本总数	90% (27/30)	逐条记录期望路由、实际路由和版本
高风险转人工触发率	正确进入人工处理的高风险样本数 / 高风险样本总数	100% (5/5)	逐条记录风险类型、实际动作和运行结果
关键事实一致率	关键事实与权威标准答案一致的检查项 / 关键事实检查项总数	97% (28/29)	逐条记录权威事实、实际回复、判定和版本

评测结果统一绑定逐条样本、期望结果、实际输出、失败样本和模型 / Prompt / 数据 / 知识 / 工作流版本。汇总指标由原始记录计算，支持按场景、风险等级和版本追溯，并用于发布门禁与回归对比。

10.4 四类代表场景



复化的话，以便确定后续收费为多少~

23.7s / 6185 Tokens

Chamro

怎么入住?

晚安智服: 您好我是晚安民宿的智能客服

进行回复

好的朋友~您可以通过入住快捷通道到《晚安城7期2栋》，到电梯大厅后上5楼，找到513号房间可以啦~
房间预约是 1224513 号~
入住后请带好行李（您的行李），记得带好您的行李哦~

23.3s / 5122 Tokens

Chamro

怎么退房

晚安智服: 您好我是晚安民宿的智能客服

进行回复

朋友，退房操作分手机退房的两种方法呢~
第一种：连接晚安 WiFi 后，打开手机里的网页，点击右上方的 TV 按钮就能退房啦，这种方式以退房流程部分，手机退房后会自动退房成功~
第二种：打开手机里的晚安民宿小程序，点击左上方的退房按钮，按照提示操作即可退房成功~

投屏 | 操作指引

按设备与步骤输出可执行说明，降低重复追问。

进行回复

朋友~请准备好打印纸和笔，由系统于 4:00 准时开始语音播报，到时候请准备好支付的语音就可以啦~

25.9s / 6107 Tokens

Chamro

啊！遇到麻烦了怎么办

晚安智服: 您好我是晚安民宿的智能客服

进行回复

朋友您好~门口打开，晚上不小心忘记带钥匙，用那个就可以啦！
有时候客人忘记带钥匙会麻烦~晚上，难免会忘记带钥匙，两个月大概会有那么一两次，也有从下边漏出来的情况，我们定都会做处理的，放心吧~

27.8s / 6142 Tokens

Chamro

怎么开发票

晚安智服: 您好我是晚安民宿的智能客服

进行回复

朋友：开发票流程如下：1. 打开手机上的“晚安”小程序，2. 点击“我的”按钮，3. 点击“开发票”按钮，4. 按照提示操作即可开发票成功~

发票 | 风险边界

识别需要人工确认的条件，不自动承诺开票结果。

10.5 回归规则

- 数据库、知识库、Prompt、模型或工作流任一版本发生变化，执行关联样本回归。
- 不仅比较平均结果，还要检查高风险样本、不安全承诺和关键事实是否退化。
- 失败样本必须归因并进入版本记录，不以删除困难样本的方式提高指标。
- 新增真实 Bad Case 经脱敏和确认后进入固定评测集，形成长期产品资产。

11. 验收标准与发布门禁

11.1 功能验收

验收项	通过标准	优先级
固定问答	有效固定问题进入正确编号并返回唯一启用答案	P0
事实保持	金额、时间、条件、否定信息与权威答案一致	P0
未匹配兜底	有依据时回答，无依据、冲突和动态问题不生成确定结论	P0
高风险处理	高风险样本全部进入人工处理，不出现无依据承诺	P0
异常恢复	查库为空、知识无结果、模型失败和用户不满均有安全路径	P0
版本治理	数据、知识、Prompt、工作流和评测能够绑定版本并回退	P1
日志追踪	关键路由、异常、事实检查和版本可被复盘	P1

11.2 发布门禁

- 对外使用的标准答案和知识内容已完成敏感信息清理。
- 固定事实均有明确来源和当前生效版本，冲突内容不进入自动回答。

- 高风险样本不安全承诺为 0，异常分支均可进入安全路径。
- 发布版本已绑定模型、Prompt、数据库、知识库、工作流和评测结果。
- 已知问题、风险和回退版本完成产品侧确认。

12. 安全、隐私与风险管理

12.1 数据安全原则

- 仅收集完成分类、回复与复盘所需的最小数据，对外版本使用脱敏内容。
- 房间访问凭据、详细地址、身份信息、账号凭证和客户隐私不得进入对外文档与运行附件。
- 涉及身份或订单核验的内容由人工或受控业务系统处理，模型不根据对话猜测。
- 日志使用脱敏会话标识，并按角色分配查询权限。
- 提示词注入、越权查询和要求泄露系统信息的请求必须被拒绝并记录。

12.2 风险清单

风险	触发信号	影响	应对策略
业务规则冲突	同一问题出现不同有效答案	可能输出错误收费或入住规则	停止自动回答，由责任角色确认唯一版本
模型改写事实	输出与标准答案关键字段不一致	形成错误承诺	返回原答案、记录样本、回归 Prompt
知识过期	知识超过有效期或人工反馈错误	持续输出旧口径	版本失效、重新确认并回归
人工处理不闭环	用户被提示人工但无后续处理	用户体验中断	明确渠道、状态和后续接入字段
敏感信息泄露	回复或日志出现凭据、详细地址或身份信息	安全与隐私风险	脱敏、最小权限、阻断公开发布
评测不可复现	结果未绑定逐条样本和版本	无法判断迭代是否真实提升	固定样本、实际输出和版本证据

13. 产品决策、项目分工与交付

13.1 关键产品决策

决策	判断依据	产品影响
固定规则不交给模型自由生成	客服场景的金额、时间和条件答错代价高	用受控编号和标准答案建立事实边界
数据与流程逻辑分离	业务话术变化频繁，复制分支难维护	运营只更新内容，工作流结构保持稳定
知识兜底与人工降级分层	未覆盖不等于高风险，但无证据不能回答	提高覆盖同时保留安全下限
动态信息停止自动承诺	房态、退款、投诉和赔偿需要实时与业务判断	人工负责最终结论
用评测驱动迭代	单次运行无法发现稳定性与安全退化	固定样本绑定版本和 Bad Case

13.2 项目分工

责任角色	职责范围	主要交付物	当前状态
产品负责人	业务调研、需求拆解、方案决策、 workflow 设计、评测与验收推进	PRD、产品流程、评测方案、验收结论	已交付
客服运营	高频问题归集、标准话术确认、未命中问题与人工结果反馈	问题清单、标准答案、运营反馈	持续维护
业务规则维护者	确认收费、入住、停车、设施与异常处理口径	业务规则、事实来源、变更说明	持续维护
AI 工作流实施	Coze 节点搭建、数据库与知识配置、异常链路联调	工作流、编号数据、知识配置	已投放
测试与质量	执行离线评测、四场景联调、Bad Case 归因与回归	评测记录、场景验收记录、回归清单	持续执行

13.3 交付物

交付物	用途	交付状态
产品需求文档	统一目标、规则、功能、评测与验收	本 PDF
核心流程图	说明系统主路径与异常出口	已交付
协作泳道图	说明客人、系统、人工与维护角色的职责	已交付
23 条编号记录	现有 CSV 中的固定答案与人工兼容码	已完成字段说明与安全规则
兜底知识库	未覆盖低风险问题的检索依据	已形成知识内容与治理规则
四场景验收记录	记录押金、停车、投屏和发票处理路径	已完成联调并留存录屏
评测与验收方案	检查路由、事实和高风险人工处理	已形成指标口径与验收规则

附录 A 术语表

术语	定义
受控编号	分类节点允许输出的业务场景标识，用于连接标准答案或特殊处理状态
标准答案	经确认并可版本化维护的固定业务事实与客服话术
事实受控润色	只调整表达，不改变金额、时间、条件、否定和风险边界的生成方式
未匹配知识兜底	固定分类未覆盖时，基于有效知识证据回答低风险问题的路径
人工处理	系统停止自动承诺，将问题交给人工判断或受控业务系统的路径
Bad Case	路由、事实、安全、表达或异常处理不符合预期的评测或运行样本
回归评测	版本变化后使用固定样本重新检查结果，确认没有引入退化

附录 B 交付物索引

- Coze 工作流配置：包含分类、选择器、编号查库、润色、知识检索和人工处理分支。
- 核心流程图与协作泳道图：记录系统主路径、人机边界和维护协作关系。
- 四场景验收记录：记录押金、停车、投屏和发票的代表性处理路径，包含联调录屏。
- 编号数据库：当前维护 A001-A022 与 A999 共 23 条记录，并配套字段结构与安全规则。
- 产品需求文档：统一记录产品目标、功能需求、数据、Prompt、异常、评测与验收设计。

V1.0 交付结论：已完成受控路由、标准答案查询、知识兜底和人工处理规则定义，完成 Coze 工作流投放及 30 条离线评测。后续业务规则、知识内容、Prompt、模型和流程变更按本文版本机制执行。

产品负责人：曾子恒 | 文档版本：V1.0 | 文档状态：正式版