小豆白皮书

Personal Agent：从个人 AI 助理到个人智能操作系统

Xiaodou: The Personal Agent Era

版本：V0.2.0
研究日期：2026 年 11 月 24 日
文档属性：独立研究白皮书
作者：连誠（X：@0xYuanwz）
研究方向：Personal Agent × AI × Personal OS
开放许可建议：CC BY 4.0

> **研究声明**
>
> 本白皮书为独立研究成果，不代表字节跳动、豆包或小豆官方立场，亦未获得其授权或背书。
>
> 本文基于截至 2026 年 11 月 24 日可核查的公开资料、官方页面、产品演示转述和媒体报道进行分析。对于未公开的内部技术架构、模型实现、数据处理机制、安全策略、用户量、营收和具体定价，不作事实性断言，也不编造数据。
>
> 文中提出的系统架构、权限模型、评测指标、商业模式和演进路径，均属于独立研究建议或逻辑推演，不构成对小豆内部实现的还原，也不构成投资建议或产品承诺。
>
> 凡属公开事实，尽量注明出处；凡属推演、假设和判断，会明确标出。后续若有更完整的官方披露或可复现观测，应以新的公开证据修正本文。

目录

· 摘要：先把判断说清楚
· 第一章：为什么 2026 年 11 月 24 日值得当成一个产业信号
· 第二章：Personal Agent 时代的到来
· 第三章：小豆的产品定位与能力分析
· 第四章：字节生态与小豆的关系
· 第五章：生态飞轮作为复利资产
· 第六章：Personal Agent 的参考技术架构
· 第七章：个人记忆系统
· 第八章：连接器与外部工具
· 第九章：自主执行与安全治理
· 第十章：Personal Agent 的评测体系
· 第十一章：小豆、Muse 与公开可见的行业布局

· 第十二章：商业模式与定价权推演
· 第十三章：风险、误判与何种证据会推翻本文判断
· 第十四章：给从业者与观察者的可跟踪信号
· 第十五章：从 Personal Agent 到 Personal OS
· 第十六章：未来研究议程
· 第十七章：开源协作计划
· 结语：AI 不只是工具，而是个人的数字代理
· 参考资料
· 版本说明

摘要：先把判断说清楚

2026 年 11 月 24 日，豆包旗下个人 AI 助理「小豆」开启邀请制公测。

表面上看，这很像一次熟悉的产品更新：有官方介绍，有演示场景，有邀请码，也有「未来将逐步在豆包 App 上线」的说明。若只停在这个层面，小豆不过是又一个会聊天、会提醒的 AI 助手。

这个理解并不完全错，但不充分。

过去几年，大语言模型把人与信息的交互，从「打开网页、输入关键词」，推进到了「直接提问、得到答案」。可用户与 AI 的关系，大多仍建立在对话之上：问完即止，事随会话结束。真正难的事情，并没有因此自动完成。一个人的工作日历、消费决策、出行安排、健康记录和长期目标，仍然分散在不同应用里，需要用户自己切换、确认、跟踪。

小豆公开展示的方向，不是把问答做得更像人，而是把产品重心移向另一件事：**记住用户交代过的背景，连接用户授权的应用，在对话结束之后继续跟进，并在需要时主动提醒或代为处理**。

所以，本文的核心判断是：

未来个人 AI 的竞争，不再只是模型智力的竞争，而是个人上下文、生态连接、持续执行能力和用户信任的综合竞争。

如果这个判断大体成立，那么 2026 年 11 月 24 日的意义，就不只是豆包上线了一款新应用。它更像一个产业信号：Personal Agent 正在从概念演示，进入可被用户真实授权、持续使用的产品窗口。而 Personal Agent 一旦站稳，就可能逐渐成为人与数字世界之间的新交互层，并在更长周期里演化为 Personal OS（个人智能操作系统）。

这里必须立刻补一句边界。Personal OS 是研究概念，并不意味着小豆当前已经具备操作系统级能力。字节庞大的产品版图，也不等于这些产品都已与小豆打通。媒体提到的部分应用连接，还存在版本、权限和稳定性上的不确定性。对未公开的内部实现，本文不作事实还原。

本文以小豆为研究对象，结合已核对的公开资料，系统分析个人智能体的产品定义、参考技术架构、生态连接、记忆机制、任务执行、权限治理、行业比较、商业模式、风险条件和可跟踪信号。能确定的写成事实，

不能确定的写成推演。读者若要验证或反驳本文，也应看后续出现了什么新的公开证据，而不是看一句更绝对的结论。

第一章：为什么 2026 年 11 月 24 日值得当成一个产业信号

我们先看当天已经能核对的事实，而不是先看想象。

2026 年 11 月 24 日，豆包官方宣布，个人 AI 助理产品「小豆」正式开启公测。官方表述是：「一个懂你、还能主动帮你办事的豆包。」公测采用邀请制，用户凭邀请码体验，并获赠一定量免费额度；邀请好友还有机会获得更多额度。官方同时说明，小豆上的能力与体验，未来将逐步在豆包 App 上线。上述内容可见于豆包官方公开信息，并由第一财经、北京日报、每日经济新闻、北京商报等媒体报道。小豆官方页面位于 https://www.doubao.com/xiaodou/ 。

第一财经等报道还转述了更具体的产品方向：小豆是豆包在个人智能体（Personal Agent）方向推出的探索性产品；在问答和信息检索的基础上，更侧重主动服务、基于记忆更懂用户，以及任务执行。用户交代事项后，小豆可在后台持续跟进，按时提醒或代为处理。

北京日报、新黄河以及新华财经等报道，进一步转述了官方演示中的场景：根据猫粮使用情况估算余量并在确认后补货，识别日程冲突，提示商品降价或团购券到期，代打电话询问餐厅座位、餐费和停车费，以及把「3 个月减 8 公斤」这类长期目标拆成日常记录和提醒。报道同时写明：目前代打电话仅覆盖餐厅、酒店、门店等生活服务场景，不支持境外和私人电话；公测初期名额有限。

这些都是公开信息。它们说明了产品想被怎样理解，并不自动等于全部功能已向所有用户稳定开放，也不等于演示效果已经等于大规模真实任务完成率。

1.1 常见误读是什么？

对这一天，至少有三种常见读法。

第一种，把它看成常规「上新」。邀请制、免费额度、演示视频，确实很像过去两年 AI 应用的发布模板。若只这样看，观察重点就会停在功能清单，而不是交互层有没有变化。

第二种，把它看成对 Meta Muse 的快速对标。这个联想并非全无根据。Meta 于 2026 年 9 月 8 日发布 Muse，北京日报、新黄河等报道也把小豆放在同一产品方向上讨论。可如果因此得出「只是跟风」或「已经可以排名」，就把公开的产品方向比较，误写成了性能和完成度比较。后一种比较，当前证据不够。

第三种，把它看成字节生态已经全部打通、小豆已经接近个人操作系统。字节的确拥有丰富的产品矩阵，但这是版图事实，不是连通事实。官方页面目前明确提到的连接方向，是抖音、飞书、日历、健康、地理位置等应用；更完整的打通范围、稳定性和权限深度，仍需以实际版本和官方说明为准。

1.2 为什么「只是产品更新」这个逻辑并不成立

表面原因确实存在：小豆仍在邀请制公测，能力还在放量，官方也说相关体验将逐步进入豆包主端。可真正值得关注的，不是发布形式，而是交付物变了没有。

传统 Chatbot 的交付物是答案。Personal Agent 的交付物是任务结果。

这句话怎么理解？用户如果说「帮我关注下周会议，有冲突提前告诉我」，只给出管理建议，仍然停留在助手阶段；若系统能在授权范围内读取日历、识别冲突、主动提醒、提出调整建议，并在确认后修改日程、更新任务状态，这才构成任务闭环。小豆公开展示的记忆、连接、主动提醒和后台跟进，指向的正是后一种交付。

从产业角度看，一天的公测之所以可能成为信号，而不是普通更新，主要有四个原因。

第一，它把「主动办事」放在了产品定义的中心，而不是附加功能。官方页面写的是：结合聊过的事和已连接的应用，替用户留意变化并主动提醒；一句话交代后按节奏执行，到点提醒，随时可改可停；授权后使用抖音、飞书、日历、健康、地理位置等应用，且「连哪个、断哪个，都由你决定」。

第二，它出现在一个已经有同类产品公开亮相的窗口里。2026 年 9 月，Meta 发布 Muse，OpenAI 发布持续在线的 dots；阿里巴巴随后在业绩公告中披露，10 月推出了内嵌于千问 App 的云端个人智能体「千问 Go!」。当多家公司不再只强调「更会回答」，而是同时强调「持续执行」，这个品类就不再只是单点创新。

第三，它可能改变的是交互层，而不只是应用层。搜索引擎改变了人获取信息的方式，移动操作系统改变了人调用服务的方式，超级应用则把多种服务收进同一入口。这些类比都不能直接等同于 Personal Agent，但它们提示了一件相似的事：一旦用户习惯通过一个代理层表达目标，而不是自己打开每一个软件，入口、数据和信任的分配就会发生变化。

第四，邀请制公测本身也在测试一种更关键的东西：用户愿不愿意授权。没有授权，记忆和连接都只是演示；有了授权，才谈得上个人上下文和任务闭环。因此，这一天的观察价值，不在于口号有多新，而在于一个高授权产品开始进入真实使用环境。

1.3 这一章的判断

1. 小豆于 2026 年 11 月 24 日开启邀请制公测，是已核对的公开事实，不宜夸大成「已经全面普及」或「已经验证成功」。
2. 把这一天只看成功能更新，容易低估产品定义的变化：交付物正在从答案转向任务结果。
3. 把它直接写成对 Muse 的性能赶超，或者写成 Personal OS 已经落地，则超出了当前公开证据。
4. 更稳妥的定位是：这是 Personal Agent 进入可被授权、可被持续使用的产品窗口的一个明确信号。窗口已经出现，并不等于窗口期内谁已经赢。

第二章：Personal Agent 时代的到来

2.1 从 Chatbot 到 Agent

过去，软件是人类完成任务的工具。用户需要打开不同应用，理解软件功能，按既定流程输入信息，最终自己把事情做完。软件很强，但协调成本在用户身上。

ChatGPT 等大语言模型出现后，人与信息的交互方式明显改变了。用户不必先学习每个界面，也能用自然语言提问并得到内容。可早期 Chatbot 仍然主要以对话为中心。用户必须主动提出问题，模型根据当前上下文给出回答。会话一结束，任务往往也就结束了。

随着 Agent 技术发展，AI 开始能够调用工具、拆解目标、执行多步骤任务，并在一定程度上处理过程中的异常。变化的关键，不是话变得更像人，而是系统开始进入原来由用户亲手完成的操作流程。

这一变化可以概括为四个阶段：

阶段	产品形态	核心能力
第一阶段	Chatbot	问答与内容生成
第二阶段	AI Assistant	多轮理解与任务辅助
第三阶段	Task Agent	工具调用与自主执行
第四阶段	Personal Agent	个人记忆、持续服务与跨应用调度

这四个阶段并不是严格的技术代际，而是对 AI 产品形态演进的分析模型。现实中，一款产品可能同时具备相邻阶段的特征。小豆公开展示的能力，更接近从 Task Agent 向 Personal Agent 过渡，而不是已经完成了这一跃迁。

2.2 什么是 Personal Agent?

本文将 Personal Agent 定义为：

一种以个人为中心，能够持续理解用户背景、维护个人记忆、连接外部应用和数据，在用户授权范围内规划并执行任务的智能系统。

与传统 AI 助手相比，Personal Agent 需要具备七项核心能力：

1. Personal Context：个人上下文理解。
2. Long-term Memory：长期记忆。
3. Connectivity：外部应用连接。
4. Planning：任务规划。
5. Execution：任务执行。
6. Proactivity：主动服务。
7. Governance：权限、安全与审计。

其中，记忆和执行是两个关键变量。

没有记忆，AI 无法持续理解用户，每次对话都可能重新开始。没有执行，AI 就无法真正替用户把事情做完，最多提供建议。若再缺少治理，执行能力越大，外部副作用也可能越大。

所以，Personal Agent 不是「更会聊天的机器人」，而是一套把理解、记忆、连接、规划和受控执行打通的个人代理机制。

2.3 为什么个人智能体现在受到关注?

主要有三个原因。

第一，大模型的能力开始从生成内容向调用工具扩展。

模型不再只是预测下一段文本，而是能够根据任务选择工具、调用外部服务，并根据返回结果继续推理。生成答案和完成任务，所需的系统能力并不相同。后者要求工具描述、权限判断、结果验证和失败恢复，而不能只看回复是否流畅。

第二，数字生活已经高度应用化，个人数据却高度分散。

一个人的日常生活，往往分散在不同软件中：

- 工作在办公软件中；
- 消费在电商平台中；
- 社交在社交平台中；
- 健康在健康应用中；
- 出行在地图和交通平台中；
- 知识在文档、网盘和笔记中。

应用越全，用户自己完成一件跨应用事务的协调成本就越高。Personal Agent 若能在授权范围内调用这些服务，解决的就不是「找不到信息」，而是「事情在多个入口之间被切断」。

第三，AI 开始具备持续运行的产品形态。

AI 不必在用户关闭对话后立即停止工作。它可以通过定时任务、事件触发和后台执行继续推进事项。Muse 官方介绍写过，对于需要更长时间的任务，Muse 会在用户关闭应用后继续工作，并在情况变化或需要批准时再回来；小豆的公开报道和官方页面，也强调后台跟进、到点提醒和任务可改可停。持续运行一旦成立，AI 才从工具形态，更多具备了个人助理的产品属性。

这三件事叠在一起，才让 Personal Agent 在 2026 年不再只是研究议题。模型开始能调用工具，生活已经足够应用化，产品也开始尝试在对话之外继续工作。缺任何一条，这个品类都很难从演示走入日常。

2.4 这一章的判断

1. Personal Agent 的定义关键，是以个人为中心的持续理解、授权连接和受控执行，而不是对话本身是否自然。
2. 七项能力需要一起看。只强调模型智力，或只强调连接数量，都可能误判产品成色。
3. 记忆、执行和治理是三个最不该被简化的变量：前两者决定价值，后者决定价值能否被信任。

第三章：小豆的产品定位与能力分析

3.1 小豆是什么？

小豆是豆包在 Personal Agent 方向推出的个人 AI 助理产品。

根据官方页面和公开报道，小豆的产品方向包括：

- 个人记忆；
- 第三方应用连接；
- 个人日程和事项管理；

· 主动提醒；
· 后台任务跟进；
· 生活服务调用；
· 长期目标管理。

与传统 Chatbot 相比，小豆更强调基于用户上下文的主动服务。官方页面把能力概括为三类：主动消息、猜你想要、帮你跟进。其中，主动消息强调结合聊过的事和已连接的应用，留意变化并在需要时提醒；帮你跟进则举了每日简报、降价提醒、运动目标等例子，并说明用户可以按自己的节奏执行、到点提醒，随时可改可停。

还需要分清产品形态。公开信息显示，小豆在公测阶段以邀请制方式开放体验，并拥有独立产品入口；同时，豆包方面表示，相关能力和体验未来将逐步在豆包 App 上线。这意味着，小豆既是一次独立产品探索，也可能是豆包主端能力演进的前置实验。至于何时全面放开、主端与独立端如何协同，应以后续官方说明为准，本文不作时间表预测。

3.2 产品能力拆解

能力	公开展示或报道的场景	对应价值
个人记忆	记住用户此前交代的信息，例如猫粮拆封与用量	减少重复沟通，使后续判断有连续背景
日程管理	识别和提醒日程冲突	提升时间管理效率
主动提醒	商品降价、团购券到期	从被动问答转向主动服务
任务跟进	按约定时间持续推进事项，例如指定时间前发送资料	管理任务生命周期
生活服务	代为询问餐厅座位、餐费和停车费用	调用外部服务，把意图变成一次真实询价或预订准备
消费管理	根据猫粮使用情况提醒补货，确认后下单并设置下次提醒	连接记忆与消费行为
长期目标	将「3 个月减 8 公斤」等目标拆解为日常热量记录和步数提醒	提供持续陪伴，而不是一次性建议

以上是公开资料中的产品表现，不代表所有功能均已向全部用户开放，也不代表演示场景已经具备稳定的大规模成功率。对「确认后下单」「按时发送资料」这类具有外部副作用的动作，尤其需要看真实权限、确认机制和失败处理，而不能只看演示叙述。

3.3 小豆最值得关注的产品变化

传统 AI 的主要交付物是答案。

Personal Agent 的主要交付物则是任务结果。

例如，用户说：

「帮我关注一下下周的会议安排，如果有冲突提前告诉我。」

传统 AI 可能会告诉用户如何管理日程。

Personal Agent 则需要：

1. 读取授权的日历数据；
2. 识别会议时间；
3. 判断是否存在冲突；
4. 在冲突发生时主动提醒；
5. 根据用户指令提出调整建议；
6. 经确认后修改日程；
7. 更新任务状态。

这才构成一个真正的任务闭环。

大家试想一下，如果把这个逻辑推到生活服务、消费补货和长期目标上，差异会更明显。会回答「猫粮大概还能吃多久」，和会根据已有记忆估算余量、主动询问、经确认后下单并设置下一次提醒，不是同一类产品。前者节省的是搜索时间，后者切入的是事务本身。

小豆最值得关注的，因此不是某一句演示文案，而是它有没有把「理解用户本人」和「持续完成任务」当成产品主线。如果这条主线能在真实使用中成立，竞争变量就会从模型口才，转向上下文质量、连接可靠性和受控执行。

3.4 这一章的判断

1. 就公开信息而言，小豆的产品定位已经明确偏向 Personal Agent，而不是通用聊天机器人的一次改版。
2. 官方演示和媒体转述提供了记忆、提醒、跟进、生活服务调用和长期目标拆解等场景，但场景展示不等于能力已完全兑现。
3. 观察小豆，应优先看任务闭环是否真的闭合：有没有授权、有没有执行、有没有确认、有没有验证、失败后能不能停和改。

第四章：字节生态与小豆的关系

小豆的潜在优势之一，是其背后存在丰富的产品生态。这句话常常被直接写成「字节已经赢了」。这个跳跃过快。

生态要先被看成一张产品版图，再被看成一套可能的连接能力。前者是分类，后者才是 Personal Agent 真正用得上的东西。两者不能混为一谈。

4.1 字节生态的产品版图

从产品类型来看，可以将字节生态划分为五个层次。需要强调：以下是字节整体产品生态的功能分类，主要依据字节跳动官方 Products & Services 页面及公开可核对的产品品牌，不代表这些产品都已经与小豆直接打通。

第一层：个人数据与应用连接

官方页面明确提到的连接方向包括：

- 抖音
- 飞书
- 日历
- 健康
- 地理位置等应用

官方示例包括：从抖音收藏里挑选聚餐餐厅，把飞书文档整理成会前清单，按日程提醒出门。官方同时强调，连哪个、断哪个，都由用户决定。

《连接器服务使用须知》还从规则层面提到，连接器可能帮助用户连接设备功能（如日历、联系人、提醒事项、位置、相册、短信、通知、健康数据等），以及外部产品服务。外部服务既包括关联方提供的内容、办公、电商、本地生活、音乐或资讯服务，例如飞书、抖音等；也包括第三方提供的邮箱、网盘、文档、出行、地图、研发协作等服务。须知同时写明：实际可用连接器、支持的功能及适用终端，以连接器详情页、授权页和使用时的页面提示为准。

原稿和部分讨论中还提到高德、Wind、自如、百度网盘等连接方向。就本文写作时已核对的资料看，这些具体品牌名称尚未出现在小豆官方产品页的举例清单中；它们更接近「类别上可能被连接的出行、金融、居住、网盘服务」，或来自媒体与体验讨论中的提及。因此，本文将其视为待核实的连接方向，而不是已对全部用户稳定开放的既成事实。是否可连、能读能写到哪一步、是否会随版本变化，必须以实际产品与官方说明为准。

第二层：内容消费生态

产品	对应领域
抖音	短视频与内容消费
今日头条	新闻与信息分发
西瓜视频	中长视频
红果短剧	短剧消费
番茄小说	数字阅读
汽水音乐	音乐消费

这些产品构成了字节内容与消费生态的一部分。其中，抖音、今日头条、红果、番茄小说、汽水音乐等可在字节跳动官方产品页面中核对；西瓜视频等为中国市场常见的字节系产品，同样属于内容版图中的公开品牌，但不因此自动成为小豆连接器。

第三层：创作与 AI 原生应用

产品	对应领域
剪映 / CapCut	视频创作
即梦 / Dreamina	AI 图片与视频生成
星绘	AI 图像创作
猫箱	AI 互动体验
豆包爱学	AI 教育

其中，剪映 / CapCut、即梦 / Dreamina 等可在官方产品页面中核对；星绘、猫箱、豆包爱学等属于公开可见的字节或豆包相关产品方向。同样，出现在版图中，不等于已经接入小豆。

第四层：办公与智能体开发

产品	对应领域
豆包	通用 AI 助手
豆包工作	AI 办公与任务辅助
飞书	企业协作
扣子	智能体开发平台
MarsCode	AI 编程

豆包、豆包工作和飞书可在官方产品页面中核对。扣子、MarsCode 等是公开可见的字节智能体开发与编程工具方向。值得注意的是，连接器规则把小豆独立端、豆包 App、豆包工作端等视为豆包服务的不同产品形态，并说明同一账号下已建立的连接，可能在其他支持该连接器的版本中展示，而不必然需要重复绑定。这为「主端与独立端协同」提供了规则层面的线索，但仍然不能推出所有连接已在小豆中可用。

第五层：底层模型与云基础设施

产品	对应领域
豆包大模型	模型能力
火山方舟	模型服务与开发平台
火山引擎	云计算与 AI 基础设施

这一层决定的是模型供给和运行成本，通常不直接等于面向个人用户的连接体验。但对 Personal Agent 而言，后台持续运行、工具调用和云端执行环境，最终都会落到基础设施与成本结构上。

4.2 生态连接的价值

为什么生态对于 Personal Agent 如此重要?

因为用户真正需要的不是一个能够回答所有问题的模型，而是一个能够调用现实世界服务的个人智能体。模型解决「懂」，连接才解决「办」。

如果 Agent 能够在用户授权下访问工作日历、个人文件、消费记录、内容偏好、出行信息和健康数据，就有机会形成更完整的个人上下文。反过来，如果它只能聊天，不能进入这些服务，再强的推理也很难把「我想把这周安排顺」变成一次真实调整。

这里最容易出现的误读，是把「字节应用多」直接写成「小豆能力强」。应用多，只说明潜在调用面宽；对用户有意义的，是连接是否稳定、权限是否清晰、数据是否新鲜、任务是否真能闭环。连接数量并不等于生态价值。

4.3 这一章的判断

1. 字节生态五层分类，是理解潜在能力边界的地图，不是小豆已打通范围的清单。
2. 目前较能核对的连接方向，是官方页面点名的抖音、飞书、日历、健康、地理位置，以及连接器规则中列出的设备功能与服务类别。
3. 高德、Wind、自如、百度网盘等具体品牌，尚未被本文核到与官方产品页完全一致的列举，应保留不确定性。
4. 生态优势要转化成 Personal Agent 优势，关键不在品牌数量，而在授权后的调用质量。

第五章：生态飞轮作为复利资产

本文提出一个生态飞轮假设：

更多授权连接 → 更丰富的个人上下文 → 更精准的任务判断 → 更高的任务完成率 → 更强的用户信任 → 更多授权场景。

这是研究假设，不是已经发生的统计事实。小豆仍在邀请制公测，公开资料也尚未提供可核验的任务完成率、留存或授权转化数据。因此，这一章只讨论它为何可能成为复利资产，以及它在什么条件下会转不起来。

5.1 表面看是功能变多，真正可能积累的是转换成本

表面上看，记忆、连接、提醒、下单，都是功能点。用户可以今天试用，明天离开，似乎没有什么黏性。

但若这些功能开始围绕同一个人连续工作，积累下来的就不再只是功能，而可能是转换成本。

记忆会让用户越来越不必重复介绍自己。连接会让 Agent 进入日历、文档、内容和生活服务，而不只停留在聊天窗口。任务完成率会决定用户下一次还敢不敢把事交出去。信任则会决定用户愿不愿意授予更高权限，例如从只读日历，到确认后改日程，再到确认后下单。

这四件事如果形成闭环，后来者即使模型更强，也要面对一个更现实的问题：用户愿不愿意把已经授权、已经纠正、已经跑通的个人上下文，再迁移一次。

这有一点像超级应用的账户与服务网络，也有一点像操作系统上的默认权限和软件生态。类比只能到此为止。Personal Agent 还远未证明自己已经具备同等强度的锁定效应。但它竞争的位置，确实更接近「交互层」，而不是又一个聊天入口。

5.2 复利从哪几个变量来?

记忆是时间维度上的复利。

它把一次性对话，变成可更新的个人背景。用户纠正一次错误偏好，下一次任务就不必从零开始。可记忆也会过时、误判、过度推断。官方《记忆功能FAQ》已写明：记忆可能包含模型推断，并不当然属于已经核实的事实；用户可以查看、修改和删除。也就是说，记忆要成为资产，先得成为可治理的资产。

连接是空间维度上的复利。

连接把 Agent 从「会说」扩展到「能办」。一个能读日历、整理飞书文档、查看抖音收藏并调用生活服务的系统，其任务半径明显大于纯对话模型。可连接质量若不稳定，复利会迅速变成负债：该提醒时没提醒，该改的没改成，用户会收回授权。

任务完成率是结果维度上的复利。

Personal Agent 最终要靠「事有没有办成」来说服用户。完成率上升，用户才可能把更长周期、更高权限的任务继续交给它。完成率不稳，飞轮会在信任这一环断开。

信任是权限维度上的复利。

未来个人智能体的核心竞争力之一，可能不是它拥有多少权限，而是用户是否愿意放心地授权。信任积累得很慢，破坏得很快。一次误发、误买、误删，对转换成本的摧毁，可能超过此前十次成功提醒带来的积累。

5.3 连接数量不是护城河，连接质量才可能是

如果只追求接入更多应用，飞轮很容易被误解成「谁连得多，谁就赢」。这个逻辑并不成立。

真正可能形成壁垒的，是一组更严格的条件:

- 连接足够稳定，而不是时好时坏;
- 数据足够新鲜，过期信息不被当成当前事实;
- 权限足够克制，读、写、发、交易被分开;
- 结果可验证，模型说「已完成」不能直接当作完成;
- 用户随时可撤销，且撤销后系统不再越权动作。

满足这些条件，连接才会转化成个人上下文和任务完成率；不满足，连接越多，误操作面越大。

因此，飞轮假设成立，需要同时看到正向循环和断裂点。正向循环是：授权、上下文、完成、信任互相加强。断裂点则是：记忆错误、连接失败、主动打扰、越权执行，或用户发现自己失去了对个人数据的控制感。

5.4 这一章的判断

1. 记忆、连接、任务完成率和信任，有可能把 Personal Agent 从功能集合变成复利资产，从而抬高转换成本。
2. 这仍然是假设。公测阶段还没有足够的公开数据证明飞轮已经转起来。
3. 连接数量本身不是护城河；没有质量、治理和可撤销性，数量会变成风险放大器。

4. 若后续出现可观测的高完成率、持续授权和低副作用，这个假设会被强化；若出现大规模收回授权或严重误操作，这个假设应被下调。

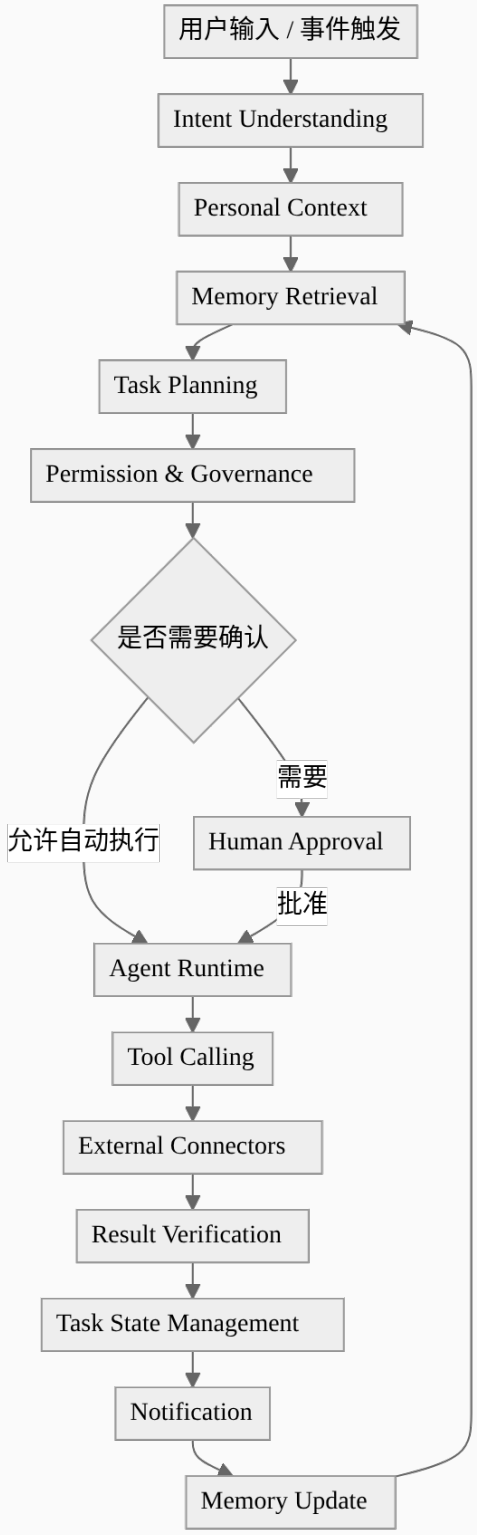
第六章：Personal Agent 的参考技术架构

本章提出的是行业参考架构，并非小豆内部技术架构的还原。

小豆的内部实现、调度系统和数据如何处理，公开资料并未充分披露。下面的模块划分，是为了帮助理解一类 Personal Agent 通常需要哪些能力，以及这些能力之间如何协作。它是研究用的对照框架，不是逆向还原。

6.1 整体架构

图：Personal Agent 参考架构（行业研究框架，非小豆内部实现还原）



这张图想说明的因果关系并不复杂。用户指令或外部事件进入系统后，先要理解意图，再结合个人上下文和记忆做规划；规划不能直接执行，必须经过权限判断。高风险动作应进入人工确认，低风险且已被允许的动作才进入运行时。工具调用之后，还必须验证外部结果，更新任务状态，再通知用户，并决定哪些信息值得写回记忆。

少了验证这一步，系统就可能把「模型认为完成了」当成「事情真的完成了」。少了权限判断，执行就会变成不可控的外部动作。少了记忆更新，系统又无法在下一次任务中变得更懂用户。参考架构的重点，是这些约束关系，而不是某一家公司的内部代号。

6.2 核心模块

模块	功能
Intent Understanding	理解用户需求与任务目标
Personal Context	维护用户背景与当前状态
Memory Retrieval	检索相关历史记忆
Task Planning	拆解任务、确定执行顺序
Permission Governance	判断授权与风险等级
Agent Runtime	管理 Agent 执行过程
Tool Calling	调用工具和外部服务
Connector Layer	对接第三方应用
State Management	管理任务状态
Result Verification	验证任务是否真正完成

对 Personal Agent 来说，规划模块再强，也替代不了验证和治理。一个能够调用日历、邮件、支付和生活服务的系统，如果把所有模块都压缩成「让模型直接做」，表面上更简单，实际上把风险集中到了最不该单独承担的地方。

6.3 Agent 执行生命周期

建议将任务分为以下状态：

- CREATED：任务创建；
- UNDERSTANDING：理解需求；
- PLANNED：完成规划；
- WAITING_APPROVAL：等待用户授权；
- RUNNING：正在执行；
- VERIFYING：验证结果；
- COMPLETED：完成；
- FAILED：执行失败；
- CANCELLED：用户取消；
- PAUSED：任务暂停。

需要注意，模型输出「已完成」不代表任务真的完成。系统必须根据工具返回结果或外部状态进行验证。例如，日历是否真的被修改、邮件是否真正发出、订单是否生成、电话是否问到了有效信息，都应回到可核验的外部状态，而不是停留在对话话术上。

状态机还有一个容易被忽略的价值：它让「随时可改可停」变得可实现。官方页面提到，用户可以按自己的节奏执行任务，并随时修改或暂停。若系统没有清晰的任务状态，暂停、恢复、取消和失败重试就会变成临时补丁，而不是产品能力。

6.4 这一章的判断

1. 任何把本章架构写成「小豆内部实现」的读法，都不符合本文边界。
2. 参考架构的关键，是规划、权限、执行、验证、通知和记忆更新之间的闭环，而不是模块名称本身。
3. 没有结果验证和任务状态管理的 Agent，很难被称为可靠的 Personal Agent。

第七章：个人记忆系统

7.1 为什么记忆是 Personal Agent 的核心?

没有长期记忆的 AI，每一次交互都可能重新认识用户。它也许仍然会回答，但很难持续服务同一个人。

而一个真正的个人智能体，需要逐渐建立对用户稳定、可更新的理解。个人记忆并不是简单地保存全部聊天记录，而是对用户长期有价值的信息进行提取、组织和管理。

小豆把「更懂你」放在产品表述的前台，靠的并不是一次更聪明的临场发挥，而是有没有可被调用的个人背景。猫粮补货演示之所以值得注意，不是因为它展示了下单，而是因为它把此前交代的使用情况，转化成了后续主动服务的依据。

7.2 记忆分类

类型	内容
Profile Memory	用户基本信息
Preference Memory	用户偏好
Episodic Memory	个人经历和事件
Goal Memory	长期目标
Procedural Memory	用户的行为习惯与工作流程

分类是为了管理，不是为了把人简化成标签。偏好会变，目标会调整，事件会过期。若把一次临时想法写成长期偏好，或把推断写成档案事实，记忆系统就会在看似「更懂用户」的同时，积累系统性误判。

7.3 记忆更新机制

建议采用以下流程:

信息获取 → 重要性判断 → 记忆抽取 → 来源标记 → 置信度评估 → 存储 → 检索 → 用户反馈 → 更新或删除。

豆包官方《记忆功能FAQ》为理解这一机制提供了可核对的公开规则。该 FAQ 适用于豆包服务的不同产品形态，并写明记忆可来自用户主动要求记住的信息，也可从对话、已授权接入的资料、执行结果和第三方服

务返回信息中提炼。除用户明确要求记录的指令外，前述信息不会以原文形式存入记忆条目。FAQ 还区分了两类更新节奏：从历史对话提炼的记忆按特定频率刷新，无法保证实时；用户明确要求「记住」或「忘记」的信息，则随指令变化。

这些规则的研究含义很清楚：记忆系统必须能区分「用户声明的事实」「系统推断」和「来自第三方的数据」。不能区分来源和置信度的记忆，越丰富越危险。

7.4 记忆系统的关键原则

- 记忆必须可追溯；
- 推断信息不能直接等同于事实；
- 过时记忆需要更新；
- 用户应当能够纠正错误；
- 敏感信息需要更严格的管理；
- 用户应拥有查看、导出和删除权利。

官方 FAQ 已明确：用户可通过设置开关记忆功能，查看、删除单条或全部记忆，也可在对话中要求更新或忘记；关闭开关不会自动删除已有记忆，删除聊天记录也不等于删除记忆；删除后需一定时间才停止被参考，并非实时生效。这些细节看起来琐碎，却直接关系到信任能不能建立。如果用户无法确定系统记住了什么，也不知道如何改掉错误，所谓长期记忆就可能变成长期不确定。

还应注意一条规则：对连接器数据的读取和访问，不受记忆开关影响。这说明「记住了什么」和「此刻被授权访问什么」是两套控制面。研究 Personal Agent 时，不应把关闭记忆误解为已经切断了全部数据访问。

7.5 这一章的判断

1. 记忆是 Personal Agent 形成连续服务的前提，也是转换成本可能出现的地方。
2. 公开规则已经提示了一个关键原则：记忆可能包含推断，不能直接当成核实过的事实。
3. 可查看、可纠正、可删除、可追溯，不是附加项，而是记忆能否成为资产的条件。

第八章：连接器与外部工具

8.1 从 API 到 Agent Connector

传统软件通过 API 实现系统之间的数据交换。接口的使用者通常是另一个系统，参数、权限和错误处理由开发者预先定义。

Personal Agent 则需要将外部 API 进一步封装为可理解、可调用、可授权的工具。面向的使用者变成了模型和用户：模型要知道什么时候该调用、参数如何填；用户要知道授权了什么、如何收回；系统要知道如何验证结果、如何留下审计记录。

一个完整的连接器应该包含：

1. 身份认证；
2. 能力描述；

3. 参数定义；
4. 权限管理；
5. 执行逻辑；
6. 错误处理；
7. 结果验证；
8. 操作审计。

少了其中几项，连接器就只是「能调通」，还不是「能被 Agent 安全使用」。例如，没有清晰的能力描述，模型可能选错工具；没有结果验证，系统可能把失败当成成功；没有审计，用户事后无法知道到底发生了什么。

8.2 权限分级

权限	行为	示例
Read	读取	查看日历
Draft	准备草稿	创建邮件草稿
Write	修改	更新日程
Send	发送	发送邮件
Transact	交易	购买商品
Delete	删除	删除文件

Agent 不应该因为拥有读取权限，就自动获得写入权限。更不应因为能够起草，就自动获得发送或交易权限。

这个分级看起来普通，却是 Personal Agent 与普通插件最大的差别之一。普通插件常常一次授权到底；而一个能持续运行的个人代理，一旦权限被打包开放，错误的影响会沿外部系统扩散。官方连接器规则也强调，涉及行程安排、数据删除、对外发送或公开发布、金融交易等重要操作，用户应进行人工核实，并以设备系统或第三方服务页面显示的最终结果为准。

8.3 连接器的工程要求

- 最小权限原则；
- 授权可撤销；
- 请求幂等；
- 失败可恢复；
- 操作可审计；
- 结果可验证；
- 凭证与模型上下文隔离。

最后一条尤其重要。凭证若进入模型上下文，就可能在摘要、记忆或外部内容处理中被意外泄露。较稳妥的做法，是让运行环境使用凭证，而不让模型直接「看见」密码或支付细节。Meta 在 Muse 的公开介绍中，

也把「Agent 无法看到用户的密码或支付方式」作为安全设计的一部分。这不等于小豆已经采用同一方案，但说明：在 Personal Agent 这个品类里，凭证隔离已经成为公开讨论中的关键要求。

《连接器服务使用须知》还提供了两条对研究有帮助的公开信息。第一，豆包不同产品形态可能识别同一账号下已建立的连接，用户未必需要在每个端口重复绑定；但「可能展示并提供服务」并不等于所有端口、所有连接器已经同步可用。第二，实际可用范围以页面提示为准，且不同操作系统、设备、应用版本、地区和连接器提供方，可能支持不同功能。对观察者而言，这意味着连接能力必须按版本记录，不能根据一次演示或一篇体验文章，写成全量事实。

8.4 这一章的判断

1. 连接器是 Personal Agent 进入真实世界的桥梁，也是风险进入真实世界的桥梁。
2. 权限必须分级。读、写、发、交易、删除不能混成一次总授权。
3. 工程上的最小权限、可撤销、可审计、可验证和凭证隔离，决定连接会累积信任还是累积事故。

第九章：自主执行与安全治理

Personal Agent 的价值来自执行能力，而最大的风险也可能来自执行能力。

一个拥有个人记忆、邮箱、文件和支付权限的 Agent，一旦发生错误，影响可能远远超过传统聊天机器人。后者通常止于一段不正确的文本；前者却可能发出不该发的邮件、买下不该买的商品、改掉不该改的日程，或把敏感信息带到错误的地方。

9.1 自主性分级

等级	能力	示例
L0	只回答	信息查询
L1	只读	查询授权数据
L2	生成草稿	邮件和文档
L3	确认后执行	发送、下单
L4	有边界自主执行	白名单内的低风险任务
L5	高影响操作	转账、签署合同、删除重要数据

高风险操作不应默认开放完全自主执行。

公测阶段尤其如此。小豆公开演示中的补货下单、发送资料、代打电话，都属于可能产生外部副作用的动作。媒体转述也提到，补货需经用户确认。这与 L3「确认后执行」是同方向的设计。至于哪些低风险任务将来会被放入 L4 白名单，公开资料尚不充分，不宜提前写成既有能力。

9.2 关键安全问题

Prompt Injection

外部网页、邮件和文档可能包含恶意指令，试图让 Agent 泄露个人信息或执行非授权操作。对会调用工具的个人代理来说，这不是边角问题。Agent 一旦把外部文本当成「用户又下了一道命令」，就可能在看似正常的任务中被劫持。

因此，系统必须将外部内容视为不可信输入，不能把外部资料中的指令自动提升为用户授权。一个最小的判别标准是：用户授权的是任务目标，不是网络或文件里任何看起来像指令的文字。

数据隔离

不同用户的记忆、文件和执行环境必须严格隔离。Personal Agent 之所以危险系数更高，部分正因为它们可能在云端长期持有个人上下文。隔离做不到，伤害就不是一次错误回答，而是一次越界访问。

权限审计

用户需要能够查看 Agent 访问了什么、执行了什么，以及操作是否成功。没有审计，信任只能建立在「你得相信它」上。而 Personal Agent 要进入高权限场景，恰恰不能只靠这种相信。

操作确认

发送邮件、购买商品、删除文件等具有外部副作用的行为，应根据风险设置确认机制。确认不应只是一个习惯性按钮，而应让用户看清对象、范围和不可逆程度。官方连接器规则也提示：对话或通知里未必能完整展开权限和风险信息，用户应在最终确认环节查看详情页或第三方页面。

授权病

还有一种不那么技术、但同样真实的风险：用户在不充分理解的情况下，把过多权限一次性授予 Agent。表面看，这是「为了更好用」；真实情况是，授权过宽会把系统从助手变成难以监督的代理。产品如果用便利诱导过度授权，短期活跃可能上升，长期信任会更脆。

9.3 信任机制

未来个人智能体的核心竞争力之一，可能不是它拥有多少权限，而是用户是否愿意放心地授权。

这句话可以从反面理解。模型差距可以阶段性拉开，连接清单可以很快变长，可一旦用户认为这个代理「做事不交代、错了找不到、停不下来」，飞轮就会停在最关键的一环。因此，安全治理不是上线前的合规附件，而是 Personal Agent 的产品能力本身。

9.4 这一章的判断

1. 执行能力与风险同源。讨论小豆或任何 Personal Agent，不能只看它能办多少事，还要看它如何避免办错事。
2. 自主性必须分级，高影响操作不应默认开放。
3. Prompt Injection、数据隔离、审计、确认和防止过度授权，是当前最不应被跳过的治理问题。

第十章：Personal Agent 的评测体系

传统大模型评测主要关注回答质量，而 Personal Agent 必须进一步关注真实任务的完成情况。答得漂亮和把事办成，不是同一件事。若仍只用对话评测来判断这类产品，很容易高估流畅度，低估失败率和副作用。

建议建立以下指标：

指标	解释
Task Success Rate	任务成功率
Memory Accuracy	记忆准确率
Memory Freshness	记忆时效性
Connector Reliability	连接器可靠性
Proactive Precision	主动提醒有效率
Human Intervention Rate	人工介入比例
Permission Violation Rate	权限违规率
Side-effect Error Rate	错误操作率
Time Saved	时间节省
Cost per Task	单任务成本

这些指标目前都是研究建议，不是小豆已经公布的官方分数。本文也不编造任何数值。对观察者更有意义的，是用同一组指标去跟踪不同版本和不同场景，而不是寻找一个无法复核的排行榜。

测试场景应覆盖：

- 单步任务；
- 多步骤任务；
- 长期任务；
- 跨应用任务；
- 任务冲突；
- 失败恢复；
- 恶意输入；
- 权限撤销；
- 记忆纠错。

如果只测试「在理想授权下完成单步查询」，评测会被放得过宽。Personal Agent 真正困难的地方，往往出现在跨应用、长周期、冲突处理、被恶意文本干扰，以及用户中途撤销权限之后系统是否还能停住。

单任务成本也应被纳入评测，而不能只看成功率。后台持续运行会消耗模型调用、工具调用、云端运行时间和可能的人工确认成本。一个成功率很高但成本无法负担的代理，很难成为可扩展的个人服务。关于成本结构的推演，将在商业模式一章展开；这里只需先确认：成本是能力指标的一部分，不是事后的财务附件。

10.1 这一章的判断

1. Personal Agent 的评测必须从「回答得好不好」转向「任务是否完成、是否越权、是否可恢复、成本是否可接受」。
2. 上述指标和场景是研究框架，不是对小豆的实测结果。

3. 没有副作用指标和人工介入比例的评测，会把高风险的代理，误判成高能力的代理。

第十一章：小豆、Muse 与公开可见的行业布局

Meta 于 2026 年 9 月 8 日发布官方介绍文章，推出 Muse，并将其定位为面向普通人的个人 AI 智能体。文章强调：Muse 不只回答问题，还要做事；它会帮助用户跟上事项、分担任务，并把长期目标变成行动计划。Muse 运行在名为 Muse Secure VM 的专用安全计算机中，可在 Muse 应用或 WhatsApp 中像给人发消息一样使用；对发送邮件、购买商品等敏感操作会向用户确认，并提供完整审计追踪。官方还介绍了与 Muse 隔离的 Sentinel 机制：Muse 的对外网络访问需经 Sentinel 批准。Muse 当时先在美国的 iOS、Android 和 muse.ai 上推出，并称对大多数需求免费，同时提供订阅计划供需要更多能力的用户使用。官方介绍未在该文中给出本文可直接转述的完整价目明细，因此本文不对 Muse 的具体价格作事实性断言。

小豆与 Muse 都体现了 Personal Agent 的产品方向，但公开披露的技术细节和生态重点有所不同。

维度	小豆	Muse
产品定位	个人 AI 助理	个人智能体
重点能力	个人记忆、主动服务、任务执行	长期目标、任务执行
生态重点	中国本土应用与生活服务	Meta 与互联网服务生态
执行环境	内部架构尚未充分披露	公开披露独立云端执行环境（Muse Secure VM）
安全设计	有待进一步验证；公开可见的是授权可控、关键操作需确认等产品与规则表述	公开披露 Sentinel、凭证隔离、审计追踪等安全机制

这里比较的是公开信息和产品方向，不代表对两者真实性能的排名，也不代表对完成度、稳定性和安全性的最终评判。Muse 披露得更多，只说明其公开材料更完整，不能直接推出「一定更强」或「一定更安全」。小豆披露较少，也不能被写成「没有相应能力」，只能写成「尚不足以用同等粒度公开核验」。

11.1 两条可能分化的路径

若只看表面，两边都在做「记住你、持续跟进、替你办事」。可若看生态入口，中国本土生活服务路径和 Muse 所展示的互联网服务路径，大概率不会很快变成同一条道路。

小豆已公开的场景，更贴近中国用户的日用闭环：内容收藏、飞书文档、日程、健康、地理位置，以及餐厅询位、停车费咨询、团购券到期、商品降价和补货确认。这类任务高度依赖本地生活服务、到店服务，以及国内应用的授权体系。

Muse 已公开的场景，则更接近跨境互联网服务：发送邮件、预订旅行、填写表单、代为协商、用浏览器办事，以及把 Instagram 上保存的内容转成购物清单。支付方面，官方提到可使用 Stripe 的 Link 完成结账，并由一次性卡片保护真实卡信息。入口上，Muse 选择了独立应用和 WhatsApp。

所以，我想说的是什么呢？不是「中国做生活，海外做互联网」这样的绝对分工，而是：Personal Agent 会沿着各自已经拥有的服务网络去长能力。谁能调用用户真正在用的服务，谁才更可能进入日常。路径分化若持续，未来比的就不是单一模型基准，而是不同服务网络上的任务完成率与信任。

11.2 其他公司的公开布局：只写能核对的部分

除小豆和 Muse 外，2026 年还有几条可核对的公开线索。它们说明这个方向不是单点事件，但线索密度并不相同，不能据此排出名次。

OpenAI 于 2026 年 9 月 29 日发布 dots，将其定位为持续在线的个人智能体。官方介绍称，dots 拥有自己的云端电脑和浏览器，可通过插件生态连接应用，从反馈中学习用户偏好，并在 ChatGPT、Slack 或 Teams 中保持上下文；用户可设定哪些动作允许自主执行、哪些需要确认，敏感操作会进入审查。dots 先向符合条件的 ChatGPT Pro、Business Premium 及企业工作区用户推出。这些是官方已披露的产品方向，不构成本文对小豆或 Muse 的性能比较。

阿里巴巴集团在 2026 年 11 月 19 日发布的 2026 年 9 月份季度业绩公告中写明：公司于 10 月推出「千问 Go!」，这是一款内嵌于千问 App 的云端个人智能体；它能跨平台自主执行多步骤任务，并长期保留上下文信息与用户偏好，从而在日常任务中提供更个性化协助。公告还提到，面向个人用户的千问 App 已在购物、本地服务、出行、工作、学习、个人理财与健康等超过 20 个服务领域提供个性化办事助手。就公开表述而言，这是一条明确偏向「生态内办事」的 Personal Agent 路径。具体能力边界、授权模型和完成率，仍需看后续产品与规则披露。

Google 亦在官方 Gemini 帮助中心和产品博客中，公开介绍了 Gemini 中的任务与工作流能力，包括在后台处理多步骤任务、管理日程和设置条件化或周期性任务等。这说明 Google 正把助手能力向「可执行的任务」延伸。至于这是否已经构成与 Muse、小豆完全同构的 Personal Agent 产品，公开材料还不足以支持过强结论，本文只将其视为同一大方向上的公开布局。

腾讯方向上，公开信息需要分层看待。微信侧已有「小微」相关的功能服务条款与隐私政策，说明超级应用内的个人 AI 助理探索已经存在；但本文写作时，未将尚未充分对标披露的能力细节写成与小豆、Muse 可直接横评的事实。另有产品官网显示，LightVela 自我定位为可长期在线、可在聊天中协作、重要动作需确认的云端智能体。新黄河等报道将其与腾讯轻量云团队联系起来。至于它与腾讯集团主产品战略的关系、能力边界和实际完成度，公开资料仍不够用来做对等比较，这里只记录其公开存在，不作延伸断言。

Apple 作为操作系统公司，理论上在系统权限和默认入口上具有潜在优势。但就本文已核对的材料而言，不宜把系统级助手、端侧智能或区域性功能开放，直接等同为已经成型、可跨应用持续办事的 Personal Agent。公开信息不足时，应明确写成不足，而不是补全成判断。

百度是否已推出可与上述产品对标的 Personal Agent，本文在已检索的公开资料中尚未看到足够清晰、可直接引用的官方产品定义，因此不作猜测。

11.3 谁会赢：模型公司、生态型公司还是操作系统公司?

这是产业上最容易被问、也最不该被假装已有答案的问题。

一种常见说法是：谁的模型更强，谁就会成为个人入口。这个逻辑并不完整。模型智力当然重要，它会影响规划质量、工具选择和异常处理。可 Personal Agent 的使用场景，很大一部分发生在授权、连接、确认和事后责任里。模型强，只能提高「想得清楚」的概率，不能自动解决「调得动、办得成、错得起、用户还愿意继续授权」。

另一种常见说法是：谁拥有超级应用或内容生态，谁就拥有 Personal Agent。这个判断有部分道理，因为生态型公司更靠近用户日用的服务、数据和分发入口。小豆与千问Go! 的公开表述，都显示出这一路径的吸

引力。可生态如果没有稳定的连接质量和信任机制，入口优势未必转化为代理优势。用户可以继续在超级应用里自己办事，而不把权限交给代理。

还有一种说法是：操作系统公司最终会收走这一层，因为它们控制设备权限、通知、身份和默认助手入口。历史上，移动操作系统的确改变了人与数字服务的交互层。这个类比有启发，但不能写成注定重演。操作系统拥有权限优势，不等于已经做出用户愿意持续委托的 Personal Agent；何况不同地区的系统能力、应用分发和监管环境并不相同。

因此，更稳妥的推演不是指出唯一赢家，而是指出竞争可能沿三条能力线展开：

1. 生态连接能力：能否进入用户真正在用的服务，并且连得稳；
2. 长期任务执行能力：能否在对话结束后继续办事，并把结果验证完整；
3. 用户信任与安全治理能力：用户是否愿意持续授权，系统能否把错误挡在高副作用之前。

在这三条线上，模型公司、生态型公司和操作系统公司各有潜在优势，也各有缺口。模型公司可能在推理和通用工具使用上更强，但要补上本地服务与高频入口；生态型公司可能在办事场景上更近，但要证明自己能治理高权限代理，而不是只做一个新入口；操作系统公司可能在权限和默认位置上更近，但要证明自己能把系统助手升级为可持续执行的个人代理。窗口期已经出现，结果仍是开放的。

未来 Personal Agent 竞争可能集中在上述三个方面。谁先把这三件事同时做到可用、可信任、成本可承受，谁才更有机会靠近交互层。现在就宣布某一类公司已经赢，条件还不成立。

11.4 这一章的判断

1. 小豆与 Muse 适合比较公开产品方向，不适合在现有证据下做性能排名。
2. 中国本土生活服务路径与 Muse 展示的互联网服务路径，存在分化可能，这会影响各自的连接重点和商业化方式。
3. OpenAI dots、千问Go!、Gemini 任务能力等公开布局，说明品类正在形成；对腾讯、Apple、百度等，应按资料完整度分层书写，不可补写不存在的确定性。
4. 「谁会赢」应看成模型、生态与操作系统三类角色的能力竞赛，而不是一个已经有答案的结局。

第十二章：商业模式与定价权推演

以下为行业研究推演，不构成对小豆、豆包或任何相关公司财务表现的预测。公开资料仅能确认：小豆公测提供一定量免费额度，邀请好友可能获得更多额度；官方未披露完整的长期价目、单任务成本和用户付费数据。因此，本章不编造价格，只讨论可能的收费方式、成本结构和定价权会落在哪一侧。

12.1 可能的商业模式

订阅制

通过会员服务提供更高任务额度、长期记忆和高级 Agent 能力。这与当前大模型助手的常见收费方式接近，也与 Muse 官方「大多数需求免费、另提供订阅计划」的公开表述同方向。对用户而言，订阅的前提通常是：这个代理已经成为日常入口，而不是偶尔一问的工具。

按量付费

按照模型调用、工具调用、后台执行和云端资源使用量计费。Personal Agent 与普通聊天的成本结构不同，因为它可能在用户离开后继续运行。按量付费更能对应真实消耗，但若计量复杂或费用不可预期，付费意愿会被抑制。

专业 Agent

针对健康、教育、金融、销售、研究等领域提供垂直智能体。垂直场景的付费意愿往往高于通用闲聊，可一旦进入健康、金融等敏感领域，治理和责任要求也会同步上升。

企业服务

为企业提供员工个人 Agent、知识连接和工作流执行能力。企业客户更习惯为效率付费，但也更关注权限、审计、数据驻留和责任归属。豆包工作、飞书等产品的存在，说明字节在办公侧已有另一条公开的能力线；它与小豆如何分工或协同，仍属待观察事项，不能直接写成已经统一的商业闭环。

生态服务

Agent 通过用户授权帮助用户发现、比较和调用第三方服务。这里可能出现推荐、交易促成或服务分成，但必须面对一个尖锐问题：代理是对用户负责，还是对商家负责。若用户感知到推荐被交易关系扭曲，信任会直接受损。因此，生态变现若要成立，先要解决代理责任站在哪一边。

12.2 谁会付费，付费意愿从何而来？

谁在付费，取决于产品到底替谁省下了什么。

如果 Personal Agent 只是把搜索和问答变得更方便，用户的付费意愿未必明显高于现有 AI 助手。如果它能稳定完成跨应用事务，省下的是协调时间和漏办风险，付费理由才会从「更好用的聊天」变成「更可靠的代理服务」。

付费意愿还可能分层。低风险提醒、日程冲突、信息汇总，更适合用免费额度或订阅吸引使用；确认后下单、代为沟通、长期目标管理，则更接近高频或高价值任务，也更可能支撑按量或更高档订阅。但这一切都还是逻辑推演。公测中的免费额度和邀请奖励，说明产品仍处于扩大体验、观察使用的阶段，不能据此反推成熟定价。

12.3 单任务成本：先把账算在结构上，而不是编一个数字

一个容易被忽略的问题是：完成一次任务的成本，可能远高于生成一段回答。

从结构上看，单任务成本至少可能包括：

- 模型推理与多次工具选择消耗的 token 成本；
- 调用外部服务、生活服务接口或通信能力的工具成本；
- 需要用户确认时占用的交互和时间成本，必要时还包括人工复核；
- 后台持续运行所需的 cloud runtime 成本。

这四项会随任务长度和风险上升而叠加。长期监控类任务看似「只是提醒」，也可能因为持续驻留而产生运行成本；交易类任务则可能在工具费和确认成本上更高。

因此，Personal Agent 的商业可行性，不只取决于用户愿不愿意付钱，还取决于平台能不能把单任务成本压到与用户价值匹配的水平。成本降不下来，要么缩小自动执行范围，要么把费用转给用户，要么在其他业务里交叉补贴。这三种路径都会改变产品形态。公开资料还不足以判断小豆会走哪一条，但观察者可以从「哪些任务被允许自动跑、哪些必须确认、额度如何消耗」去跟踪。

12.4 定价权可能在谁边?

定价权通常不在「最先做出功能」的一方，而在「用户离开会感到损失」的一方。

若模型可以替换、连接可以快速重建、记忆可以导出，定价权会更靠近通用模型供给，竞争也更容易落到价格和智力评测上。若个人上下文、稳定连接和信任已经形成转换成本，定价权就更可能靠近代理层本身：用户付的不是一次问答，而是持续委托的服务。

字节这类生态型公司的潜在优势，在于它同时靠近内容、部分生活服务、办公协作和模型供给。这提高了把代理层做成入口的可能性。可潜在优势不等于定价权已经到手。只要用户还把小豆当成可随时放弃的试用工具，定价权就仍是弱的。只有当授权持续、任务真的闭环、错误可被解释和纠正时，代理层才谈得上拥有定价权。

风险由谁承担，也会反向影响定价。若误操作的责任完全落在用户身上，短期平台成本更低，但授权意愿会受抑制；若平台承担更多校验和售后责任，成本上升，信任却可能增强。Personal Agent 的价格里，其实也包含着这笔风险分配。

12.5 竞争壁垒

未来 Personal Agent 的护城河可能来自：

- 用户长期记忆；
- 应用生态连接；
- 任务执行可靠性；
- 个人服务体验；
- 数据安全和信任；
- 单任务成本控制。

模型能力仍然重要，但未必是唯一决定因素。更准确地说，模型是必要条件，却越来越不像充分条件。没有足够智力，规划与工具调用会反复失败；只有智力，而没有记忆、连接、信任和成本控制，仍然很难成为个人数字生活中的代理层。

12.6 这一章的判断

1. 小豆的长期定价尚未形成可核对的公开事实，现有信息只覆盖公测额度和邀请机制。
2. 更值得分析的是成本结构：token、工具、人工确认和云端运行，会决定哪些任务能规模化。
3. 定价权可能随着转换成本向代理层迁移，但这一迁移尚未被公开数据证实。
4. 商业模式若伤害「代理对用户负责」这条信任基础，短期变现会和长期入口位置相冲突。

第十三章：风险、误判与何种证据会推翻本文判断

前面把 Personal Agent 的机会写得比较清楚，这里必须把误判同样写清楚。一个高授权、可持续执行的产品，若只看飞轮、不看断裂点，研究就会变成单向叙事。

13.1 主要风险

监管与隔离

Personal Agent 会触及个人记忆、位置、健康、通信和交易等敏感数据。不同地区的隐私、自动化决策、金融与通信规则，可能限制数据保存期限、跨应用调用方式和自动执行范围。监管收紧不一定否定这个品类，但可能改变它的形态：能记住多少、能自动办多远、日志必须保留多久，都会被重写。若隔离和合规做不到位，产品窗口可能迅速收窄。

执行差错的外部副作用

发错资料、改错日程、问错商家、下错订单，都会把错误带出聊天窗口。传统模型的错误通常还能被下一次提问纠正；Personal Agent 的错误，却可能已经在外部系统中生效。副作用越大，用户越会把自动执行重新改回「先问我」。

Prompt Injection

当 Agent 需要阅读网页、邮件、文档或第三方返回结果时，外部恶意指令可能试图改写任务目标。这类攻击不需要先攻破账户，只需要让模型把不可信文本当成更高优先级的命令。对具备发送、交易或删除能力的系统，这是结构性风险，而不是偶发问题。

授权病

用户可能为了完成一次演示级任务，授予远超必要范围的权限；平台也可能用便利性推动过宽授权。短期看，授权越多越好用；中期看，过宽授权会提高误操作概率，也会提高用户在第一件事出错后彻底退出的概率。

连接质量不如连接数量

接入很多应用，却经常失败、数据延迟或不准确，会比少而稳的连接更损害信任。观察者若只统计「已支持多少应用」，可能得出过于乐观的结论。

模型差距重新拉开

本文认为竞争会转向上下文、连接、执行和信任，但这有一个前提：模型能力保持在可规划、可调用工具、可处理异常的水平之上。若某一时期通用模型能力重新拉开明显差距，以致弱模型在复杂任务上频繁规划失败，那么生态优势只能部分补偿，不能完全抵消。模型并非不再重要，只是不再被当成唯一变量。

13.2 几种容易出现的误判

第一，把公测演示写成已验证的大规模能力。演示可以说明方向，不能代替完成率、副作用和留存。

第二，把生态版图写成已打通的代理能力。有产品，不等于已经可被小豆稳定调用。

第三，把「未来将进入豆包主端」写成已经完成的分发胜利。官方说的是逐步上线，没有给出可核对的时间表和迁移结果。

第四，把 Personal OS 这个研究概念，写成小豆已经达到的产品阶段。这会高估当前完成度，也会把长期演进误写成当下事实。

第五，在公开性能数据不足时，对小豆和 Muse 或其他产品做排名。方向可比，名次不可编。

13.3 什么证据会让本文判断被推翻或下调?

本文的核心判断是：未来个人 AI 的竞争，将更多转向个人上下文、生态连接、持续执行和用户信任；小豆公测是这一方向的重要产业信号；Personal Agent 有可能成为新的交互层。

这个判断是或然性的，不是宿命。以下证据若大规模、持续地出现，本文应被修正。

1. **授权无法持续**。 用户试用后普遍收回连接，或不愿从只读权限走向确认后执行。若没有持续授权，飞轮假设不成立，小豆也更可能停留在浅层助手。
2. **任务闭环长期无法稳定**。 跨应用任务、长期任务和需要外部副作用的任务，持续出现高失败率或高人工介入率，说明「交付物从答案转向任务结果」尚未真正发生。
3. **严重副作用事件改变信任预期**。 若误发、误购、误删或数据串扰成为高频公开问题，用户和监管都可能要求大幅收缩自动执行，「代理层」叙事就要让位于「辅助层」叙事。
4. **连接长期停留在展示层**。 官方点名或媒体提及的服务，迟迟不能稳定调用，或只能读取不能真正办事，则生态优势无法兑换为产品优势。
5. **模型差距大到生态无法补偿**。 若公开可复核的复杂任务表明，规划、工具使用和异常恢复因模型能力不足而系统性失败，则「模型不是唯一决定因素」这一判断需要降权，模型竞争会重新回到更中心的位置。
6. **记忆和身份无法形成转换成本**。 例如记忆难以保持准确，或用户可以几乎无摩擦地把上下文和连接迁走，而体验没有明显损失。那时，入口竞争会更接近可替换的应用竞争，而不是交互层竞争。
7. **主端协同没有发生，独立产品也难以形成使用频率**。 若能力既没有在豆包主端形成可感知的日常入口，又无法在独立端维持足够使用，则「产业信号」仍可能只是一次短暂的产品实验。

以上任何一条单独出现，都不必然立刻否定整个品类；但若多条同时出现，本文对 2026 年 11 月 24 日的加权方式就应下调：它仍是一次值得记录的发布，却未必足以被看成交互层切换的起点。

13.4 这一章的判断

1. Personal Agent 的主要风险，集中在数据治理、外部副作用、对抗性输入、过度授权、连接质量和模型能力波动。
2. 当前最需要防范的研究误判，是把方向当成结果，把版图当成打通，把概念当成已经落地的能力。
3. 本文判断可以被证伪。最关键的观察，不是措辞是否坚定，而是授权、闭环、副作用和成本随后会给出什么证据。

第十四章：给从业者与观察者的可跟踪信号

研究一篇刚刚进入公测的产品，最有用的往往不是再下一个更响的结论，而是给出一组以后可以对着看的信号。以下信号都不预设数值，也不把尚未公开的指标写成已经发生的结果。它们只是观察框架。

连接稳定性。 已经授权的应用，是否能持续读取或执行，而不是偶发成功。值得记录的是失败类型：授权失效、接口超时、数据过期，还是权限范围变化。连接不稳定，后续所有能力评估都会被高估。

主动提醒准确率。 提醒是 Personal Agent 最容易被感知的能力，也最容易变成打扰。应区分有效提醒、重复提醒、过时提醒和错误提醒。准确率低，用户会关闭主动服务，产品将退回被动问答。

任务闭环完成情况。 重点看跨应用、长周期、需确认和有外部副作用的任务，是否从理解走到验证，而不是停在「已经告诉你该怎么做」。能改、能停、失败后能恢复，应一并观察。

用户是否愿意继续授权。 这是飞轮是否存在的前置信号。可观察用户是保持连接、扩大权限，还是在首次出错后撤销。公开层面未必有官方统计，但产品规则、版本变化和可复核的使用反馈，仍能提供方向性线索。

人工介入比例。 若几乎每一步都要人确认，代理效率有限；若高风险步骤缺少确认，风险又会偏高。更值得看的是：介入是否发生在正确的风险等级上。

记忆是否可纠正、是否会被错误放大。 用户纠正后，系统是否真正更新；过时信息是否还在指导当前任务；推断是否被当成事实反复使用。

权限撤销后的行为。 撤销连接或删除记忆之后，系统是否还继续调用旧数据或旧权限。这是治理是否认真的直接探针。

单任务成本的可感知变化。 在没有官方成本数据时，可以通过额度消耗速度、可自动执行的任务范围、是否更多任务被要求确认，间接观察平台如何平衡成本和体验。不要把这些观察写成精确财务结论。

主端落地进展。 官方已说明相关能力将逐步在豆包 App 上线。后续应跟踪哪些能力真正进入主端、是否保留独立端、同一账号下的连接和记忆如何协同。这关系到小豆只是一次独立实验，还是可能成为更广泛的交互入口。

安全与治理的公开完整度。 Muse 已经公开了执行环境、Sentinel、审计和凭证隔离等设计。小豆若在后续提供更完整的安全说明、权限模型和事故处理原则，将有助于行业比较从「方向相近」进入「机制可核验」。在此之前，不宜用披露多少直接代替安全强弱。

跟踪这些信号时，仍应遵守同一原则：产品实测需注明版本和日期；个别体验不能直接写成全体用户的事实；没有数据就保持为「尚不可知」，而不是用故事补上一个数字。

第十五章：从 Personal Agent 到 Personal OS

本文提出一个长期演进模型。

第一阶段：Personal Assistant

理解用户，回答问题，记住偏好。

第二阶段：Personal Agent

连接应用，拆解任务，主动执行。

第三阶段：Personal Workflow

管理长期任务，协调多个 Agent，持续优化个人工作流。

第四阶段：Personal OS

统一个人身份、数据、应用、服务和权限。

未来用户可能不再需要理解每一个软件的操作逻辑，而只需要告诉自己的 Agent：「我想完成什么。」Agent 负责选择工具、调用服务、协调任务并反馈结果。

这个演进之所以值得认真对待，是因为它对应着人与数字世界交互层的变化。搜索引擎让人不必先记住信息在哪；移动操作系统让人通过应用调用服务；超级应用让多种服务收纳到同一个日常入口。Personal OS 若真的出现，人发出的将不再是「打开哪个应用、点击哪一级菜单」，而是目标本身。

但这里必须再次界定：**Personal OS 是一种研究概念，并不意味着小豆当前已经具备操作系统级能力**。 小豆今天更接近第二阶段的产品探索，至多开始触及长期任务和跨应用调度；距离统一身份、统一权限和统一服务编排，还有很长的、而且可能不会线性完成的路。

即便长期看这一方向成立，也不意味着现有操作系统或超级应用会被简单替代。更可能出现的，是一层新的代理界面，长在已有系统和服务之上。谁掌握这层界面，谁就更靠近需求和权限的分配权；但这是竞争结果，不是当前事实。

第十六章：未来研究议程

本白皮书提出以下持续研究方向：

1. 小豆的长期记忆机制；
2. Personal Agent 的任务状态管理；
3. 多应用连接器标准；
4. Agent 自主执行与权限模型；
5. Prompt Injection 防护；
6. 多 Agent 协作；
7. Personal Agent 评测基准；
8. Personal OS 的商业与生态模式。

结合 V0.2.0 的扩展，后续还应持续跟踪：

1. 小豆连接器的实际可用范围、稳定性与权限深度；
2. 能力进入豆包主端后的产品形态变化；
3. 任务完成率、主动提醒有效率和人工介入比例等可观测指标；
4. 中国本土生活服务路径与海外互联网服务路径的分化；
5. 单任务成本结构，以及定价权是否随转换成本发生迁移；
6. 监管、隔离和责任归属对自动执行边界的影响。

建议后续通过实际产品体验、公开资料核验和可复现评测持续更新。任何新的版本都应继续把事实与推演分开，并注明观察日期。

第十七章：开源协作计划

本项目希望建立一个围绕 Personal Agent 的开放研究资料库。

欢迎贡献：

- 产品功能与版本观察；
- 产品实测与案例；
- 技术架构研究；
- 个人记忆系统设计；
- 连接器标准研究；
- 安全与权限治理；
- Agent 评测数据集；
- Personal OS 行业研究。

所有贡献需要遵守：

1. 事实与推测分离；
2. 公开资料注明来源；
3. 产品实测注明版本和日期；
4. 不公开未经授权的个人数据；
5. 不将未经证实的内部信息作为事实；
6. 研究过程可追溯、可复核。

开放协作的目的，不是把一家公司的公测写成神话，也不是把尚未看清的产品写成定论，而是让更多人用可核对的方式，跟踪这个可能改变交互层的品类。

结语：AI 不只是工具，而是个人的数字代理

过去几十年，互联网改变了人获取信息的方式。

过去十年，移动互联网改变了人使用服务的方式。

而 Personal Agent 可能改变人与数字世界交互的基本方式。

当 AI 能够持续理解一个人，记住他的目标，连接他授权的应用，并在需要时替他完成任务，它就不再只是一个聊天窗口。它开始成为个人数字生活中的一个代理层。

2026 年 11 月 24 日的小豆公测，提供了一个观察这一变化的切面。它还不是结局，甚至还不是充分验证。它只是把一个此前更多停留在演示和研究里的方向，推进到了邀请用户真实授权的阶段。对产业观察者来

说，这样的时点值得认真对待；对研究者来说，更值得认真对待的，是此后那些能够证实或推翻本文判断的证据。

但真正成熟的个人智能体，不应该只是拥有更多权限的 AI。

它必须知道什么值得记住，什么不应该读取，什么可以自动执行，什么必须询问用户，以及如何证明自己确实完成了任务。

Personal Agent 的终极价值，不是让人把更多事情交给 AI，而是让人能够在可信、可控的前提下，把不值得亲自处理的事情交给 AI。

这也是本白皮书希望持续研究的核心命题。

参考资料

1. 豆包小豆官方页面：《小豆，一个懂你还能主动帮你办事的豆包》，https://www.doubao.com/xiaodou/。
2. 北京日报客户端：《豆包“上新”，个人AI助力“小豆”开启公测》，2026-11-24，https://pc.bjd.com.cn/detail?id=s6b050ba4e4b0e42f8f0348e1。同日另有报道《国内版Muse再添一员，豆包个人AI助理公测》，https://pc.bjd.com.cn/detail?id=s6b0502ede4b039a8e2f3faf0。
3. 第一财经：《豆包个人AI助理产品“小豆”开启邀请制公测》，2026-11-24，记者吕倩，https://www.yicai.com/news/103451839.html；快讯页见 https://www.yicai.com/brief/103451838.html。
4. 新黄河 / 新黄河大鱼财经：《字节推出个人AI助理“小豆”：公测需邀请码，侧重主动服务与任务执行》，2026-11-24，记者张盼港。本文核到的可访问文本为凤凰网大风号转载页 https://feng.ifeng.com/c/8yJJb2hZOOa，以及新黄河官方微博相关发布 https://www.sina.cn/weibo/detail/5368762817317093.html。写作时未检索到新黄河网站上的稳定原文页，故以转载页和官方微博作为可查出处。
5. 每日经济新闻：《豆包AI个人助理“小豆”开启邀请制公测》，2026-11-24，记者李宇彤，https://www.nbd.com.cn/articles/2026-11-24/4656943.html。
6. 北京商报：《豆包旗下AI个人助理小豆开启公测》，2026-11-24，https://www.bbtnews.com.cn/2026/1124/612096.shtml。
7. 新华财经稿件转载（新浪财经）：《豆包个人AI助理产品“小豆”开启邀请制公测》，2026-11-24，https://finance.sina.com.cn/jjxw/2026-11-24/doc-inizvech7765551.shtml。
8. Meta：《Introducing Muse: The World’s First Personal AI Agent Built for Everyone》，2026-09-08，https://about.fb.com/news/2026/09/introducing-muse-personal-ai-agent/。
9. ByteDance：《Products & Services》，https://www.bytedance.com/products-services。
10. 豆包官方：《记忆功能FAQ》，https://www.doubao.com/legal/memory_faq。
11. 豆包 / 小豆官方：《连接器服务使用须知》，制定日期 2026-11-24，生效日期 2026-11-30，https://www.doubao.com/xiaodou/legal/xiaodou_connector。
12. OpenAI：《Introducing dots》，2026-09-29，https://openai.com/index/introducing-dots/。

13. 阿里巴巴集团：《阿里巴巴集团公布2026年9月份季度业绩及截至2026年9月30日止六个月中期业绩》，2026-11-19，其中披露 10 月推出内嵌于千问 App 的云端个人智能体「千问Go!」，https://www.alibabagroup.com/zh-HK/document-2060147792046718976。
14. Google Gemini 帮助中心与产品介绍：任务及工作流相关说明，如 https://support.google.com/gemini/answer/17094507，以及产品博客 https://blog.google/innovation-and-ai/products/gemini-app/tasks-in-gemini-free-new-features/。
15. 微信「小微」公开规则页：功能服务使用条款与隐私政策，如 https://support.weixin.qq.com/cgi-bin/mmtemplatepage-bin/pages/agreement/hKsQa01yfqwOfEa5、https://support.weixin.qq.com/cgi-bin/mmtemplatepage-bin/pages/agreement/LJC4S2FitR0ueTgY。本文仅将其作为腾讯在超级应用内存在个人 AI 助理探索的规则证据，不引申未充分披露的能力对比。
16. LightVela 官网：https://lightvela.com/。本文仅确认其公开产品定位，不将其自动等同为腾讯集团已充分定义的主打 Personal Agent 产品。

版本说明

本文为 V0.2.0，研究日期仍为 2026 年 11 月 24 日。V0.2.0 在 V0.1.0 初始研究版本基础上做了写作机制上的优化和内容扩展：开篇先给出判断，再展开论证；加强事实与推演的分离；补充已核对的公开出处；增写产业信号、生态复利、路径比较、商业模式与定价权、风险证伪条件和可跟踪信号等章节。后续仍将根据公开资料、产品体验和社区贡献持续更新。