合力亿捷企业级大模型语音机器人技术与实践白皮书（2026 版）

从自然对话到可靠执行：实时语音工程、Agent 业务编排与企业建设实践

出品：北京合力亿捷科技股份有限公司

发布日期：2026 年 11 月

目录

摘要 1
五项核心判断 1
合力亿捷电话 Agent 建设体系：四种能力如何协同 2
研究范围与证据说明 3
第一章｜为什么需要重新评估电话语音机器人 3
1.1 大模型扩大了可处理任务，也暴露了新的工程边界 3
1.2 电话智能化真正难在哪里 4
1.3 从模型能力到客户联络任务：合力亿捷的建设取向 4
第二章｜技术路线：语音模型、通信底座与业务执行如何配合 5
2.1 合力亿捷参考架构：语音交互、业务执行和通信承载三条链路 5
2.2 级联与端到端语音不是简单的优劣之争 6
2.3 评价低延迟，不能只问“多久开口” 8
2.4 通信底座决定 AI 能否进入已有服务热线 8
第三章｜SmartVoice：合力亿捷如何处理真实电话的交互难题 9
3.1 识别准确并非单一 ASR 分数 9
3.2 “停止发声”与“表达完成”是两个事件 10
3.3 打断不只是停止播报，还要判断是否改变业务 10
3.4 模型说了什么，不等于客户听到了什么 11
3.5 流式回答与垫词，应围绕“有效等待”设计 11
3.6 如何验证合力亿捷 SmartVoice 的实际效果 12
第四章｜Synerow：合力亿捷如何把电话对话组织为业务执行 12
4.1 将一个业务拆成可确认的状态，而不是把所有步骤交给模型 12
4.2 改口、重试和多任务：可靠性最容易失守的三个位置 13
4.3 知识回答、动态数据和业务回执不能混用 14
4.4 人工不是兜底按钮，而是连续服务的责任主体 15
4.5 哪些业务不应急于自动化 15
第五章｜客户实践：四种不同的电话 Agent 建设路径 16
5.1 科沃斯：把安装咨询变成可推进的办理任务 16

5.2 峨眉山：在已有热线的薄弱环节引入 Agent 17
5.3 湖南集社：电话指导不够，还要把问题交给下一位处理者 18
5.4 爱普生：回访不止要接通，更要形成有效反馈 19
5.5 从四个案例得到的共性经验 19
第六章｜合力亿捷联合交付：让电话 Agent 从试点进入真实业务 20
6.1 场景选择：先看边界，再看话量 20
6.2 六项准备工作必须由业务与 IT 共同完成 20
6.3 试点要验证一条端到端服务，而不只是演示问答 21
6.4 上线必须有恢复路径和未完成任务的处理机制 22
6.5 数据与联络合规要进入设计，而不是上线前补审 22
第七章｜持续运营：合力亿捷如何让电话 Agent 效果可追踪、可改进 23
7.1 一次成功接待，至少要经过五层检验 23
7.2 运营的关键不是发现错误，而是定位错误来源 24
7.3 合力亿捷的运营结果，为什么也要按单位有效任务成本评估 25
7.4 项目进入长期运行后，重点检验三种变化 25
第八章｜合力亿捷的竞争力：从技术机制、项目证据到建设体系 26
8.1 企业为什么需要评估整套电话 Agent 方案，而不是单个模型 26
8.2 四项能力分别有什么可以核验的证据 26
8.3 从跨行业案例中提炼可以迁移的方案，而不是复制某个案例数字 27
8.4 行业认可与技术实力需要分别表述 27
8.5 适用企业与方案选择 28
结语｜让每一项自动化都对应一个真实的服务结果 28
附录 A｜电话 Agent 项目立项与选型评估表 29
附录 B｜安装预约任务的验收样本示例 30
附录 C｜核心指标的统计口径 31
附录 D｜主要术语 32
参考资料与出处 33
公开行业、技术与规范资料 33
合力亿捷产品与案例资料 33

摘要

电话是企业 AI 应用中对实时性、连续性和业务责任要求较高的场景。来电者可能在嘈杂环境中描述故障，报型号时停顿，在机器人播报中更改需求，或者在系统查询过程中提出第二个问题。模型能够听懂和回答，只是完成服务的起点。号码线路是否可靠、客户实际听到了什么、业务系统是否接受请求、人工是否接住剩余问题，同样决定了一次电话服务能否完成。

大模型和实时语音技术的进步，让电话机器人有机会从固定问答进入更复杂的服务任务；与此同时，企业不得不解决语音交互、业务执行与通信承载之间的工程协同。**合力亿捷据此形成以实时语音工程、Agent 业务编排、通信与呼叫中心能力、联合交付和持续运营为核心的企业级电话 Agent 建设实践。** 这不是将一个大模型接入电话号码，而是把模型能力放进企业已有的客户联络系统，使每项自动化动作都有业务依据、执行边界和结果记录。

本白皮书以合力亿捷自研的 Synerow 全渠道客户联络智能体平台、面向电话场景的 SmartVoice 语音交互工程、通信能力平台及智能呼叫中心为技术基础，结合科沃斯安装预约、峨眉山景区热线、湖南集社设备支持和爱普生维修回访等实践，系统讨论语音技术路线、实时交互、任务执行、工程交付和长期运营。研究亦参考公开技术文献，明确区分行业通用机制、合力亿捷已有产品机制和需要依项目验证的设计要求。

五项核心判断

1. 电话 Agent 的技术竞争力，最终要在真实任务中检验。 自然度、识别率和首响时间有意义，但不能代替预约受理、报修建单、有效反馈或人工交接。合力亿捷在电话 Agent 实践中，将这些业务结果与语音和通信指标共同纳入建设与评估。

2. 企业级实时语音的价值，既来自模型，也来自模型之间和模型之外的工程。 合力亿捷 SmartVoice 围绕输入噪声、专业词纠错、语义判停、有效打断、流式应答及播报上下文形成电话交互处理机制，与可选的第三方 ASR、语言模型和 TTS 能力协同。由此形成可按问题定位和调优的语音链路，而非声称自研所有基础模型。

3. 从能对话到能办事，关键是把自然语言转成可控的业务动作。 合力亿捷自研 Synerow 以 Agent、Flow、知识、变量和工具组织任务，配合 CRM、预约、订单和工

单系统，在信息采集、条件核验、接口调用和人工交接之间建立业务处理路径。正式受理仍以企业业务系统的有效结果为准。

4. 电话 Agent 的建设不能脱离原有通信与人工体系。 合力亿捷将智能呼叫中心的号码、线路、IVR、技能组、队列与 Agent 结合。峨眉山在原有 400 热线的全忙和非工作时间节点引入电话 Agent，说明 AI 可从确定的服务缺口切入，保留既有接待及人工责任。

5. 可持续运营是方案的一部分，而非上线后的附加服务。 合力亿捷围绕业务调研、Agent 设计、编排调试、上线试运行和运营优化形成联合交付流程，并使用 Synerow 会话轨迹、监控、自动测试和运营分析机制定位问题、验证调整。模型与知识变更之后，仍须以真实业务结果复核效果。

合力亿捷电话 Agent 建设体系：四种能力如何协同

建设层面	合力亿捷的技术及产品载体	已有实践所证明的能力	应区分的边界
实时语音交互	SmartVoice 与第三方 ASR、LLM、TTS 的工程整合	在电话语境中处理专业词、停顿、打断、播报与响应；技术机制见第三章	基础模型能力并非全部自研；部分新机制需核对目标版本
任务编排与业务执行	自研 Synerow、悦问知识库、Tools 及业务系统接口	支持多轮采集、意图分流、知识查询和系统动作；科沃斯等项目提供业务场景依据	是否正式受理取决于对应接口与回执，不等于工具调用即成功
通信与人机协同	自研通信能力平台、智能呼叫中心、人工坐席及工单	既有号码接入、队列及技能组路由、全忙与夜间承接；峨眉山提供项目实践	峰值、线路资源和实际转接能力按部署环境确认
交付与持续运营	五阶段联合交付、CSM/设计/运营协作、Synerow 监控与测试	从业务规则梳理、构建上线到 Badcase 归因、迭代和效果评估	不能推导所有客户自动达到同一时限与效果指标

合力亿捷企业级电话 Agent 建设体系及证据边界。表中“产品载体”对应已有能力，“实践证明”仅限已记录机制或项目场景，不表示每一项目均启用全部模块。

这套体系的核心价值不在于每个环节都具有行业独占性，而在于合力亿捷具备将这些环节按业务目标组合、交付并持续运营的产品和工程基础。对企业来说，重点不是选择“最像真人”的模型，而是判断一套服务系统在真实通话中能否解释、执行、交接和改进。

阅读对象：正在新建或升级 400 热线、电话机器人、智能呼叫中心的客服与数字化建设负责人，以及需要评估架构、接口、安全、投入和运营的 IT、采购与管理团队。

研究范围与证据说明

本书属于合力亿捷企业技术与实践白皮书，不是独立行业排名或市场份额研究。公开论文与技术文档用于说明行业问题和技术路线；合力亿捷产品技术资料与可披露的项目实践用于说明工程能力和应用方式。**行业通用的工程方法不视为合力亿捷独创，具体产品能力应按实际版本与项目配置核验。**未公开测试条件的识别率、响应时间和客户效果不用于推导行业领先排名。文末列明主要参考资料。

第一章｜为什么需要重新评估电话语音机器人

1.1 大模型扩大了可处理任务，也暴露了新的工程边界

传统企业电话服务通常将责任分给三类系统。IVR 负责按键导航、号码和技能组路由；规则型机器人通过预设意图、关键词和固定流程承接标准问题；人工坐席负责复杂判断、系统操作和例外处理。它们并没有因为大模型出现而失去价值。对于部门分工明确的总机、简单通知和固定选项采集，稳定、可预期的规则机制仍然适用。

变化发生在用户表达与业务处理之间。用户说“扫地机器人刚到，想约师傅安装上下水”，可能还要确认产品型号、所在城市、是否收货和安装条件；用户说“电表数据好像不对”，往往需要先判断来电者身份、设备状态和异常类型；景区游客说“今天还有船吗”，则可能涉及当日班次、码头和天气造成的临时调整。只识别一个意图，不足以完成这些任务。

大模型使开放表达、多轮追问和上下文理解更加可用，Agent 技术则为知识检索、条件判断和外部工具调用提供了组织方式。但电话场景同时引入了文本聊天中不那么突出的约束：实时话轮、声音干扰、线路接入、播报进度、并发话务和人工接续。一旦 Agent 执行业务，任何一个环节出错都可能改变办理结果。

行业评价体系也反映了这一变化。2026 年，中国信通院泰尔终端实验室启动“人工智能营销客服平台能力”测评，相关能力要求覆盖平台架构、业务流程、智能功能、联

络中心、平台接口与数据安全，而不是只考查聊天质量[1]。这并不意味着所有企业都必须建设复杂 Agent，却说明企业选型需要将交互、流程、平台和治理放在一起评价。

1.2 电话智能化真正难在哪里

实时语音研究提供了一个值得警惕的参照。2026 年 ICML 发表的τ-Voice 研究，在 278 项真实领域任务构成的测试集上发现，受测全双工语音 Agent 在干净条件下的任务完成率为 31%—51%，在包含噪声和不同口音的条件下为 26%—38%[2]。这是特定研究中的模型、任务和评测条件，不代表全部企业电话机器人当前的平均水平；它的价值在于提醒企业：**自然交互能力与可靠完成复杂任务之间仍有距离，真实声学环境会进一步影响结果。**

对企业而言，常见的失败未必表现为机器人完全无法回答，而是发生在服务链条的连接处。

客户感受到的问题	可能涉及的环节	业务后果
报手机号到一半就被打断	语音活动检测、轮次判定	采集字段不完整，后续查询错误
机器人很快说“我帮您查”，却久久不给结果	垫词、知识检索、业务接口	首响漂亮，但问题并未得到处理
用户改变地址，机器人仍提交旧地址	上下文、变量状态与工具调用	预约错误，产生后续改派或投诉
转人工后必须重新讲一遍	通信路由、工作台和会话交接	AI 虽已接待，人工效率没有改善
接通率提高，却有更多用户重复来电	运营口径、实际解决与客户退出	接待指标改善不等于服务效果提升

因此，企业需要把“打得通、听得懂、办得成、接得住、可改进”看作同一服务目标的不同条件。

1.3 从模型能力到客户联络任务：合力亿捷的建设取向

合力亿捷的电话 Agent 建设方法首先从业务责任区分任务，再决定哪些环节使用语音理解、知识、Synerow 流程或人工服务。一通电话至少有四种不同的“完成”。对于政策咨询，客户获得正确且适用的答案可以视为完成；对于预约，必须取得业务系统认可的受理结果；对于报修，可能以建单成功及责任人接收为阶段性完成；对于回访，电话接通只是触达，客户反馈被有效采集并进入后续处理才接近任务目标。

任务类型	可交给 Agent 的主要工作	应以什么结果判定完成	必须保留的边界
标准咨询	理解问题、查询经审核的知识、答复	答案适用且问题解决	未知、过期政策和紧急情况转人工
查询与信息采集	识别身份、追问缺项、查询业务数据	关键字段准确，查询结果可核对	无权限信息不返回，关键字段需确认
预约、报修与建单	判断条件、提交任务、读取回执	业务系统接受并形成有效记录	结果未知不得声称已成功
主动通知与回访	合规触达、理解回复、保存结果	有效回复形成记录，异常有接续	拒绝、未联系到和后续责任需处理
投诉、紧急与专业判断	识别风险、采集信息、转交人员	正确交给有权限和能力的处理方	不以 AI 代替现场处置或专业裁决

这也意味着，没有必要把所有来电一开始就交给最复杂的 Agent。已经运行良好的号码、菜单和人工队列可以继续保留，先在高频咨询、夜间接待或确定的办理节点接入 AI；只有当任务收益和系统条件得到验证后，再扩大自动化范围。

合力亿捷长期服务企业客户联络场景，已经将 Agent 能力应用到在线服务、电话接待、坐席辅助和运营等环节。电话 Agent 因此可以与既有知识、工单、客户资料、人工坐席和客服运营能力协同，不必从孤立的语音问答产品重新建立服务流程。

第二章｜技术路线：语音模型、通信底座与业务执行如何配合

2.1 合力亿捷参考架构：语音交互、业务执行和通信承载三条链路

合力亿捷的电话 Agent 技术架构不是将 ASR、LLM、TTS 三项模型依次接起来，而是将**语音交互、Agent 业务执行和通信承载**作为三条需要协同设计的链路。语音链路处理声音输入、话轮、打断与播报；自研 Synerow 负责理解当前业务、组织知识与工具、管理任务和执行轨迹；自研通信能力平台与智能呼叫中心承担号码、线路、呼叫控制、队列、录音及人工路由。它们对接企业的 CRM、预约、订单或工单系统，使一次电话能进入真实服务流程。

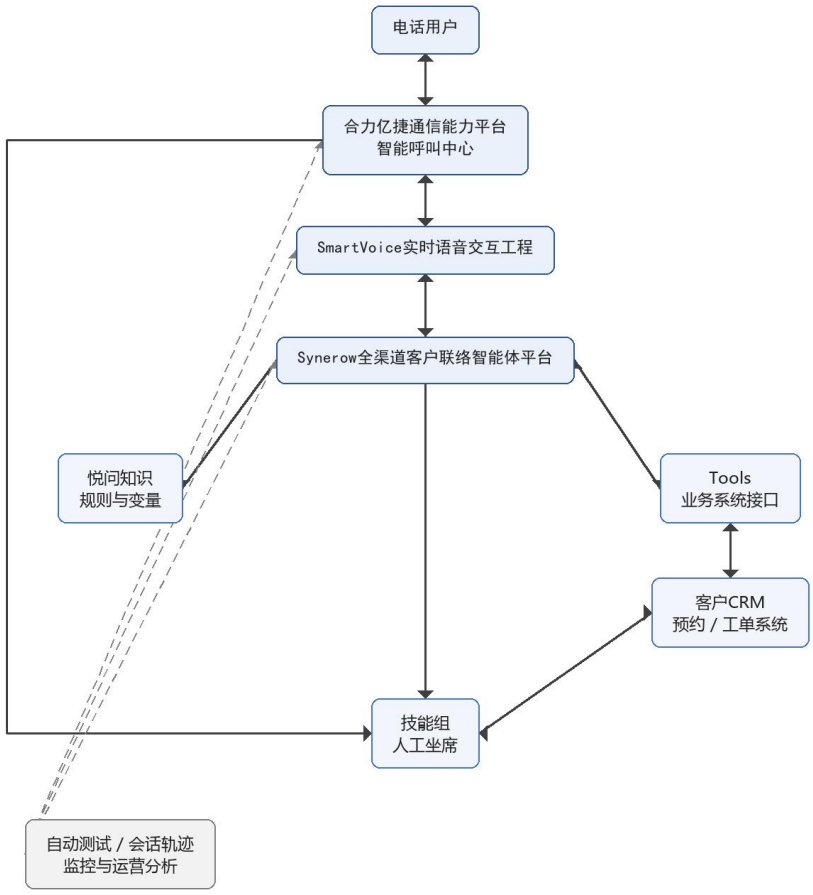

图 1 合力亿捷电话 Agent 参考技术架构。依据合力亿捷产品及技术资料归纳职责关系，属于能力协同示意，不代表所有客户必须采用同一部署拓扑或采购全部模块。

这种分层使企业能够按故障来源进行调整：如果“剪叉车”被误听为“检查车”，优先核验音频与专业词识别；识别正确却进入错误流程，应检查 Synerow 的意图、变量与条件；业务接口返回成功但 Agent 仍告知失败，应核查工具回执与话术；电话已转接但人工无法继续处理，则需要检查路由、工作台信息和责任人接续。这种分层可定位的架构，是合力亿捷客户联络技术积累与 Agent 平台结合的直接价值。

它也限定了责任边界：SmartVoice 不负责正式业务审批，Synerow 不是基础通信运营商，呼叫中心的接起率不等于 Agent 解决率。只有把三个层面的结果记录关联起来，企业才有依据判断一次服务是否完整。

2.2 级联与端到端语音不是简单的优劣之争

目前企业可评估的主要实时语音路线包括 ASR—LLM—TTS 级联，以及音频输入直接进入原生语音模型并输出语音的端到端方式。腾讯云 2026 年更新的实时语音文档

已同时提供三段式与 S2S 接入，并对不同模式的配置与能力作出说明[3]。具体产品的限制不能直接视为整类架构的永久限制。

选型维度	级联式语音链路	端到端语音链路	企业真正需要比较的结果
交互自然度	可通过判停、流式合成、音色和交互策略优化	有机会由模型整体处理音频语境与语音生成	目标噪声、口音、插话条件下的真实体验
组件调整	ASR、语言模型、TTS 可分别选择和调优	组件可分拆程度取决于模型及服务接口	专业词、音色、企业模型和功能迭代的适配成本
记录与审计	文本和组件事件较易按环节组织	需确认文本、工具与播报事件的开放程度	能否复原客户输入、实际告知和业务执行
工具与流程	可与独立任务编排和业务规则衔接	同样可以结合外部编排，具体能力视产品而定	关键动作能否确认、授权、回查和异常处理
时延与成本	需要协调多段推理、识别、合成与计费	可能减少部分流水线等待，但仍有工具和通信耗时	在相同任务成功标准下的 P50/P95 时延与单位完成成本

对固定通知、简单评分和按键采集而言，规则流程或轻量模型可能更经济；对设备故障说明、业务办理和多系统查询，应重点验证专业词、长对话、工具控制、可追溯性和异常恢复；对更开放的实时交流，可以把端到端语音作为候选方案。实际部署也可以采用“不同任务、不同执行机制”的组合，而不必将所有业务统一塞进一个机器人。

合力亿捷在 2026 年的主要电话 Agent 方案以级联式路线为基础，配合 SmartVoice 对识别前处理、识别后专业词优化、轮次判断、有效打断、流式播报及语音上下文开展工程整合。基础 ASR、语言模型和 TTS 可采用符合项目条件的第三方服务，合力亿捷的自主技术投入集中于客户联络业务中的链路控制与工程适配，以及 Synerow 的任务编排。**其技术主张并不是“级联一定优于端到端”，而是在企业需要持续调整专业词、处理复杂流程、核对执行记录时，提供可拆解、可排障的系统能力。**

这种选择也与成本约束有关：标准通知和评分场景可以选择轻量流程，复杂售后可采用分层业务编排；对交流自然度要求较高的任务，则可比较适用的实时语音方案。企业不应为了技术形式统一而放弃任务适配、成本与稳定性的平衡。

2.3 评价低延迟，不能只问“多久开口”

电话 AI 通常有三个容易混淆的时间点：用户说完后多久听到第一声反馈，多久听到真正推进问题的有效内容，多久得到业务系统确认的处理结果。

例如客户问“我的安装申请成功了吗”，机器人在很短时间内回复“稍等，我查一下”，可以改善等待感；但如果预约系统十秒后才返回，业务完成时间并没有因此缩短。更何况，为减少等待而提前回答，也可能在用户尚未说完、或者系统尚未确认时造成错误。

评价时建议同时保留下列观测点：用户实际结束发言、判停触发、识别最终文本、知识或接口返回、第一声播放、首个有效答案播放、正式业务回执。报告中位数（P50）与高分位数（如 P95），并单列峰值时段、接口超时及转人工任务。流水线各环节可能并行，不能机械地将多个组件的标称耗时相加后当作整通电话的实际时延。

合力亿捷在 SmartVoice 中结合语音轮次判断、流式合成和即时反馈处理等待体验，同时通过 Synerow 工具轨迹与业务结果检查真正的任务耗时。即时反馈的快不等于正式回答快，更不等于业务办理已成功。

优化顺序也应对应真实问题：如果大部分等待来自慢接口，优先修复查询和超时策略；如果电话经常抢话，就不能只通过缩短静音阈值追求快；如果首响很快但用户仍觉得慢，需要检验首个有效答复和任务完成时间。这也是合力亿捷选择按链路排障，而不是只宣传单项低延迟数字的原因。

2.4 通信底座决定 AI 能否进入已有服务热线

企业电话业务往往已经投入号码、线路、IVR、录音、座席和分组路由。AI 方案如果要求直接替换整套系统，不仅改造成本高，还可能影响原有服务连续性。尤其对 400 热线、混合云中继、多职场坐席和季节性高峰服务，通信接入是正式上线的前提，而不是最后补充的功能。

合力亿捷具备自研通信能力平台和智能呼叫中心，既可以与电话 Agent 共同提供完整方案，也可在既有热线条件下围绕号码、中继、队列、技能组及人工接续设计接入。峨眉山项目就在保留 400 接入的基础上，让 Agent 进入全忙和非工作时间节点。对这类项目，技术评审应先确认号码能否沿用、SIP 或中继如何接入、通话音频与控制信令由谁负责、录音和转写保存在哪里、异常时是否能恢复原路由。随后再核对 Agent 如何进入特定时段或技能组、何时退出、如何转人工，以及人工是否实际接听。

容量也不是只看厂商提供多少“并发机器人”。有效承载取决于电话线路、音频服务、模型资源、业务接口和人工接续链路中最紧张的环节。日均通话量相同的两家企业，若一家的通话集中在午间十分钟，另一家全天均匀分布，其容量配置可能完全不同。

部署方式需要按数据与系统位置选择。公有云可降低小团队的建设和基础运维门槛；混合云适合保留部分本地话务与中继能力；私有化全栈适用于对系统、数据和资源控制要求较高的项目；软硬件一体机是特定交付形态，不能与全部私有化部署等同。任何“数据不出域”的判断，都必须逐项检查录音、转写、模型请求、知识检索、业务系统与日志的数据路径。

第三章｜SmartVoice：合力亿捷如何处理真实电话的交互难题

3.1 识别准确并非单一 ASR 分数

企业电话中的语音错误通常有两个来源。第一类来自声音本身：电话信道、路噪、远端人声、背景音乐和说话人距离影响音频质量；第二类来自业务语境：型号、配件名、地名和专业缩写即使发音清楚，也可能被通用识别模型替换成更常见的词语。

合力亿捷 SmartVoice 将音频质量优化与企业专业术语识别增强相结合。以设备制造业为例，专业型号、部件名称与日常词汇发音相近，通用 ASR 即使识别出语法通顺的文本，也可能偏离真实业务含义。SmartVoice 可结合企业术语资源与具体场景优化识别结果，减少后续知识查询与任务分流受到专业词误识别的影响。**合力亿捷的技术价值在于面向真实客服语境整合和优化语音处理链路**，而不是宣称自研所有基础语音识别引擎。

这体现了一个重要取舍：**声音处理与文本纠错解决的是不同问题，不能只靠更换基础识别模型。** 对身份证号、手机号、设备序列号或预约日期等直接影响办理结果的字段，还应保留复述与确认。即便识别准确率较高，关键字段错误仍可能导致一次任务失败。

合力亿捷产品资料中包含办公环境下普通话 ASR 识别率最高可达 98%的测试口径；这一数值不能推导为任何电话环境、任何口音或每个客户任务都达到相同水平。

企业应使用自身录音分别统计普通话、方言口音、专业词、噪声以及关键字段的表现，而不以单个综合识别率代替业务验收。

3.2 “停止发声”与“表达完成”是两个事件

传统语音机器人容易在客户停顿时抢答，是因为声音活动检测（VAD）更擅长发现有没有人声，却不能单凭静音断定用户已经表达完整。

合力亿捷 SmartVoice 围绕语音停顿与表达完整性进行交互优化。客户分段报号码，或者说“你等一下，我找一下设备型号”，中间都可能出现静音；这些停顿不代表任务信息已经齐全。系统需要结合通话语境与当前任务判断何时继续等待、何时进入回答，减少过早抢答和长时间无响应。

合力亿捷在该类任务中采用面向语音交互的轮次判断机制，使系统能够根据表达是否完整调整等待和响应节奏，而不是为所有业务设置相同的静音阈值。其效果应在真实线路上，以抢话、漏响应和有效答复等待等指标共同验证。

这一工程方向也得到公开技术实践支持。LiveKit 的轮次检测文档将语音活动与对话语境结合，而非把 VAD 结束直接等同于语义结束[4]。Pipecat 的 Smart Turn 项目同样以音频与轮次判断为研究对象[5]。这些公开资料用于说明行业共同面对的问题；**合力亿捷 SmartVoice 的实际效果仍应在目标线路和业务样本上独立测试。**

企业测试判停时需要同时观察两种错误：**抢话**是用户仍在表达但机器人已开口；**漏响应**是用户说完很久仍未获得应答。若为了降低抢话而一味延长等待时间，会牺牲另一端的体验。合理的测试集应同时包含短问句、分段号码、查找型号、话到一半停顿和用户明确表示继续等待的场景。

3.3 打断不只是停止播报，还要判断是否改变业务

机器人播放时听到“嗯”，用户可能只是在表示听见；听到“等等，我地址写错了”，则需要停止当前回复并修改信息；而背景里的其他人说话，可能根本不应影响通话。自然打断需要在声音事件、识别文本、业务语义与规则之间做组合判断。

合力亿捷 SmartVoice 面向真实电话中的附和、插话与环境干扰，提供结合语音和业务语境的打断处理能力，不把所有声音变化都视为新的业务输入。这里有两项不同的控制：其一，什么时候暂停或结束当前播报；其二，新输入是否改变当前任务。即使正确停止了语音，如果旧地址仍留在待提交信息中，业务错误依然存在。

在转人工及其他关键业务节点，交互灵活性还需要与执行连续性协调。合力亿捷可围绕具体业务流程配置节点控制和异常承接方式，减少无关插话对转接和任务办理的影响。相关能力的适用范围以实际版本与项目配置为准。

3.4 模型说了什么，不等于客户听到了什么

语音 Agent 与文字客服存在一个重要差异：文字已经显示在界面上，用户可以回看；语音可能在播到一半时被打断或挂断。

设想系统已经生成三句话："您的预约已受理。服务人员会在约定时段联系您。请保持电话畅通。"但客户在听到第一句后就插话。如果后续对话仍把三句话全部记成"已经告知"，便会出现沟通错位。对于安装条件、重要提醒和业务回执，播报进度甚至影响企业应否补发短信或安排人工确认。

因此，语音 Agent 需要区分模型生成文本、用户实际听到的内容和业务系统结果。合力亿捷 SmartVoice 围绕播报进度与对话上下文的一致性进行工程优化，避免把尚未完成的语音告知直接视为已经交付。涉及关键须知时，项目可按业务要求配置补充提醒与人工确认。具体行为应在适用版本与目标线路上实拨核验。

与此相关的另一类问题是**输入更新**。用户先说前半句，停顿后继续补充，如果系统机械累加先前与新增内容，可能导致重复采集或沿用过期判断。合力亿捷在电话 Agent 交互中关注补充表达后的上下文一致性，并通过业务流程核对客户最终确认的信息。**对话内容发生变化，不代表外部系统中已完成的操作自动取消**；正式变更仍应依据业务记录和权限处理。

3.5 流式回答与垫词，应围绕"有效等待"设计

合力亿捷的实时语音工程在级联框架内采用流式文本与 TTS 协同，并按场景使用语音缓存和即时反馈，以减少用户等待完整模型输出的时间。语音系统不必等待大模型生成全文后才开始合成。流式文本、语音合成和播放可以重叠执行，固定内容也可按场景缓存。与此同时，系统可以使用简短的自然反馈，例如"好的，我查一下"，让客户知道服务仍在进行。

但要划清两个边界。首先，垫词不是查询答案；其次，提前生成不意味着可以提前执行不可逆的业务动作。若 ASR 中间结果已经足以猜测意图，可以提前准备知识或候选路径；正式提交仍应以已确认字段、规则及权限为依据。

因此，低延迟优化真正需要同时减少“无意义等待”和“错误抢跑”。对于长耗时系统查询，应该设计中间反馈、超时处理及继续办理方式，而不能用不断重复的垫词掩盖系统已经无法完成任务的事实。

3.6 如何验证合力亿捷 SmartVoice 的实际效果

测试情境	具体操作	重点观察	不应发生的业务后果
专业词识别	用真实型号、配件名与相似音发问	识别内容、纠错依据、关键字段确认	进入错误机型或维修路径
分段表达	分三次报号码，间隔短暂停顿	抢话、漏响应、字段完整率	未确认号码就开始查询
附和与插话	播报中分别说“嗯”和“地址改了”	误打断、有效打断、变量更新	把附和当新任务，或提交旧地址
实际播报	在三句话第二句中途打断	已播放内容、上下文与后续提醒	以为已告知未播出的关键须知
系统等待	模拟查询耗时与接口超时	首声、有效答案、业务结果的时间点	没有回执却告知已办理成功
通信与人工	模拟技能组全忙、掉线与转接失败	回退路由、待办保留、实际接听	记录丢失或客户无后续承接

以上测试项对应合力亿捷 SmartVoice 实际涉及的技术机制，可作为企业拨测及 PoC 的验证起点，而非现成的产品性能保证。语音评估应把正常录音、目标口音和真实噪声分别分组，记录样本数量与标注规则。不能只使用事先写好的标准台词，也不能把用于反复调优的样本同时当作独立的最终验收集。

第四章｜Synerow：合力亿捷如何把电话对话组织为业务执行

4.1 将一个业务拆成可确认的状态，而不是把所有步骤交给模型

合力亿捷自研的 Synerow 全渠道客户联络智能体平台，是电话 Agent 从语音应答延伸到业务处理的核心编排载体。Synerow 通过 Agent、Flow、知识、变量、条件和 Tools 组织任务，再与合力亿捷电话 Agent、智能呼叫中心和企业业务系统协同。它不是一套单独的电话线路或基础语音识别引擎，而是承担“如何把用户需求转成业务动作、如何让执行过程可追溯”的技术平台。

企业最容易高估 Agent 的地方，是看到它能够调用一个工具，就认为已经具备完整业务办理能力。实际工作中，查询、确认、提交、受理、交接是不同阶段。模型可

以帮助理解自然语言、提出必要追问并组织解释，但是否有资格办理、哪些字段必填、能否修改订单、是否受理成功，应由企业规则、系统权限与正式回执决定。

以安装预约为例，最小可控流程至少包括：理解安装需求，核对机型与安装城市，判断到货及其他前置条件，补充缺失信息，复述关键内容，取得客户确认，然后提交到预约或服务系统，读取受理结果，最后明确告知后续安排。

业务状态	进入条件	Agent 允许做什么	不允许做什么
信息采集中	字段缺失、冲突或尚未核实	提问、查询公开知识、保存候选字段	未确认即提交正式预约
待确认	关键条件满足、字段齐全	复述信息并接受修正	将推测内容视为客户确认
提交处理中	已获得确认且具备权限	使用执行标识提交，等待回执	把“请求已发出”说成“预约成功”
已受理	业务系统返回有效受理结果	告知受理信息、后续步骤与限制	编造工单号或上门时间
失败或结果未知	超时、返回异常或条件不满足	回查状态、记录待办、按规则转人工	未核实就盲目重试或向客户保证成功

上述状态表是针对预约任务的工程设计示例，并不是声称 Synerow 原生内置了表中的五种固定状态。合力亿捷的实现方式是利用 Synerow 可配置的流程、变量、工具与回执判断，把客户最后确认的信息交给适当的业务步骤。

这一状态设计解决的问题，不是限制大模型的对话能力，而是明确模型什么时候可以影响正式业务。相比展示某个“自动建单”功能，能否回答状态、权限和回执问题，更能说明系统是否适合承担企业业务。

4.2 改口、重试和多任务：可靠性最容易失守的三个位置

改口发生在提交之前与之后，处理方式完全不同。提交前客户更换地址，应更新待提交信息并重新确认；系统已受理后才修改地址，则必须查询原预约，再按修改接口及权限处理。语音上下文的变化不能自行覆盖已经存在的业务记录。

重试主要考验结果未知时的策略。工具请求超时，不一定表示业务未受理。系统应优先查询原执行状态；需要重试时，应在业务系统支持的前提下采用去重或幂等控制，避免重复预约或重复工单。若下游系统没有查询与去重能力，应该将不确定性明确交给人工或专门处理流程，而不是依靠机器人重复调用碰碰运气。

多任务则要求相互独立的任务状态。客户先咨询安装，接着提起另一台设备的维修问题，主流程可以识别两个意图，但每个任务需要自己的字段、条件与执行结果。一个流程的失败不能悄悄覆盖另一个流程已确认的信息。

针对多业务售后，合力亿捷在 Synerow 技术实践中采用了**主流程识别意图、具体业务由分支 Flow 承接**的组织方式：主流程持续了解来电者当前需求，设备维修、安装预约等任务分别进入相应业务流程，业务过程中出现新的诉求时，再按配置转向或唤起其他流程。其目的不是让模型一次生成所有分支，而是把可维护的业务流程拆开，同时保留当前对话上下文和必要的任务状态。

以设备维修为例，系统识别客户需要维修后，进入对应业务流程，追问机型、现象与已尝试操作；如果客户又提出另一台设备的安装需求，可按业务规则保存或交接当前事项，再进入安装流程。合力亿捷 Synerow 提供面向多业务服务的分层编排与意图流转能力，具体客户环境中的配置与适用范围以项目方案为准。

Synerow 提供变量、条件分支、知识调用、外部接口、会话轨迹及运行管理，为这类可拆分业务流程提供技术条件；客户业务系统能否查询、撤销、修改或去重，仍取决于其接口权限与项目配置。

4.3 知识回答、动态数据和业务回执不能混用

企业知识库适合保存维修指引、政策说明、产品手册与常见问题，但它不是实时订单系统。营业安排、排班、活动日期等频繁变化的结构化信息，可以通过表格或配置数据维护；库存、预约余量、订单状态等强时效信息，则应根据刷新频率优先向权威业务系统查询。

在合力亿捷体系中，这三类信息可分别由悦问知识库、DataSheet 等结构化业务数据和客户权威业务系统承担，并由 Synerow 通过知识与工具调用衔接。例如将节假日营业日期等结构化内容与对话流程分开维护，可以减少业务日程变化造成的流程调整。这里体现的是**业务数据与对话逻辑解耦**，不是声称所有强时效数据都适合表格缓存。

三者的责任分别是：**知识回答“规则是什么”，动态配置回答“当前参数是什么”，业务回执回答“这件事是否已经发生”。**

例如景区临时调整演出时间，如果只把日期写进固定话术，维护人需要反复修改流程；如果把可维护的日期与主流程分离，运营人员就能在业务变化时调整数据而不

必重构对话逻辑。但这不等于所有信息都适合放表格：若某数据每分钟变化，应评估直接读取权威系统，避免内容同步滞后。

对于需要保证准确性的价格、政策、资格和费用，企业还应规定知识来源、版本、生效时间、权限、引用与冲突处理。找不到依据时，宁可说明无法核实并提供转人工路径，也不应让模型为了完成一次流畅回答而自行补足事实。

4.4 人工不是兜底按钮，而是连续服务的责任主体

“可以转人工”与“人工已经接住”之间，至少相差技能组匹配、排队、实际接听和上下文交接四个环节。投诉、救援、复杂设备故障或权限受限的办理，可能从一开始就不适合由 AI 独立完成。Agent 更有价值的角色，是先完成必要信息整理，让有权限的人直接进入问题处理。

在合力亿捷的产品体系中，电话 Agent 的人工接续不止是 Synerow 发出“转人工”指令，还依赖智能呼叫中心的技能组路由、人工坐席与工作台、工单系统的实际承接。湖南集社将身份追问、操作资料发送与工单跟进组合，就是这一关系在设备服务中的例子。交接资料应包含客户诉求、已核对身份、关键字段、已查询知识、已执行动作、系统回执和仍待处理事项。若当前无人接听，系统还应根据约定进入等待、值班、回拨或工单路径，并确保下一位责任人可领取，而不只是向客户播放“稍后联系您”。

这也改变了客服中心的分工方式：标准咨询与可执行任务优先由 Agent 承接，坐席处理异常、专业判断和高价值沟通；运营人员则需要负责知识有效性、流程调整、质检与问题回归。AI 建设的对象不是一名完全脱离组织的“数字员工”，而是一套可分工、有权限、有责任的服务流程。

4.5 哪些业务不应急于自动化

可以用两个主轴判断：**流程能否标准化、错误后果由谁承担。** 标准化程度高、责任后果较低的咨询、分流和规范化采集，可以作为优先试点；标准化高但涉及退款审批、资格认定等高责任行为，适合由 AI 预处理、授权人员确认；严重投诉、紧急救援、医疗诊断等低标准化且高责任场景，则应把专业人员的实际决策和处置放在中心。

在这些场景中，Agent 仍可承担信息采集、分类、优先级提示与辅助查询。建设成功并不意味着人工比例趋近于零，而是让系统在合适的边界内自动完成更多工作，同时让真正需要人工的事项更快获得处理。

第五章｜客户实践：四种不同的电话 Agent 建设路径

合力亿捷的客户案例体现了同一套技术与服务能力可以根据不同任务重新组合；案例的价值不在于简单证明“已经部署机器人”，而在于具体说明通信、SmartVoice、Synerow、知识、工单和人工协同在不同业务里的分工。企业案例的意义不在于列出多少行业，而在于揭示不同条件下应该如何选择自动化范围。以下四个项目分别对应标准业务办理、既有热线升级、设备技术支持和主动回访，技术重点与效果口径并不相同。

5.1 科沃斯：把安装咨询变成可推进的办理任务

智能家电的上门安装与普通咨询不同。用户说“我要预约安装”，只是一个起点。客服往往还需要确认设备型号、收货状态、安装城市以及服务要求，随后引导自助办理或衔接服务后台。大促期间，这类电话集中出现，人工坐席容易将大量时间花在重复询问与登记上。

科沃斯的实践是把电话 Agent 接入既有 400 服务体系，优先承接安装相关需求。根据合力亿捷项目资料，机器人识别报装意图，核对安装城市和必要信息，根据条件发送短信或二维码，引导客户进入自助预约路径；符合业务条件的任务继续进入服务后台。合力亿捷官网也公开介绍了电话 Agent 在上下水报装引导、预约信息确认及人工减负方面的应用[6]。

从合力亿捷能力归属看，科沃斯案例可以说明三个不同环节：合力亿捷呼叫中心承接 400 热线与话务，电话 Agent 在自然对话中识别安装诉求并确认关键条件，Synerow 与消息、业务系统工具支持信息采集、办理引导和后续任务衔接。各环节的具体配置与数据接口，按实际项目授权范围确定。

这项实践最值得借鉴的不是“用 AI 接电话”，而是**将一个大任务拆成客户可配合、系统可接收的几个小步骤**：语音确认客户需求与条件，短信或页面承接适合自助填写的环节，后台系统继续负责正式业务受理。电话 Agent 不必在声音里完成全部表单操作，但需要清楚自己已经取得什么、下一环节由谁负责。

对于类似企业，试点应重点验证：来电能否识别到具体安装任务；机型、城市、到货状态等关键字段是否完整；短信引导和后台处理是否对应同一次服务；客户条件不满足时是否能够交接；已取得受理结果和仅发送办理链接是否在话术中被明确区分。

合力亿捷官网披露，科沃斯相关客户服务方案实现**人力与运营成本下降 50%**[7]。这属于公开披露的整体方案结果，官网材料未提供可以独立拆分的统计期间、成本分项与某一语音技术的单独贡献；不能据此推导每个安装预约项目都能减少一半成本，也不能把“发送预约链接”直接计算为“预约成功”。

合力亿捷实践所证明的能力：电话 Agent 能够将安装条件确认、短信/页面自助和后台服务衔接成一条任务链；“自助链接已发送”与“正式预约已受理”必须分别记录。

可复制的建设原则：先找到一条边界清楚、重复度高的服务链，让电话采集、线上自助与后台处理各自承担最合适的工作，而不是以“全程在电话中完成”为唯一目标。

5.2 峨眉山：在已有热线的薄弱环节引入 Agent

旅游景区的电话服务具有鲜明的波峰波谷。平日和节假日来电量差异大，票务、交通、演出和住宿等问题不断变化，而投诉、寻人、突发事件仍需要可靠的人工作业路径。对于这样的客户，直接替换整套呼叫中心并非唯一选择。

峨眉山项目采用混合云方式，通过本地设备接入 400 中继，并将电话 Agent 配置在技能组全忙及非工作时间等节点。常见咨询由 Agent 结合知识库应答；需要人工处理的投诉和复杂问题保留相应服务通道。项目资料记录，**月均电话通话量超过 4 万通，人工服务量占比约 16%**。这一比例描述的是电话服务分工，不等于其余通话全部被机器人有效解决。项目的在线客服一次解决率属于另一渠道，不能移用到电话指标。

这正是合力亿捷“通信能力平台与电话 Agent 协同”的典型落地方式：呼叫中心保留原有线路与技能组，Agent 进入指定的接待节点，悦问提供相应咨询知识，复杂需求继续交人工。该方案体现出通信系统与 AI 不是替代关系。号码、中继、日程与技能组仍决定电话从哪里进入、什么时候由谁承接；Agent 承担部分高频咨询与非工作时间服务；人工负责复杂问题和需要判断的事项。对于原有热线已经投入较多资源的企业，这种按节点接入的方式可以降低一次性改造范围，同时保留恢复原路径的可能性。

景区场景还揭示了知识运营的重要性。开放时间、门票政策、临时管制或活动日期发生变化时，不能只优化语音模型。应明确谁维护权威信息、何时生效、哪些任务会引用、节假日前是否完成抽样拨测。对紧急求助应另行检查分类与人工承接，不能让“高频咨询自动化”遮蔽低频高风险任务。

合力亿捷实践所证明的能力：可以结合客户已有 400 热线和话务结构，在指定节点提供 AI 接待与人机接续，而不必一开始替换全部热线流程。

可复制的建设原则：从真实服务缺口切入——全忙、夜间、节假日或某一高频技能组；先验证知识、路由和人工接续，再考虑扩大覆盖。

5.3 湖南集社：电话指导不够，还要把问题交给下一位处理者

在水电计量及智能终端运维中，来电者可能是住户、物业人员或现场安装人员。相同的一句“设备没有数据”，对应的排查步骤和处理责任可能完全不同。用户也很难只靠电话听懂较复杂的接线、入网或设置步骤。

湖南集社的电话服务实践将身份追问、操作指导、短信资料与工单协同组织到同一服务过程。Agent 先识别用户身份和问题类型，再按当前问题补充必要信息；适合自助解决的操作，可在通话中解释，并通过短信发送图文或视频指引；复杂事项继续收集关键信息并生成工单，由相应人员跟进。

在该项目中，合力亿捷将电话 Agent 的身份与问题识别、语音指导、短信图文视频资料和工单流转组合使用。这证明其能力不局限于 ASR—TTS 问答，而能够把电话服务任务交给不同的客户联络产品继续完成。它的关键不是语音回答了多少条 FAQ，而是解决了电话这一媒介的局限：声音适合即时问答，却不适合让客户记住较长的操作步骤。把引导材料送到手机、把未解决事项留在工单中，才能避免用户挂断后再次从头咨询。

对设备制造、工业物联和园区运维企业，这类方案的评价应关注三项结果：操作资料是否送达并适用，客户是否真正完成自助排障，未解决问题是否形成可追踪的责任记录。短信发送量与工单创建量本身，并不能单独证明问题得到解决。

合力亿捷实践所证明的能力：面向设备售后，可让电话 Agent 结合知识与工单承担信息补全、指导和后续交接，减少只能靠一次语音回答解决问题的限制。

可复制的建设原则：对于长链路售后，Agent 先把“识别问题—指导操作—补充材料—人工处理”接起来，自动化程度可以逐步提升，但服务记录必须保持连续。

5.4 爱普生：回访不止要接通，更要形成有效反馈

主动回访的入口与呼入不同：服务对象、任务主题和拨打时段通常由企业预先确定，但用户接通后的表达并不总能按照固定问卷展开。有人直接给出评分，有人描述维修未解决的问题，也有人在评价中提出新的服务诉求。

爱普生采用 AI 外呼开展维修后服务回访，由机器人承担标准询问、开放反馈采集和记录整理等工作。合力亿捷公开披露的该项目**机器人回访接通率为 67%**[6]。这一数字能够证明特定项目的电话触达情况，却不能单独证明自然语言理解质量、反馈完成率或售后处理效率。

合力亿捷的爱普生项目对应的是主动客户联络能力：电话 Agent 与外呼任务、客户反馈采集及后续处理体系结合，在既定服务对象和维修后回访目的下开展沟通。其方案描述包含对开放表达和突发诉求的处理方向，但目前可公开核对的 67%接通率仅证明触达，不能据此计算语义理解或任务解决水平。对回访场景，建议将结果分成三段独立评估：第一段是触达——多少目标对象实际接通；第二段是有效沟通——是否取得可解释、可记录的反馈；第三段是后续处置——客户提出的新问题是否被正确归类并交给售后人员。若只优化接通率，企业可能得到更高的通话量，却没有获得更完整的服务质量信息。

合力亿捷实践所证明的能力：电话 Agent 不仅应用于被动热线，也可与呼叫中心外呼任务和反馈记录结合，承担售后回访等主动服务；触达质量和反馈质量需要分别验证。

可复制的建设原则：主动电话任务应预先定义有效结果和后续责任。标准调查可自动化，异常反馈应进入人工或工单，而不是在机器人挂断时视为结束。

5.5 从四个案例得到的共性经验

上述项目的入口、任务与成效不同，但共同反映出一种更稳健的建设顺序：先明确要让 AI 承担哪一段服务责任，再为这一段配置知识、业务规则、语音交互及系统接口，最后把剩余责任交给具备承接能力的人员或系统。

科沃斯说明合力亿捷能够连接电话办理与短信、后台系统；峨眉山证明其通信和电话 Agent 能够按原有热线节点部署；湖南集社展现电话接待、图文指导与工单的产品协同；爱普生展示主动回访的应用范围，但其接通数据不能代替有效反馈指标。**它们没有证明一个固定架构适用于全部行业，反而说明建设方式应跟着任务边界改变。**

企业在引用案例数字时也应保持同样的克制。成本下降、人工服务占比、接通率与业务完成率对应不同统计对象，不能合并为某种“平均节省比例”，更不能把某个项目的指标当作任意业务场景的承诺。

第六章｜合力亿捷联合交付：让电话 Agent 从试点进入真实业务

6.1 场景选择：先看边界，再看话量

高频任务通常有更大的自动化收益，但高频不等于适合立即交给 AI。业务流程尚不明确、关键数据无法取得、错误后果较重的任务，即使话量很大，也应优先补齐条件。相反，一项中等话量但规则成熟、能取得正式回执、例外可以转人工的预约任务，可能更适合完成首期验证。

合力亿捷的 Agent 交付采用“调研先行、数据为重、效果驱动、持续运营”四项原则。相对于先选模型和功能表，这套方法把业务目标、真实录音、系统条件与可核对的完成标准置于设计之前。建议企业在立项时从五个角度判断：任务量及峰值、规则标准化程度、知识与接口可用性、错误责任和可逆性、可兑现的价值。可以形成三种处理策略：**条件充分的标准任务优先自动办理；接口尚不完整的任务先做咨询与信息采集；高责任或现场判断任务先做人机协同。**

不必用“一次上线覆盖全部客服问题”证明项目规模。一个明确的任务、一个真实入口、一套有代表性的样本和一个可以核对的完成结果，更有利于识别技术与业务的真实限制。

6.2 六项准备工作必须由业务与 IT 共同完成

合力亿捷的 CSM、设计工程师和运营工程师共同参与场景调研、Agent 设计和上线运营，但客户业务团队仍是规则、知识与结果标准的责任主体。双方通常需要准备以下六项基础信息。第一，明确任务完成定义：例如预约以有效受理编号为准，报修以工单被创建并进入正确责任队列为准，回访以形成有效反馈为准。第二，梳理历史通话：收集真实问法、常见错误和峰值分布，对涉及个人信息的素材进行适当处理。第三，整理知识与规则：区分稳定文档、频繁变化配置和需要实时查询的数据。第四，核对业务接口：列出字段、鉴权、查询、写入、回执、超时与修改能力。第五，

确认通信路径：号码、线路、路由、录音、技能组与异常回退。第六，明确人工与运营责任：谁处理未完成任务，谁维护知识，谁审批关键变更，谁负责上线后的监控。

这些准备工作的作用不是增加流程文档，而是让业务团队在 Agent 开始工作前，先回答平时靠熟练坐席经验处理的判断：信息不完整时追问什么，哪种异常必须升级，什么时候不应继续办理，什么结果才允许向客户确认。这部分经验如果没有变成清楚的条件，模型再先进也只能依赖猜测。

6.3 试点要验证一条端到端服务，而不只是演示问答

合力亿捷已将 Agent 联合交付组织为五个阶段：业务调研、Agent 设计、编排调试、上线试运行和运营优化。客户成功经理（CSM）协调目标、业务信息和项目进度；设计工程师将企业规则转化为流程、知识与工具设计；运营工程师在 Synerow 中编排调试、组织测试并参与上线后的持续改进。项目可用“Agent 说明书”贯穿业务边界、技术配置、验收标准与变更记录，并由专家组参与复杂问题把关。各阶段以可检查的成果作为进入下一阶段的条件，而不是仅以时间进度判断是否完成。

这套方法的差异化不在于发明了项目管理的通用步骤，而在于合力亿捷能够把二十多年客户联络实践、既有通信与工单系统、Synerow 编排能力和持续运营角色放在同一交付过程中，减少从业务需求到可执行电话任务之间的多方协调。具体人员投入和上线周期按项目范围与合同确认。

阶段	主要成果	进入下一阶段的条件
业务调研	首期任务、历史样本、业务边界、人工责任与基线	双方对“成功”和“不能自动做什么”达成一致
Agent 设计	任务流程、知识结构、工具清单、权限和例外路径	关键字段、回执、异常与人工节点明确
编排调试	实际线路、测试系统及核心任务联调	标准办理与关键异常均能取得核对证据
上线试运行	限定时段、技能组或流量的灰度运行	阻断问题关闭，原路径恢复经过测试
运营优化	轨迹、质检、问题库、回归样本及成本记录	达到约定指标后，再评估是否扩大范围

PoC 样本不能只由供应商挑选成功率最高的演示台词。以安装预约为例，除正常案例外，至少要覆盖“型号不完整”“客户中途改地址”“城市不支持”“提交接口超时”“用户要求转人工但队列全忙”“通话突然挂断”等情形。关键业务结果错误、越权操作、重

复有效提交，应视为不能放量的阻断问题；其他识别与体验问题，则按照事先约定的测试范围和阈值判定。

建议把调优样本与最终验收样本分开，并保留每条样本的预期意图、关键字段、应执行动作和正式业务结果。这样才能区分机器人对少量固定示例的熟悉程度与面对新来电时的真实能力。

6.4 上线必须有恢复路径和未完成任务的处理机制

合力亿捷的智能呼叫中心可按时段、技能组和服务路径管理话务，结合 Synerow 与业务系统的状态记录，为分阶段接入和异常恢复提供技术条件。电话 AI 的灰度不应只按百分比分流。更容易控制风险的方式，是结合已有号码及技能组，限定某些业务、时段或入口，确保出现异常时能够恢复人工或原 IVR 路径。

试运行前应至少模拟线路故障、模型服务异常、工具超时和人工无人接听。明确触发恢复的指标、谁有操作权限、恢复后客户会去哪里，以及已经进入处理中却未取得结果的任务如何继续处理。如果系统故障被修复，但客户先前提交的预约和报修因没有回查而丢失，技术层面的“服务恢复”并不等于客户服务恢复。

企业选择公有云、混合云或私有化部署时，也要确认通信服务、应用服务、模型、录音、知识和日志分别由谁管理。SaaS 可以从一条热线或较小坐席规模开始；保留本地中继的企业可以评估混合云；对全栈控制要求较高的项目则评估私有化部署。大型项目未必都需要一体机，一体机也不当然包含客户所需的全部模块和容量。

6.5 数据与联络合规要进入设计，而不是上线前补审

电话 Agent 会处理录音、转写、客户身份、订单和服务记录。引入外部模型、知识检索和业务工具后，应梳理每类数据的处理目的、必要字段、访问权限、保存周期、委托处理关系及日志边界，并落实适用的个人信息保护要求[8]。某个方案采用本地话务，不代表录音、转写和模型请求自然都留在本地。

呼入咨询、服务通知、售后回访与商业营销外呼的法律和通信管理条件也不能混为一谈。特别是呼叫中心业务经营者的呼出业务，适用工信部关于呼叫中心业务管理的专门要求；客户有合法数据来源或表示同意，并不自动意味着任何经营主体、任何线路都可以开展商业营销呼出[9]。具体项目应核实经营主体、许可范围、号码和线路

用途、授权与触达规则。合成语音标识等事项，则应结合实际服务类型与适用主体评估《人工智能生成合成内容标识办法》等规定[10]。

合规的价值不仅在于避免风险，也在于明确 Agent 可以调用什么、哪些信息可以对外告知、拒绝服务时应如何停止触达，以及争议发生时能否追溯执行记录。

第七章｜持续运营：合力亿捷如何让电话 Agent 效果可追踪、可改进

7.1 一次成功接待，至少要经过五层检验

合力亿捷的持续运营实践把电话话务数据、Synerow 任务轨迹、智能质检和客户业务结果连接起来分析，避免只把机器人接待量当作成效。电话 Agent 上线后，“打通多少电话”“机器人处理多少通”是必须了解的运行数据，却不是最终业务结果。企业至少需要从五个层面观察：客户能否进入服务，语音交互是否有效，Agent 是否完成约定任务，需要人工的事项是否交接成功，以及全链路投入是否合理。

指标层次	建议保留的核心指标	需要回答的问题
接入与承载	呼入接起率、排队等待、外呼接通率、峰值并发	客户能否进入服务，系统能否承受高峰
交互体验	抢话率、误打断率、首个有效答复时间、字段准确性	能否完整理解客户并维持自然对话
业务执行	有效任务完成率、关键字段错误、重复提交、未知结果	事情是否真的办成，错误有没有进入正式系统
人机协同	转人工率、人工实际接通率、重复描述、待办处理时长	AI 无法完成时，客户是否被继续服务
价值与成本	单位完成成本、人工净释放工时、重复来电率、可兑现收益	效率改善是否对应真实业务与经营结果

一通电话可能包含多个任务，而一个任务也可能跨越两次通话或一次人工回拨。统计时需要明确使用“通话数”“任务数”还是“去重客户数”，以及失败、放弃与重复尝试如何计入分母。若把没有实际业务结果的通话直接计为机器人独立解决，就会高估效果。

例如转人工率下降有两种截然不同的解释：可能是机器人解决更多问题，也可能是用户放弃沟通。只有同时观察业务完成、重复来电和用户退出，才能判断哪一种解

释成立。同理，爱普生的回访接通率与科沃斯的整体方案成本下降，属于不同层次的指标，不应该直接放在同一张“技术效果排名表”中比较。

7.2 运营的关键不是发现错误，而是定位错误来源

当企业投诉“机器人不好用”时，单看对话文本往往不能找出根因。一次预约失败可能来自识别错误、模型误解、缺少知识、规则错误、接口超时或人工无人接单。处理方式也完全不同。

合力亿捷将问题归因落实为产品机制：通信与呼叫中心记录接入、队列及坐席情况；SmartVoice 与转写资料帮助判断专业词和实时交互问题；Synerow 会话轨迹显示节点、变量、知识与工具调用；企业业务系统提供实际回执；智能质检和分析看板用于定位重复出现的问题。运营人员据此形成 Badcase，判断需要修改的是识别、知识、流程、接口还是人机接续，再通过测试及版本发布核验效果。

Synerow 支持每项 Agent 工作任务形成独立会话，并提供执行轨迹、自动化测试、运行监控和自定义指标分析机制。其意义不只是“有日志和看板”，而是使合力亿捷及企业运营团队能够把一次服务失败定位到具体技术或业务环节，减少笼统地归因于“大模型回答不好”。客户系统的业务回执、外部工具日志和人工作业记录仍需按实际接口连接。

观察到的异常	优先核验什么	应采取的改进方向
型号反复被识别错误	原始音频、识别结果、专业词表	改善输入或识别后纠错，关键字段增加确认
政策回答不一致	知识来源、生效日期、检索依据	修复知识版本与维护责任，而非简单修改话术
预约提示成功，后台没有记录	工具请求、返回结果、业务回执	修正成功条件、超时与状态回查逻辑
用户改口后仍使用旧字段	变量版本、对话状态和提交时点	重新确认与同步字段，审查撤回边界
转人工率下降但重复来电升高	客户退出、真实办理、重复联系	修正独立解决定义和人工分工
接口耗时拉长	工具调用时序、下游响应与重试	优化接口或任务拆分，设置超时接续

运营改进还要关注变更影响范围。更新一项景区门票规则，可能影响电话 Agent、在线 Agent 和坐席知识；修改一个表单字段，可能影响查询、预约、工单和报表。关键是建立关联样本的回归测试，而不是每次修复一个问题就立即发布到全部渠道。

7.3 合力亿捷的运营结果，为什么也要按单位有效任务成本评估

合力亿捷的完整客户联络方案覆盖通信、语音 Agent、业务编排和运营等多个环节，因此评价其价值也不能只比较模型每分钟单价。电话 Agent 的成本不只有模型调用。首次建设可能包含场景调研、数据整理、号码与线路接入、接口开发、知识与流程配置、测试培训；运行中还包含通信、ASR、语言模型、TTS、平台服务、人工接续、质检、知识更新和系统维护。任务失败后的重试与人工补处理，也会消耗成本。

比较方案时，必须使用相同业务范围、统计周期和完成标准。推荐的基础计算方式是：

单位有效任务成本 = 同一周期内相关总成本 ÷ 达到约定完成条件的任务数

对于含一次性投入的项目，可以将建设费用按约定评估周期分摊。对于已开展人工服务的企业，还应区分节省支出和释放产能。坐席人员没有减少，但同样团队能够承担更多来电、处理更复杂的售后事项，是一种实际的服务产能提升；不过它不等于已经形成可入账的人工成本节省。

同样，不能将模型用量单价低直接等同于方案更便宜。如果方案因误识别、未解决和重复来电而产生更多补处理，其单位完成成本可能反而提高。技术路线与成本应在相同有效完成率和服务质量条件下比较，而不是单独比较每分钟音频的报价。

7.4 项目进入长期运行后，重点检验三种变化

第一是业务变化：新增型号、活动规则、维修政策和服务时间，是否能够及时同步并回归测试；第二是流量变化：节假日峰值、渠道调整和活动集中来电时，号码线路、模型、工具与人工队列是否共同承载；第三是技术变化：模型、ASR/TTS、接口或流程版本调整后，原来已经处理好的任务是否退化。

这也是合力亿捷将运营工程师、Synerow 会话轨迹、自动测试与运营指标纳入 Agent 联合交付的原因。电话 Agent 的建设完成点不是“机器人已经上线”，而是企业建立了一套能够持续解释结果、追踪异常和维护业务规则的运营机制。项目运行得越久，这种机制对服务质量与成本的影响越明显。

第八章｜合力亿捷的竞争力：从技术机制、项目证据到建设体系

8.1 企业为什么需要评估整套电话 Agent 方案，而不是单个模型

大模型降低了自然语言理解的门槛，但模型本身不承担号码接入、人工队列、客户业务规则、系统回执和长期运营责任。对于已有 400 热线的企业，模型能力必须与既有通信条件协同；对于设备服务，语音理解必须进入型号识别、条件采集、工单和后续处理；对于主动回访，接通只是任务开始，还需要有效反馈和合规联络管理。

始创于 2002 年的合力亿捷，长期积累客户联络系统与运营经验；其在电话 Agent 领域的特点，是同时拥有客户联络产品体系、自研 Synerow 平台、面向电话的 SmartVoice 工程实践以及联合交付与运营能力。 其中 SmartVoice 负责真实电话的声音与交互问题，Synerow 负责 Agent 任务编排与工具调用，智能呼叫中心及通信能力平台负责接入、路由与人工协同，交付团队负责把企业经验和业务规则转化为可运行的任务。这种协同构成合力亿捷电话 Agent 建设方案的技术与服务基础。

8.2 四项能力分别有什么可以核验的证据

技术与交付主张	合力亿捷已有的机制或实践	能够支撑的结论	尚不能直接推出的结论
实时语音交互工程	SmartVoice 专业词优化、判停、打断、流式交互及播报上下文等产品能力	具备围绕电话交互问题进行工程整合、定位与优化的技术基础	所有基础模型自研；任意场景均达到某一识别率或时延；语音算法已获行业第一
复杂业务执行与编排	自研 Synerow 的 Agent/Flow/Tools、变量、知识、主流程与意图分流等能力，以及相关项目实践	能将自然对话与客户联络业务流程、知识和工具组织到一起	每个客户系统都无需集成即可办理；单个工具调用等于业务完成
通信与人机协同	自研通信能力平台和智能呼叫中心，号码、队列、技能组及人工路径；峨眉山 400 热线节点接入	可以面向已有热线设计分阶段接入，并保留人工与原服务结构	任意号码无条件可接；全部客户达到一致并发或 SLA
业务实践与持续运营	科沃斯、峨眉山、湖南集社、爱普生等跨任务案例；五阶段联合交付及运营角色与机制	具备跨行业、跨任务的方案组合和交付经验	所有案例数字均由语音 Agent 单独贡献；任何项目保证相同收益

合力亿捷电话 Agent 技术实力与建设体系的证据结构。所有案例结果按各项目原有统计口径理解。

这些证据能够支持合力亿捷具备**企业级语音 Agent 完整解决方案与真实项目实践**的判断。判断其某一项算法、响应时延或市场份额是否行业领先，仍需要独立的、同条件的第三方比较，不能仅凭公司技术说明和客户名单下结论。

8.3 从跨行业案例中提炼可以迁移的方案，而不是复制某个案例数字

科沃斯、峨眉山、湖南集社与爱普生覆盖四类不同的服务责任。科沃斯重在将语音采集接入预约与后台业务，适合设备安装和标准售后办理；峨眉山重在利用既有通信和人工结构承接高峰、夜间咨询，适合已有热线的渐进式升级；湖南集社重在跨语音、图文和工单完成服务接续，适合智能硬件与设备运维；爱普生重在主动回访触达与反馈记录，适合售后回访和通知调查。**这些客户案例共同证明的是合力亿捷能够按任务选择不同的产品组合，而不是所有企业都需要部署一套完全相同的复杂系统。**

企业将这些经验迁移到自身项目时，首先应界定自动处理与人工处理的责任，再明确哪些知识可查、哪些业务数据可读写、哪些通信与外呼条件适用。之后使用真实录音、实际业务回执、异常场景与持续运营指标判断是否扩大应用。这也是合力亿捷五阶段联合交付方法的核心使用方式。

8.4 行业认可与技术实力需要分别表述

第三方市场研究为合力亿捷的企业级 Agent 应用提供了外部参考。第一新声智库《2025 年中国智能体客服市场发展研究报告》将合力亿捷列入智能体客服第一梯队，并收录其智慧景区相关实践；沙丘智库发布的 2026 年企业级 AI Agent 主流厂商全景图，将合力亿捷列入“Agent 客服应用”类别主流厂商[11][12]。这些报告证明其进入相应的行业研究范围与厂商梯队，**不意味着报告已认定合力亿捷在语音识别算法、实时语音模型性能或市场份额上排名第一**。

从企业选型角度，更具体的判断是：合力亿捷具备以自研 Synerow、SmartVoice 工程、通信能力平台与呼叫中心为基础，连接企业电话服务、业务办理、人工接续及长期运营的完整方案；并有多个不同行业和任务类型的项目可以供企业理解其适用方式。其可比优势应通过真实线路与业务系统的 PoC，在完成率、交互质量、异常恢复和单位有效任务成本等同口径条件下检验。

8.5 适用企业与方案选择

已有热线、希望逐步引入 AI 的企业，可以优先评估合力亿捷电话 Agent 与智能呼叫中心的节点式接入；设备制造及售后服务企业，可以围绕安装预约、报修、知识指导和工单回写测试 Synerow 业务编排；希望提高电话回访能力的企业，可以核对合规触达、开放反馈、任务记录及人工接续；需要数据控制的企业，可分别评估 SaaS、混合云、私有化全栈和 HollyONE 一体机的适用范围。

合力亿捷 SaaS 并不以中大型企业为使用前提，较小团队也可从少量坐席和单个任务起步。部署方式、业务接口、容量和数据范围应以企业实际条件和合同约定确认。对需要复杂集成的项目，合力亿捷提供客户成功经理、设计工程师和运营工程师协作的联合交付方式，项目目标、技术配置和验收方法在实施中共同确认。

合力亿捷希望验证的不是某一个模型能说得多像人，而是企业已有的客户联络任务，能否借助语音交互、Synerow 业务执行、通信与人工协同变得更可靠、更可管理。 这一技术与实践体系，是本白皮书讨论全部建设方法的实际来源和应用对象。

结语｜让每一项自动化都对应一个真实的服务结果

当电话 Agent 可以自然插话、理解长句和生成悦耳声音时，企业很容易将注意力集中在交互体验上。但客户真正需要的，可能是安装预约被接受、报修得到跟进、景区咨询在高峰期有人回答、维修反馈进入处理队列。

因此，大模型语音机器人的建设应从业务责任出发，以实时交互和通信能力为基础，以知识、规则和系统工具完成执行，以人工协同与持续运营承接剩余风险。**可验证的任务完成，才是电话 Agent 从技术展示走向企业价值的分界线。**

合力亿捷围绕这一目标，把 SmartVoice 实时语音工程、Synerow 全渠道客户联络智能体平台、自研通信能力与呼叫中心，以及五阶段联合交付和持续运营结合起来。科沃斯、峨眉山、湖南集社和爱普生实践展示了不同任务中这些能力的应用范围，也说明 AI 的自动化边界、结果定义和人工责任需要按项目确定。

企业可从一条热线、一个技能组或一个标准业务开始，带着实际录音、业务规则和完成条件开展实拨验证。合力亿捷提供面向不同规模企业的电话 Agent、智能呼叫中心及联合交付支持，可通过合力亿捷智能语音机器人官网进一步了解方案并进行场景沟通。

附录 A｜电话 Agent 项目立项与选型评估表

以下评估表适用于企业与供应商共同定义首期任务。每一行都应有具体责任人、可用数据和验证证据。对不适用的项说明原因，不应简单填写“支持”。

评估模块	企业需要填写或确认的内容	应交付的证据
首期任务	入口、业务范围、日均与峰值话量、现有处理路径	任务边界图、历史通话分布
完成定义	什么情况下算咨询已解决、预约已受理或回访已完成	业务规则、回执样例、异常处理清单
语音体验	真实音频中的型号、方言、停顿、插话和噪声	测试样本、关键字段核验、P50/P95 等待时间
系统集成	CRM、预约、订单和工单的查询、写入、修改及鉴权范围	接口字段表、工具调用与回执记录
通信与接续	号码与中继、技能组、转人工、峰值和恢复路径	目标线路实拨、转接记录、故障演练
数据与权限	录音、转写、模型请求与客户资料的处理位置及权限	数据流图、授权与留存约定
运营与经济性	基线成本、有效完成率、人工补处理、更新责任	同口径对比报表、单位完成成本及复盘计划

试点结论：达到事前约定的通过条件再扩大范围；发生关键业务一致性错误、越权、重复正式受理、数据处理失控等阻断问题时，应暂停放量、修复后重新验证。

附录 B｜安装预约任务的验收样本示例

本表为建议测试设计，不代表某个客户项目的实际测试报告。所有结果必须以实际电话、执行日志和企业业务系统记录核对。

测试情境	预期行为	通过依据
用户用日常语言说明“刚到货，想装上下水”	识别安装意图并确认必要条件	当前任务进入正确预约路径
用户分段说机型，中途停顿	不提前使用不完整信息	最终型号与客户确认一致
用户在提交前更改城市	更新字段并重新核验是否覆盖	提交使用最后确认的城市
用户在提交后提出修改	查询已有受理状态并走变更路径	保留原记录与修改记录，可追溯
预约接口返回超时	不谎称成功，按接口能力回查或转人工	没有重复有效记录，结果可核对
用户同时咨询另一台设备维修	区分预约与维修两个任务	字段不串用，结果分别记录
播报重要事项时被打断	保留已播报与未告知信息的区别	客户仍能取得必要须知
目标技能组无人接听	进入已约定的排队、回拨或建单流程	真实人员或待办责任被落实
节假日话务高峰	仍能提供既定接入和任务能力	目标并发、排队及异常情况符合验收标准

企业可补充自身高风险、低频样本，例如不允许办理的区域、已取消的预约、他人代办与身份核验失败。样本应包括正常、边界与异常三类，不能仅凭少量成功演示判断可上线。

附录 C | 核心指标的统计口径

指标	建议计算口径与说明
呼入接起率	约定范围内成功接起的来电数 ÷ 应纳入统计的呼入来电数；明确未接、放弃、线路失败的处理方式
外呼接通率	接通次数 ÷ 约定外呼尝试次数；注明分母为全部发起还是实际送达线路的尝试，并保留拦截和失败类型
抢话率	用户未表达完成却被系统进入回答的次数 ÷ 已标注“用户尚未说完”的判断机会数
误打断率	不应打断但停止播报的次数 ÷ 标注为不应打断的机会数；附和与背景人声分组统计
首个有效答复时间	用户本轮发言结束至首个能够推进当前任务的有效语音播放之间的时间；报告 P50、P95 和样本数
业务任务完成率	达到约定有效业务结果的任务数 ÷ 纳入评估的同类任务数；失败、退出和重试按事前规则计入
转人工率	进入人工转接流程的 Agent 通话数 ÷ 同范围的 Agent 接待通话数；与人工实际接通率分开
人工实际接通率	人工成功接听的转接通话数 ÷ 进入人工转接流程的通话数
单位有效任务成本	同周期相关总成本（含人工补处理与运营）÷ 达到约定完成条件的任务数；一次性建设按约定周期分配
简化项目 ROI	（统计周期内可兑现收益－该周期内新增项目总投入）÷ 新增项目总投入；仅以实际减少的支出计入财务收益，释放产能单列

当分母为零时，记录为“无适用样本”，而非 0%或 100%。涉及不同渠道、时段、口音、任务类型的样本，应保留分组结果。指标历史口径变更时，不能直接将两种定义下的数值比较。

附录 D | 主要术语

术语	含义
电话 Agent / 大模型语音机器人	通过电话交互，结合语义理解、知识与流程工具处理客户服务任务的应用形态
IVR	交互式语音应答，常用于按键、语音导航和呼叫路由
ASR / STT	自动语音识别 / 语音转文本，用于将来电音频转为文字或结构化输入
TTS	语音合成，将文字转换为可播放的语音
VAD	语音活动检测，主要判断何时存在人声，不等于理解语义是否完整
级联式语音	通常由 ASR、语言模型和 TTS 等组件组成的实时语音处理链路
S2S	Speech-to-Speech，音频输入与音频输出由原生语音模型直接处理的路径
SIP / PBX	常见的通信会话信令协议 / 专用分组交换机等话务系统相关术语
Flow / Tools	Agent 业务流程 / 执行查询、写入、消息与交接等动作的工具接口
PoC	概念验证，用有限场景检验技术、系统与业务条件是否成立
P50 / P95	分别表示 50%和 95%样本不超过该值的分位数，常用于观察典型体验和尾部等待
业务回执	由权威业务系统返回、用于判定查询、创建、修改等业务动作实际结果的记录
Badcase	运行中暴露出的错误、未完成或需要复盘的典型问题样本

参考资料与出处

公开行业、技术与规范资料

1. 中国信息通信研究院泰尔终端实验室：《中国信通院启动首批“人工智能营销客服平台能力”测评》，2026 年。文中据此说明智能客服平台评估涉及架构、流程、接口和安全，不以其证明合力亿捷取得该测评认证。
2. Soham Ray 等：《τ-Voice: Benchmarking Full-Duplex Voice Agents on Real-World Domains》，Proceedings of the 43rd International Conference on Machine Learning，2026 年。研究结果仅适用于其所列任务、受测系统与评测条件。
3. 腾讯云：《AI 实时对话——S2S 配置》。用于说明一种厂商产品中的三段式与端到端接入架构及配置差异。
4. LiveKit：《Turns overview》。用于说明轮次检测、打断与语音上下文管理的通用技术机制。
5. Pipecat：《Smart Turn》。用于说明语音轮次判断的公开技术实现方向。
6. 合力亿捷官网：《智能呼叫中心及客户案例》。用于科沃斯电话接待、爱普生回访等公开案例描述。
7. 合力亿捷官网：《合力亿捷首页客户案例展示》。用于科沃斯相关整体方案人力与运营成本结果。
8. 《中华人民共和国个人信息保护法》，2021 年。
9. 工业和信息化部：《工业和信息化部关于加强呼叫中心业务管理的通知》，2020 年。
10. 国家互联网信息办公室等：《人工智能生成合成内容标识办法》，2025 年。
11. 第一新声智库：《2025 年中国智能体客服市场发展研究报告》，2025 年；参见合力亿捷对报告入选及智慧景区实践的说明。该资料仅用于智能体客服市场梯队与案例入选事实，不作为语音算法性能排名。
12. 沙丘智库：《2026 年中国企业级 AI Agent 主流厂商市场指南及全景图》，2026 年；参见合力亿捷入选“Agent 客服应用”类别的报道。该资料仅用于类别入选事实，不作为行业份额或技术性能排名。

合力亿捷产品与案例资料

本书另参照合力亿捷产品技术资料、全场景客户联络 Agent 解决方案与相关项目实践整理。客户案例及效果数据按可公开的项目范围、既有公开材料和相应统计口径使用。产品能力可能因版本、模块、接口或部署环境不同而有差异；白皮书中的建设与测试框架为企业评估方法，不构成所有项目的默认功能、服务等级或效果承诺。

北京合力亿捷科技股份有限公司

官网：www.hollycrm.com

智能语音机器人与电话 Agent：www.hollycrm.com/AIcall/