

客户理解与行动引擎

行业调研、产品蓝图与双场景试点设计

研究截至：2026 年 9 月 20 日 | 建议设计稿

本轮已确认：现有 ASR（语音转文字）平台能够提供逐句文本、说话人和时间戳；首期销售与车主服务各选一个场景试点。本文核查了官方产品资料、模型卡、开源项目说明、许可证及部分关键源码，未运行客户数据评测，未验证长城生产接口。文中的架构、场景、资源和计划均为建议。

1. 我建议建设什么

建议将产品定位为“客户理解与行动引擎”：持续理解客户当前要办成什么、与品牌和门店发生了什么、我们还欠他什么，并把这些理解交给顾问和智能体，帮助他们把下一件事办好。

它服务两个同样重要的目标：客户在需要时得到合适的帮助；门店通过合适的成交与服务获得收入，维护客户关系。少打扰、兑现承诺和尊重客户选择，都是产品效果的一部分。

具体运行方式是：录音、聊天和业务变化进入系统，形成有出处的需求、顾虑、承诺和事项；系统结合车型、车辆、政策及门店承接能力，为当前任务准备信息；顾问与 Agent（智能体）共同采取行动；最后从预约、订单、工单和客户反馈中确认结果。

以前客户中心常常回答“这个客户是谁，有哪些标签”。新引擎还要能回答六个问题：

1. 他现在想完成什么？
2. 哪些需要和顾虑已经明确，哪些仍不清楚？
3. 最近发生了什么变化？
4. 我们答应了他什么，兑现到哪一步？
5. 现在做什么对他有帮助，门店是否能办到？
6. 做完以后，客户的问题和业务状态是否真的改变？

这六个问题应成为产品页面、数据结构和智能体工具的共同主线。

先校正三个判断

****CDP（客户数据平台）的基础能力仍然有价值。****身份关联、数据接入、授权和人群分析可以复用。需要升级的是停留在静态标签、营销名单和定时活动的使用方式。行业产品也在向对话理解、任务上下文和执行发展，并非都停在旧阶段。

****收集得更多，并不自动等于理解得更深。 **对成交真正有用的，可能是客户一句“家里还没确认充电条件”；对安抚真正有用的，可能是“昨天下午答应更新进度，却一直没有回复”。** 这些信息通常比上百个泛化标签更有帮助。

****客户引擎应同时提供理解和行动依据。 **如果只汇总资料、生成摘要，顾问仍然要自己查政策、找配件、确认时段，业务价值会停在阅读便利。引擎必须连接正式业务能力，但订单、价格、库存、索赔等最终状态仍由相应业务系统负责。**

2. 行业已经走到哪里

下表是公开产品方向及其对本项目的启发，不代表它们已在长城落地，也不代表供应商宣传的经营收益能直接复用。

核查对象	官方资料显示的进展	对我们的启发
Salesforce Data 360（客户数据与语义平台）	将非结构化内容接入客户数据，并强调给智能体提供统一业务含义	结构化交易与对话证据应能共同被查询；语义不能只靠提示词临时猜
Microsoft Customer Insights（客户洞察产品）	2026 年 7 月介绍以持续更新的客户资料支持销售、服务和旅程；部分外联优化能力仍属公开预览	客户理解要供多个业务场景共用；区分发布方向、预览和正式可用
Twilio Conversation Memory（跨渠道对话记忆）	2026 年 5 月发布持续客户记忆、按当前问题召回信息，并与企业产品及政策知识配合	客户记忆与企业知识分开维护，使用时再组合；“记得客户”和“知道能提供什么”缺一不可
Decagon（客户服务智能体产品）	2026 年推出用户记忆；9 月的 Campaign Composer 支持持续外联、等待、重试和退出控制	主动服务是跨天持续的事项，需要处理回拨、暂停、人工交接和结果，不能只做一次外呼
Tekion（汽车零售软件）	2026 年 6 月发布 T1 Pro 及销售、交易检查等智能体，强调与汽车零售业务上下文结合	汽车行业的竞争已进入业务流程；我们应突出对商机、交付、维修与客户承诺的理解
火山引擎销售助手	官方产品页提供客户建档、跟进策略、对话提取和车型知识等能力，并列出吉利销售邀约案例	国内已有“录音理解+销售辅助”的产品实践；仅做摘要、标签和话术很难形成足够区别

来源：Salesforce 非结构化数据、Salesforce 语义层、Microsoft 2026 年 7 月产品方向、Twilio 对话记忆、Decagon 持续外联、Tekion 新智能体、火山引擎销售助手。

我的判断是：方向有行业依据，但“跨渠道客户记忆+智能体”本身已不是独有概念。我们值得投入的部分，是汽车业务中可延续的客户理解、可兑现的服务安排，以及销售和售后共同使用的工作方式。

具体差别应当能演示出来：同一客户从看车到交付，再到维修，换了顾问、换了渠道，系统仍能找回与当前事情相关的需要和承诺；信息过时了会修正，事情办完了有业务回执，经验能够帮助处理下一位客户。

3. 客户中心的业务骨架

3.1 以客户为中心，围绕正在发生的事组织信息

建议保留八类业务对象。它们是逻辑模型，不意味着要采购八套系统。

对象	主要内容	汽车行业特别要处理的问题
客户与关系	可核实身份、联系方式、授权、角色、与门店的关系	购车人、付款人、车主、实际用车人可以不同；同一号码可能由家庭成员共用
车型与产品	车型、年款、配置、功能、适用政策、产品说明	售前客户可能还没有车辆识别代号，不能以车辆档案为唯一入口
车辆	VIN（车辆识别代号）、配置、履历、有效车况与适用权益	一人多车、车辆转手、公司车与个人使用必须按关系和有效时间区分
商机与服务事项	一次购车考虑、交付事项、维修问题、权益兑现等	同一客户可以同时有增购商机和未完成维修；不能用一个“生命周期阶段”覆盖全部
互动与证据	对话片段、录音位置、文件、正式业务事件	保留谁说、何时说、针对什么事，以及是否经过核实
需求、顾虑与承诺	用车需要、购车约束、待确认问题、顾问承诺	将客户原话、顾问判断和厂家政策分开；承诺须有责任人和时点
行动与履约	联系、预约、试驾、预工单、交接、升级处理	统一联系安排；正式写入相应业务系统，记录是否成功
结果与方法	客户反馈、业务结果、有效步骤及适用条件	成交或不再投诉不是唯一结果；还要知道何种帮助被使用、哪些判断需要修正

已有客户主数据、CRM（客户关系管理系统）、DMS（经销商业务管理系统）负责的正式对象，应沿用其标识和归属。引擎增加的是关联、理解、任务上下文和协作能力。

3.2 一张“活的客户理解卡”

顾问打开客户页面，首屏建议只呈现这些内容：

- **当前事项：**正在考虑的车、等待的交付、尚未解决的维修问题。
- **已明确的需要：**客户自己说过什么，有什么现实约束。
- **待确认的问题：**例如充电条件未核实、预算指裸车还是全部支出仍不清楚。
- **我们作出的承诺：**谁负责、何时更新、是否可能失约。
- **沟通偏好：**允许的用途、渠道、时间，以及近期已经联系的情况。
- **建议下一步：**为什么现在做、需要谁参与、系统已准备好什么。

每条关键内容可以展开看到原始依据，修改时保留更正记录。摘要是一种阅读方式，不是新的事实来源。

3.3 “持续连接”如何体现

持续连接不等于持续发消息。建议由需要、事件与约定共同唤醒服务：客户明确提出需要；业务状态发生有意义的变化；约定的更新时间到了；或出现值得核实的服务风险。

系统先判断是否有实际帮助、当前有没有其他团队在处理、是否适合联系、门店能否承接，再选择联系客户、安排内部任务、继续等待或结束事项。

主动招揽承接真实需求；主动安抚在延期、失约和反复解释等问题扩大前承接责任；主动关怀帮助客户把车用好、把权益用上。三类服务共享事项与联系安排，避免同一天由多个团队各自找同一客户。

3.4 平台最终应呈现哪些工作页面

工作位置	使用者要完成的事	主要呈现内容
客户工作台	顾问接待、跟进、交接客户	当前事项、关键需要与证据、未兑现承诺、下一步；支持搜索、更正和转交
今日任务	顾问知道今天该处理谁、先处理什么	有依据的优先顺序、联系安排、准备好的资料和动作，明确主责
商机与服务进度	店长安排资源、发现受阻事项	沿用现有商机阶段与工单状态，补充阻断原因、承诺风险和协作需求
客户群与方法	总部、区域与店长理解共性问题	可比较的群体、需求变化、服务障碍、方法适用条件与结果
知识与数据维护	业务负责人维护可用信息	车型及政策版本、错误关联、更正、授权与待确认内容

它们可以嵌在现有 CRM、DMS 和顾问移动端中，不必让员工再维护一套独立台账。管理视角与顾问视角按岗位、门店和用途控制范围；客户也应能在合适入口查看服务进度、调整联系选择并提出信息更正。

4. 从现有 ASR 接到客户理解

已确认的逐句文本、说话人和时间戳足以成为第一阶段的重要输入。需要补的重点是语义和关联，不是先重建语音转写。

4.1 一段对话经过六步处理

- 接收与定位：**记录会话标识、来源、门店、时间、原始片段位置和转写版本。
- 确认参与者：**说话人 A、B 是声道或角色线索，不自动等于已核实客户；与顾问、客户和事项建立可靠关联。
- 提取表达：**识别需求、问题、选择、拒绝、联系偏好、承诺和待办；保留否定、条件和不确定表述。
- 检查冲突：**与已有记录、政策和当前业务状态对照；区分新变化、转写错误与尚未核实的说法。
- 更新理解：**低风险内容可按规则自动形成有来源的记录；身份合并、价格承诺等重要变化交相关人员确认。
- 产生任务：**例如准备车型对比、核实充电条件、提醒按时更新维修进度，而不只是增加标签。

不要伪造 ASR 未提供的置信度。首期应抽样测出车型名、配置、金额、日期、否定词以及说话人分离的实际错误情况。错误片段可回听，修订转写后要更新它产生的摘要和任务。

4.2 示例：同一句话可以产生不同类型的信息

以下均为虚构设计示例。

原始表达或业务信息	应如何记录	不应变成什么
客户：“如果充电方便，我可以考虑插混。”	对插混有条件兴趣；充电条件待核实	已决定购买插混
顾问：“我争取周五给您交车。”	顾问表达交付目标，需确认正式安排	厂家已承诺周五交车
客户：“最近上班别打电话，微信留言就行。”	有时间范围的联系偏好；仍需核验渠道和用途权限	所有营销均已获得永久授权
客户：“上次修过以后又响了。”	客户再次报告同类现象，关联原事项待核查	已证实维修质量事故
DMS 返回预约编号和确认状态	预约系统已确认，记录回执及查询时间	模型仅说“我帮您约好了”

第一阶段优先提取七类内容：使用需求、明确约束、决策问题、客户选择、联系偏好、顾问承诺、未完成事项。泛化性格标签不作为首期目标。

5. 数据怎么汇总：统一理解，保留来源和归属

“一份数据湖可以服务很多业务”这个方向成立，但物理上不必把所有东西搬进一个数据库。原始媒体、正式业务记录、检索索引和群体分析的更新频率、权限和保存要求不同。

数据部分	建议放什么	使用方式
原始证据区	合法取得的录音、转写、聊天、附件及其版本	按需回看原始片段；不默认反复全量送入模型
客户与事项记录	已关联身份、需求、承诺、事项、来源和有效时间	维护可查询的当前理解；明确哪些来自正式业务系统
检索索引	关键词、语义向量、可选的关系图索引	帮助找到证据；可从原始记录重建，不能单独决定事实
分析区	经目的限定、必要处理后的群体特征与业务结果	找共性问题、比较方法、评估经营效果
产品与方法知识	车型资料、政策版本、服务方法、经审核的案例	按车型、地区、门店和时间选择适用知识

数据源按业务价值逐步接入

首期接 ASR 输出、CRM 商机与试驾记录、DMS 工单与待件状态、现有客户身份和联系权限。企业微信、App 等渠道在接口及授权明确后接入。

个人微信与企业会话存档的可获取条件不同，不能默认所有顾问私人聊天都能采集。视频先用于明确、有授权的业务证据，不从人脸、穿着或声音推断财富、健康、性格；涉及图片或视频时，按任务取片段，避免全量常驻分析。

论坛等外部内容分两条路径：公开讨论用于发现产品和服务共性问题；个人服务在用户自愿进入正式渠道、完成身份核验后，再按用途关联。昵称相同、头像相似或模糊车牌都不足以自动合并客户。

《个人信息保护法》要求有合法处理基础并遵循必要范围，对公开个人信息再利用、自动化营销等也有相应限制。项目应把用途、联系选择和撤回落实到数据查询与执行中，而不是只留一份上线前说明。具体处理基础由长城相应负责人确认。[法律原文](#)

6. 客户记忆应怎样维护

6.1 四种信息分别保存

信息类型	例子	更新原则
正式业务事实	当前预约、工单状态、报价与权益	由负责该业务的系统确认；使用前检查时效
有出处的表达	客户需求、明确偏好、顾问说过的话	保留说话人和原文；“说过”与“已核实”分开
任务中的判断	可能因等待过久产生担忧、需要进一步解释配置	带依据、未知项与期限；不静默变成永久客户属性
可复用的方法	某类需求要先问哪些问题、何时应该转技术人员	从多个有结果的案例形成，经审核后发布

客户记忆解决“哪些与他有关的信息值得保留”；任务上下文解决“这次应该用其中哪些信息”。不能将两者合成一段不断变长的个人摘要。

6.2 时间和更正是核心能力

至少记录事情发生时间、系统获知时间和信息适用时间。例如客户去年有充电条件，不代表今天仍然有；昨日预计配件今天到，也不是今天已到货。

新旧说法冲突时，先看对象、适用范围、证据和时间。无法判断则同时呈现冲突并安排核实，不能总用“最后一句覆盖前一句”。车辆转手、联系撤回、转写更正，需要同步处理关联记录、摘要、检索索引和待执行任务。

关系要允许拆分恢复。误把两个人合并的损失往往高于暂时多留一条记录；模糊匹配适合产生候选，不适合直接跨品牌、跨门店共享隐私记录。

7. 核心架构：同一个引擎，按任务提供不同帮助



图中是逻辑职责，不要求每个方框独立部署。初期可以是一个业务服务配合异步处理任务，复用既有数据库、消息和工作流能力。

7.1 我们应当掌握的核心

- 汽车客户业务模型。**定义人与车、商机和服务事项、需要和承诺之间的关系，明确每一种信息由谁确认。
- 有证据的客户理解。**将对话和业务事件持续转成可修正、可追溯的记录，而不是让某个记忆组件直接决定客户事实。
- 任务上下文服务。**按照当前任务、对象、权限和时效，组织模型需要的资料；这是整个引擎对其他产品最重要的服务接口。
- 服务决策与行动记录。**让建议对应可办的任务，知道谁接手、何时完成、从哪里获得结果。
- 业务方法与评估。**知道哪些问题需要继续取证、什么帮助在什么条件下有效，以及如何证明效果。

7.2 适合复用的外围能力

ASR、媒体存储、消息通道、模型推理、关键词和向量检索、运行追踪、既有身份与客户主数据、CRM/DMS 正式接口，都应优先复用。不要因为引入智能体，再新建一套客户账号、呼叫中心、订单库和业务审批中心。

如果以后接不同厂商 DMS，替换的是适配接口；客户事项、证据引用、上下文格式和方法可以延续。反过来，若业务含义都藏在某个厂商的数据库字段或单个大模型提示词中，后续替换就会很难。

7.3 事件与定时任务怎样配合

新对话、客户改约、配件到货、预计延期、承诺到期等变化产生事件；系统将其关联到事项，合并重复信号，重新检查是否需要行动。

每个事项有主责人和可恢复状态：待核实、待顾问处理、待客户回复、待业务完成、待结果复核、结束。状态变化主要由明确规则和业务回执推动。模型负责理解和建议，不负责随意重写流程规则。

同一客户可以有多个事项，但对外联系需要协调。重试前核验已有回执，撤回授权后取消尚未发出的任务；已拨出的电话无法真正回滚，必须依赖事前校验和异常补救。

8. 怎样给智能体准备恰当的上下文

你的判断是对的：系统可能有上百种数据能力，但当前问题只需要其中几种。建议先建立业务任务目录，再允许智能体在目录范围内补充查询。

当前任务	主要取哪些信息	应得到的结果
首次接待与车型匹配	客户明确用途、条件、车型知识、当前可售配置	找出适合讨论的车型及还应问清的问题
试驾后成交推进	试驾反馈、关键顾虑、已比较车型、报价版本、下一步约定	给出有依据的跟进建议，准备正式预约或资料
交付延期主动安抚	订单与物流状态、已作承诺、客户明确时间要求	核实影响、明确责任人与下次更新时间
维修待件主动安抚	原问题、工单、配件进度、当前承诺、近期沟通	准备事实说明、安排内部处理和后续跟进
主动招揽	可信保养规则、历史检查、实际未完成项目、可预约资源	判断是否有合理服务需要，并提供可兑现安排
主动关怀	已有权益、适用功能、维修后反馈、客户明确偏好	兑现权益、解答问题、确认恢复情况
群体经营分析	合法可用的汇总特征、暴露范围、业务结果、门店资源	找共性障碍和可验证的方法，不输出个人隐私拼盘

一份上下文至少包含这些部分

- 本次任务及其对象：针对哪个客户、哪辆车或哪个商机、哪件事。
- 与任务有关的已知事实、原始证据位置和数据截止时间。
- 当前需求、顾虑、承诺和待确认问题。
- 适用车型知识、政策与门店可用资源。
- 联系选择、近期联系情况、允许执行的业务能力。

- 信息冲突、未知项和不能据此得出的结论。

组装顺序建议为：核验范围 → 读取当前业务状态 → 查相关事项与承诺 → 混合检索少量原始片段 → 补充有效知识 → 检查冲突和时效 → 交给智能体。资料不足时，智能体可以提出具体补查，不能自行扩大客户或门店范围。

这种“预先准备关键内容 + 按需查证”的方式与 Anthropic 的上下文工程建议一致。长上下文模型可以减少截断问题，但仍需要选择相关信息，不能代替身份、时间和权限管理。[上下文工程说明](#)

给工程团队的最小接口约定

能力	输入重点	输出或作用
<code>get_customer_context</code>	任务类型、业务对象、操作者、用途、时间范围	有来源、有限范围的任务信息
<code>get_evidence</code>	证据标识、必要片段范围	原话或正式记录，以及来源与版本
<code>propose_customer_update</code>	建议变化、依据、适用范围和期限	形成待规则检查或待确认的更新
<code>prepare_business_action</code>	行动类型、对象、所需参数	校验当前资源与权限，准备可执行请求
<code>execute_business_action</code>	已满足条件的请求和防重复标识	调用正式接口，返回业务系统回执
<code>record_feedback</code>	人工修正、原因、实际业务结果	更新事项与评估记录，供方法改进使用

这些接口是建议设计，不是已存在接口。调用权限由服务端根据当前身份和业务范围签发或校验，不能由模型输出的一段“允许执行”文字决定。

9. 智能体与顾问如何协作

建议按工作职责组织四类能力，而非为每个客户常驻一个机器人。

能力	何时工作	交给顾问什么
沟通理解	一段对话结束或出现新内容时	新增需求、关键顾虑、承诺与待确认项，并附原话
客户任务助手	顾问准备跟进，或事项发生变化时	任务摘要、需要核实的问题、可用方案和行动草稿
服务执行助手	条件满足且相关人员或客户已作选择后	完成预约、创建草稿、交接、催办，并读取执行结果
经验分析助手	周期复盘或出现共性问题时	哪一步经常受阻、可能原因、下一项应验证的方法

顾问不应每天维护一大批人工标签。系统从工作中的自然对话形成候选理解；顾问只需在关键位置确认或纠正。例如“不是预算超了，是家人还没体验过”，系统就应调整本次商机的问题判断，准备家庭试驾，而不是继续推优惠。

维护也要分工：补充有出处的低风险信息、提醒过期和提取重复问题可自动完成；身份合并、正式承诺、赔付、价格变更和方法发布交有权限的人或业务规则处理。智能体不能以“自我改进”为由修改生产权限和已批准政策。

智能外呼的位置

外呼是执行渠道之一。首期优先做顾问辅助和已授权的低风险确认，例如确认是否需要改约。复杂安抚以人员承接为主，系统负责查清事实、准备说明和跟进承诺。

真正接入外呼还需要呼叫线路、身份说明、打断处理、听不清后的澄清、转人工、联系时段、跨渠道频控、号码错误处置以及预约回执。现有 ASR 的三项输出不等于这些能力已经具备。

语音里的语气可作为进一步核实的线索，但“客户表达焦急”“顾问感觉客户不满”“模型猜测客户将投诉”必须分开。优先根据等待、失约、重复报修等可观察事件提前服务，不需要先给客户打上负面情绪标签。

10. 单个客户与客户群怎样相互帮助

这里需要区分三种学习，不能把它们都叫“大模型越用越聪明”。

第一种：记住这个客户的新事实

客户改变联系偏好，试驾后明确放弃某项配置，维修问题经复核已经解决。更新的是客户与事项记录，不需要为每个人重新训练模型。

第二种：发现一类客户经常遇到的问题

例如，多位考虑某车型的客户迟迟未决定，聊天中频繁出现“家人没体验过后排”，但系统过去都把原因归为价格。可以形成待验证的方法：对此明确顾虑的客户，优先提供家庭体验安排。

群体分析必须同时看车型、门店、供给、价格政策、活动和客户选择等条件。不能看到成交者常被邀请试驾，就断言试驾邀请造成了成交；也不能把其他客户的需求直接写进当前客户档案。

第三种：把经验证的方法用于下一位客户

方法应该告诉顾问“这类情境值得问清什么、提供什么选择”，并说明适用条件、证据范围与已知局限。到个体使用时，再根据该客户的实际表达核实。

推荐路径是：个体事件及结果 → 在允许范围内去标识化汇总 → 提出共性解释 → 人工审核方法 → 小范围比较 → 发布有版本的方法 → 在新客户身上核实适用性。

预测“谁可能成交”与判断“给谁这种帮助最有增量”是两个问题。首期先用可解释规则安排服务，数据和样本足够后再评估增量效果模型。训练和验证按时间、客户或门店隔离，避免把同一事件两次出现当成独立验证。

任何成功方法都要保留失败、拒绝、无需要和不应打扰的样本。只学习回店成交案例，会逐渐把系统推向过度触达。

11. 最新模型：按任务试，不按排行榜定架构

以下是截至核查日期值得进入评测的模型及方向。模型卡能证明公开能力说明和发布状态，不能证明中文汽车对话、工具调用和门店部署效果。

候选	已核实信息	在本项目中的建议位置
Qwen3.8-27B（通义开放权重模型）	官方模型卡提供 27B 权重，支持图文及视频输入，Apache 2.0 许可	私有化候选；先试需求和承诺提取、结构化输出、客户摘要与证据检索后的回答
DeepSeek-V4.1-Flash（深度求索开放权重模型）	官方模型卡为多模态混合专家模型，MIT 许可；骨干总参数 552B	复杂理解候选。激活参数较少不代表本地部署资源很小，先核验服务方式、硬件和实际成本
GPT-6 Astra（OpenAI 模型）	官方系统卡已发布，并描述能力与安全评估范围	在允许的部署与数据范围内，作为复杂任务对照之一，不据名称宣称汽车业务最优
Claude Fable 5.1（Anthropic 模型）	2026 年 9 月公告说明普遍可用；Mythos 5.1 属受限访问计划	复杂任务对照候选；本项目不以受限的 Mythos 版本作为依赖
Gemini 3.8 Flash（Google 模型）	2026 年 9 月正式公告提供开发者访问路径	多模态与任务处理对照候选；具体区域、数据及商业条件另行核验
现有 ASR 平台	用户已确认逐句文本、说话人、时间戳	首期直接复用；如专有词或分离准确度不足，再做有针对性的改进

模型来源：Qwen 模型卡、DeepSeek 模型卡、GPT-6 Astra 系统卡、Claude 发布公告、Gemini 发布公告。

我的推荐是：先固定一组汽车业务评测样本，选择一个符合部署要求的中文模型作为基线，再加一个获准使用的强模型做对照。只有复杂任务的质量差距值得成本时，才启用分级调用。不要一开始同时运维五家模型，也不要因为模型升级就重写客户数据结构。

首期不因追求新模型更换 ASR；未来实时语音、图片与视频理解另作专项验证。处理未脱敏客户信息前，确认部署区域、数据保留及处理范围。境外产品能够调用，并不意味着这些数据可以直接传过去。

12. GitHub 开源项目怎么选

本轮结论分为 ADOPT（进入采用候选）、NARROW_BORROW（有限借鉴或隔离试验）、NO_GO（本期不采用为核心）。这些是选型建议，未安装、未运行性能测试。许可证按当前公开仓库核查，正式依赖时仍需锁定版本并复核依赖项。

项目	核查到的能力与边界	对本项目的裁决
pgvector	PostgreSQL 中的向量检索；PostgreSQL 许可	ADOPT ：若现有数据库合适，用关系查询、关键词与向量检索建立首个基线；无需先上图数据库
LangGraph	MIT；提供有状态智能体编排，检查点与跨任务存储职责有区分	ADOPT ：既有平台缺少相应能力时，用于少量任务编排；不能替代客户业务数据模型和正式流程状态
Temporal	MIT；支持可恢复工作流和失败重试	条件采用 ：确有跨天等待、定时、恢复需求且现有工作流不足时引入；不与另一套业务调度重复管理同一事项
Langfuse	追踪与评估；普通目录为 MIT，指定企业目录另有许可	ADOPT ：无既有等效能力时评估自托管；记录模型、工具、成本和结果，日志先处理敏感信息
Graphiti	Apache 2.0；支持实体、关系、证据片段和时间。源码的关系对象包含出处及有效时间字段	NARROW_BORROW ：试验跨事项、跨时间的关系检索。作为可重建索引，不将模型抽出的关系直接当正式客户事实
Mem0	Apache 2.0；提供记忆提取与检索。当前源码明确部分时间参数仅平台版支持；默认推理可决定记忆增删改	NARROW_BORROW ：做对话记忆基线比较。不要把商业平台宣传等同开源版，也不直接让其维护订单、承诺和身份事实
Memobase	Apache 2.0；侧重客户资料与事件记忆、批量处理	NARROW_BORROW ：借鉴资料与事件分开、按用户聚合处理。默认个人画像字段需按汽车业务收窄，不能照搬宽泛心理画像
Acontext	Apache 2.0；将智能体运行经验存成可读的技能文件	NARROW_BORROW ：借鉴方法候选的整理；跨客户共享前去除私人信息，并用业务结果验证，不能凭智能体自报成功发布方法
Letta / Letta Code	当前主项目指向 Letta Code，Apache 2.0；侧重有持续记忆、可修改自身上下文的智能体运行环境	NARROW_BORROW ：研究记忆分层与按需读取；不为每位客户部署一个常驻、可自改规则的智能体
Microsoft GraphRAG	MIT；当前仓库明确声明主要进入维护阶段，并说明是研究演示项目	NO_GO ：本期不作为新生产客户引擎的核心依赖；可参考群体主题分析方法

重点源码依据：Graphiti 关系与时间字段、Mem0 当前记忆接口、Langfuse 许可证范围、LangGraph 状态与记忆说明、Letta 当前仓库说明。

最小推荐组合

复用现有 ASR、业务数据库与业务工作流；增加客户理解服务、任务上下文服务、必要的关键词及向量检索、模型适配和可追溯评估。若缺少智能体任务编排，选 LangGraph 做受控任务流程。Graphiti 在隔离样本上证明检索收益以后再加入。

不建议把 Graphiti、Mem0、Memobase、Letta、Acontext 同时装成一个“大记忆平台”。它们回答的问题有重叠，全上会制造多份客户记忆、多个更新规则和不清楚的事实归属。

13. 首期两个场景怎样做

场景 A：试驾后，把真正阻碍成交的问题接住

适用对象：已完成试驾、允许继续联系、存在明确待确认问题的商机。不要把所有未成交客户都当作需要再次打电话。

输入：试驾沟通及后续聊天、试驾和商机记录、版本明确的车型与配置资料、报价及活动、顾问的下一步约定。

流程：提取明确体验与顾虑 → 与车型事实对照 → 找出仍缺的信息 → 提供下一次最有帮助的沟通或安排 → 顾问确认 → 通过正式系统执行 → 读取后续业务与客户反馈。

例如，客户明确觉得第二排空间尚未与家人确认，系统就准备家庭试驾和相关空间说明；客户对用车成本有疑问，则准备有假设、有车型依据的成本解释。系统不应统一输出“再给一个优惠”。

顾问实际获得：一份带依据的跟进准备、两三个最应问清的问题、合适的资料或预约草稿，以及不适合本次跟进的动作提示。

怎么验收：核心顾虑是否提取正确；顾问完成一次有效跟进的总耗时是否下降；约定是否兑现；有效再访、试驾或成交相对合理基线是否改善；客户拒绝与投诉是否增加。成交周期未结束时先报告过程结果，不能提前宣布成交提升。

场景 B：维修待件或延期，主动安抚并跟进承诺

适用对象：有真实工单、配件或交付进度变化，并存在等待、已约定更新时间或明显承诺风险的客户。

输入：工单与待件状态、顾问上次承诺、客户明确的使用需要、最近的联系记录、门店实际可提供的处理方式。

流程：发现承诺风险 → 找责任人核实当前阻断点 → 准备明确说明与可兑现的下一步 → 顾问联系或经确认的常规通知 → 记录新的约定 → 到期提醒、未完成升级 → 维修完成后复核问题。

“主动安抚”的关键是让客户看到事情有人负责、进度可以知道。配件到货时间未知，就说明尚未确认，并给出下一次更新时间；不能用温和语气编造一个日期，也不能擅自承诺补偿。

顾问实际获得：谁需要今天被接住、之前说过什么、现在核实到什么、哪些话不能承诺、下一步由谁负责。

怎么验收：承诺兑现、有效进度更新、重复追问、事项解决、再次报修或投诉；同时记录客户是否认为帮助有用。未再投诉不能直接当作问题已解决。

两个场景共同验证的东西

同一份对话能否正确关联到人和事项；同一个客户是否能按不同任务取得不同上下文；新事实能否纠正旧摘要；顾问能否快速确认或纠正；正式行动是否有回执；结果能否支持下一次服务和方法复盘。

只要这些公共能力得到验证，后来增加交付延期、合理保养招揽和权益关怀，就不需要每次重建一套客户资料与智能体。

14. 实施顺序与资源安排

建议用约 10—12 周作为资源讨论的初始窗口，前提是两个场景所需接口、样本和业务人员及时到位。这是规划范围，不是尚未评估就承诺的交付工期。

阶段	建议窗口	交付及进入下一阶段的条件
对齐事实与样本	第 1—2 周	明确两条业务流程、数据权威和使用范围；建立顾问认可的样本与当前基线
建立理解与上下文	第 3—5 周	接入 ASR 与必要业务记录，跑通来源、关联、需求与承诺提取；能够发现未知与冲突
顾问协同试点	第 6—8 周	嵌入实际工作位置；顾问确认后执行，具备回执、更正、交接与恢复路径
复盘与有限扩展	第 9—12 周	按场景评估效果及成本，修正失败方法；只放开证据充分的低风险自动动作

首批可以选择少量有代表性的门店，兼顾销售与售后、直营与分销条件；具体门店数由业务量和接口条件确定。不能只选最配合、数据最干净的店，再据此承诺全网效果。

需要的职责包括：一名能决定场景取舍的业务产品负责人；销售与售后各一名业务负责人；数据及接口工程、应用及前端、模型与评估、测试及运行支持。可以由现有团队兼任，先确定职责与投入，不先按一堆技术模块扩编。

预算应分成共同引擎、两个场景适配、接口与部署、试点运行四部分。核算对象是“完成一项有效业务任务”的总成本，包括模型调用、检索与存储、人工核对、失败重试和运行维护；不能只比较一次模型调用的价格。

为控制投入，媒体按用途分层保存；对话按新增片段处理；优先预计算稳定信息，执行前重查易变状态；简单事件用规则处理，复杂解释再调用模型。不要给全部客户常驻运行智能体。

15. 验证什么，才足以说客户引擎有用

15.1 先建立真实的对照

建议从约 300—500 段经授权、必要处理的互动开始建立探索样本，包含两个场景以及否定、条件表达、说话人错位、旧信息更正、拒绝联系和业务失败案例。这只是探索规模建议，不代表已经达到统计证明所需样本量。

在同一批任务、相同模型和相近预算下比较三种方式：现有顾问工作方式；把完整历史交给模型；使用本文的任务上下文与证据机制。这样才能知道收益来自模型本身、数据整理，还是任务设计。

如要评估 Graphiti 等组件，再把它与已有关系查询和混合检索基线比较。不能拿不同模型、不同样本和不同预算下的结果作组件胜负判断。

15.2 三层验收

层次	关键问题	证据
理解质量	谁说、说了什么、是否有条件、还缺什么，是否判断正确？	原文定位、业务人员标注、错误类型、分场景准确性
任务完成	顾问和系统是否完成真实工作，失败后是否能恢复？	正式预约或工单回执、人工修改、交接、重复操作及异常记录
业务效果	是否提供了更合适的帮助，且成本和打扰可接受？	完整任务耗时、客户反馈、承诺兑现、有效推进及可归因的业务变化

对身份错误、越权访问、无授权联系、擅自价格或赔付承诺设严格阻断条件。对需求提取、遗漏、证据召回、响应时间和成本，按场景基线设定验收线，并分类型披露。样本中零次严重错误不等于真实环境风险为零。

安全通知、必要服务和既有权益不能为了试验被延迟或取消。适合对照的普通商业服务，可采用门店或时间分组，并控制促销、库存和顾问差异；存在相互影响时避免简单逐客户随机比较。

基盘流失、品牌连接和长期复购需要更长观察周期。首期可以验证相关过程是否改善，但不要把三个月的通话量、消息阅读或预约变化直接换算成长期留存与收入。

16. 与未来 AI 原生 DMS 的关系

这个项目可以成为未来 AI 原生 DMS 的第 0 步，但“能延续”要通过设计和验证来保证。

可提前验证五件事：客户与业务对象是否贯通；对话能否改变工作流程；智能体能否取得恰当上下文；正式业务接口能否被可靠调用；结果能否反过来更新客户理解与方法。

未来 DMS 更换或升级后，客户引擎继续提供身份关系、事项、证据、上下文和方法；DMS 提供销售、预约、维修、配件、索赔等正式交易能力。需要变化的是接口适配和部分业务规则，不应推倒客户理解体系。

这种路径允许先在两个小场景中验证 AI 驱动工作，而不必先把全套 DMS 重写。不过，单个试点成功仍不能证明全网 DMS 已完成原生化，尤其不能替代财务、库存、索赔、迁移和复杂权限的独立验证。

如果与 Workshop（经营分析产品）协作，建议复用其“事实—问题—解释—行动—结果复核”的方法，用于群体与门店经营分析。本文没有核验它当前的运行代码，不将既有方法等同于客户引擎已经实现。

17. 现在可以确定与仍需核验的内容

现在可以确定：产品围绕客户当前任务组织；复用现有 ASR；销售和服务共用核心；采用有出处、可修正的理解；业务行动以正式回执确认；群体方法与个人事实分开；首期选择两个真实场景。

开工前需核验：ASR 的客户关联和错误情况；CRM/DMS 的具体接口及负责团队；顾问聊天数据可用范围；客户和车辆关系如何维护；允许的模型部署与数据处理方式；试点门店、业务量和现有 workflows。以上不妨碍本方案形成，但决定实际实施和报价。

对外可以这样表达：

我们希望把客户每次沟通和每次服务积累下来的信息，用到下一次需要帮助的时候。销售顾问知道客户真正还在犹豫什么，服务顾问知道客户正在等什么、我们还答应过什么。系统把相关信息准备好，能办的事情接到业务系统里，再确认结果。这样客户换了渠道、换了顾问，也能得到连续的服务。

附录 A：关键来源与适用范围

下列来源均在本轮核查。网页及默认分支会变化，正式技术采购和依赖选型应固定版本。供应商公告用于证明公开方向与能力说明；仓库和源码用于判断组件职责；均不替代长城数据、接口与业务效果验证。

编号	来源	本文使用范围
S01	Salesforce 非结构化数据说明	对话、文件等与客户数据共同使用
S02	Salesforce 语义层文章	业务含义与智能体上下文
S03	Microsoft, 2026-07-15	客户资料支撑销售、服务与旅程
S04	Microsoft 发布计划	预览与正式可用状态区别
S05	Twilio 对话记忆	按问题召回客户记忆，结合产品和政策知识
S06	Decagon 用户记忆	跨会话延续与既有记录系统配合
S07	Decagon Campaign Composer, 2026-09-02	跨天主动服务、回拨和退出控制
S08	Tekion, 2026-06-16	汽车零售智能体与业务上下文
S09	火山引擎销售助手	国内汽车对话辅助与邀约实践
S10	Anthropic 上下文工程	选择相关信息、按需取证
S11	Qwen3.8-27B 模型卡	权重、模态、许可及候选范围
S12	DeepSeek-V4.1-Flash 模型卡	权重、模型规模与许可
S13	GPT-6 Astra 系统卡	官方能力与评估说明
S14	Claude Fable / Mythos 5.1	公开可用与受限访问区别
S15	Gemini 3.8 Flash	模型发布及开发者访问说明
S16	Graphiti 仓库	时间关系、检索、开源与商业产品边界
S17	Graphiti 源码	关系对象的证据与时间字段

编号	来源	本文使用范围
S18	Mem0 仓库	开源记忆能力及商业说明边界
S19	Mem0 源码	自动提取、增删改、平台限定时间参数
S20	Memobase	用户资料与事件记忆
S21	Acontext	从运行记录整理技能与方法
S22	Letta 当前说明、Letta Code	项目路径变化及记忆运行方式
S23	LangGraph、持久化说明	状态恢复与长期存储的职责
S24	Temporal	可恢复 workflow
S25	pgvector	关系数据库中的向量检索
S26	Langfuse、许可证	运行观察、评估与企业许可边界
S27	Microsoft GraphRAG	当前维护状态与研究项目定位
S28	个人信息保护法	处理基础、公开信息与自动化营销边界

附录 B：一个任务上下文示例

以下是虚构数据，用于说明接口职责，不代表现有系统字段。

```
{
  "task_type": "repair_delay_followup",
  "customer_ref": "customer_demo_01",
  "case_ref": "service_case_demo_01",
  "purpose": "维修进度服务",
  "business_state": {
    "status": "等待配件",
    "source_ref": "dms_workorder_demo_v3",
    "as_of": "2026-09-20T14:00:00+08:00"
  },
  "commitments": [{
    "content": "今天下午再次更新进度",
    "owner_ref": "advisor_demo_01",
    "evidence_ref": "conversation_demo_02:utterance_18",
    "confirmation": "顾问表达，已确认记录"
  }],
  "unknowns": ["配件到货时间尚未确认"],
  "contact_preferences": ["客户要求通过已核实渠道留言"],
  "next_step": "先向配件负责人核实，再由顾问说明当前进度",
  "cannot_claim": ["不能承诺具体完工日", "不能自行承诺补偿"],
  "available_tools": ["query_parts_status", "prepare_followup_task"],
  "permission_note": "实际查询和执行仍由服务端实时校验"
}
```

这种上下文同时告诉智能体已知什么、未知什么、下一步能查什么。它不需要包含客户全部历史，更不需要包含与本次维修无关的个人信息。