

AI 协作记忆系统

认知架构白皮书

东莞市意图共鸣科技有限公司

作者：意图共鸣科技

发布日期：2026 年 8 月 10 日

核心命题

如果一个 AI 系统需要具备长期记忆能力，
它的信息存储和调用逻辑应该是什么样的？

文档信息

项目	内容
类型	独立认知架构 白皮书
核心主张	确认才存，双模块互审，主权归用户
定位	定义 AI 记忆系统的设计逻辑与核心机制
发布机构	东莞市意图共鸣科技有限公司
作者	意图共鸣科技
官网	www.xinirp.com
发布日期	2026 年 8 月 10 日
联络方式	contact@xinirp.com

知识产权声明：本白皮书所述认知架构、核心机制与技术方案
已提交发明专利申请并获受理，相关知识产权受法律保护。

序章

如果一个 AI 系统需要具备长期记忆能力，它的信息存储和调用逻辑应该是什么样的？

这是本白皮书要回答的核心问题。

设计理念

本白皮书的设计理念建立在两个基本判断之上。

第一，人的智能与系统的智能是两种不同的运行方式。

人的智能是涌现式的——情绪、经验、直觉、理性在同一时刻交织发生。系统的智能是计算式的——接收输入、逻辑推演、输出结果。两条路完全不同，但目标可以相同：在交互层面，让人感受到被理解。

第二，系统不需要"像人一样思考"，只需要通过规则设计，输出与人等效的交互结果。

它不需要模拟人格、假装有情感，但它需要有分寸感、有边界意识、可被信任。这决定了本白皮书的架构方向：不追求模拟，只追求适配。

路径不同，结果等效。这就是"求同存异"的底层逻辑。

这个系统的设计遵循一个基本判断：记忆不是越全越好，而是越准越好。

交互过程中会产生两类信息——探索性的和确定性的。前者是过程，后者是结果。一个好的记忆系统应当天然具备区分两者的能力：让过程自然消退，让结果沉淀为可复用的知识。这不是一个存储策略，而是一个认知原则。

三个设计困境

当前，AI 记忆系统普遍面临三个设计困境：

困境一：存储与遗忘的边界在哪里？

AI 系统是否应该记住每一次交互的全部内容？如果全量存储，信息过载会稀释真正有价值的内容；如果自动遗忘，如何判断什么值得保留？这需要一套明确的筛选机制，而非依赖上下文窗口的自然截断。

困境二：记忆如何被验证和信任？

在交互过程中，AI 会接收到大量信息——有些是经过验证的结论，有些是未经确认的假设，有些是随意的推测。如果系统不加区分地存储所有信息，后续调用时就无法判断哪些信息可靠。这需要一套验证机制，而非假设所有输入具有同等可信度。

困境三：记忆的边界如何界定？

一个具备长期记忆的 AI 系统，是否应该在所有场景下都加载全部历史记录？如果始终加载，已有的认知框架会约束它给出独立判断的能力；如果不加载，又失去了长期记忆的意义。这需要一套调用规则，而非简单的"全量加载或全不加载"。

这三个困境指向同一个问题：AI 记忆系统需要一套关于"什么值得存、以什么结构存、什么时候调用"的设计逻辑。

本白皮书提出一套认知架构——从信息筛选、存储结构、冲突处理到调用机制——作为对这些问题的回答。

第一章：典型使用流程

在进入架构设计之前，先看一个典型的日常使用流程。这有助于理解后续每一层设计在真实场景中的位置。

1.1 一次完整的交互周期

09:00 —— 开启协作

用户打开协作记忆窗口，开始日常交互。系统加载已有的存储空间，但不主动调用任何信息——它在后台待命。

09:05-10:20 —— 自由探讨

用户和系统围绕一个项目进行深入讨论。过程中有发散、有试探、有推翻重来。这个阶段系统不强制存储任何内容——探讨是过程，过程不需要被记住。

10:25 —— 系统主动提议

探讨进入收尾阶段，出现了一个明确的决策。系统主动提议："今天的讨论有几个值得整理的结论，要看一下吗？"

10:30 —— 用户确认存储

用户查看系统整理的结构化条目，确认其中三条值得存储。系统写入存储空间，标记为"确认级"，附带版本号和时效条件。其他内容不进入存储空间，自然消退。

14:00 —— 新话题启动

用户开启一个新项目。系统检索到存储空间中有两条相关历史信息，主动调用并呈现给用户："这次讨论的方向，和之前关于 X 的结论有关联，需要参考吗？"

用户选择参考其中一条，另一条暂不调用。

18:00 —— 日终汇总

系统生成简短的日终摘要："今天新增 3 条可调用信息，累计 26 条活跃信息。2 条信息进入休眠状态。"

1.2 这个流程说明了什么

这套流程不是"AI 自动记住一切"，而是"人机协同的认知积累"。系统负责识别、整理、检索；人负责判断、确认、调用。探讨性内容自然消退，确认性内容永久沉淀。

存储不是默认动作，是确认后的结果。调用不是全量加载，是按需检索。遗忘不是错误，是系统设计的一部分。

这就是本白皮书要定义的架构。

第二章：核心理念——确认才存

2.1 两种根本不同的内容

每一次交互中产生的信息可以分成两类：

类型	特征	例子
探讨性内容	临时的、发散性的、未经最终确认的	"这个方案可能行""试试看""方向对不对？"
确认性内容	经过核实的、可复用的、有明确结论的	"这个方案确定，按此执行""这条规则成立"

现有系统把这两类混在一起同等对待。结果是：系统可能把探讨中随口说的一句话，当成确认过的结论来使用。探讨是过程，确认是结果。把过程当结果，是信息混乱的根源。

探讨性内容是交互的主体——想法、草稿、试探、讨论。它们是过程，不是结果。探讨不需要被记录，就像讨论的草稿不需要被存档。信息的存储应当以"确认"为分水岭。

2.2 "确认"作为信息质量的分水岭

未经确认的信息，是"原始记录"——模糊的、可能出错的、不可靠的。

经过确认的信息，是"结构化知识"——精确的、有出处的、可验证的。

系统应当只存储经过确认的内容。确认的那一刻，一个"探讨"升级为了一个"可调用的知识"。系统在未来调用它时，才能以此为基础进行可靠的推理，而不是一堆未验证的记录上构建回答。

2.3 为什么需要确认机制

当前 AI 记忆系统的普遍做法是自动存储所有对话内容。这种做法面临一个根本困境：系统无法区分"探讨"和"结论"——用户试探性的想法和最终确定的决策，在存储层面被视为同等价值。

如果系统把一切信息都存下来，它就会变成一面镜子——用户的偏见、用户的错误假设、用户的不成熟想法，全部被忠实地记录并放大，成为后续调用的"事实依据"。

只有经过确认的内容，才具备作为推理基础的可靠性。确认机制的价值在于，它为信息入库设置了一道质量门槛——不是所有信息都值得被记住。

2.4 核心表述：存储可验证的信息

系统中存储的信息应具备可复用、可验证、可追溯的特征，而非模糊的、可能产生偏差的原始记录。存储的目标不是"全量记录"，而是"可调用知识"。

2.5 触发规则：何时提出确认建议

并非每次交互结束后都需要提出确认建议——频率过高会降低有效性。系统在以下节点可以提出确认建议：

触发类型	标准	适用场景
结构化触发	能抽象成规则、能复用于同类场景、能形成标准模板	程序性信息、技能、流程
认知触发	更新用户的认知框架、形成新视角	感悟、洞察、思维转变
用户主动触发	用户主动说"这条存一下"	任何内容

触发原则：确认建议的提出应以"是否可能产生可复用知识"为标准，而非"是否达到某个交互轮数"。用户主动触发覆盖所有类型，是最高优先级的确认方式。

第三章：系统架构——双模块设计

3.1 总体架构

本系统由两个核心模块组成：

模块 A：协作记忆窗口

功能：加载记忆空间，系统拥有持续上下文

用途：持续对话、长期项目

特点：上下文自然累积，但细节会随时间模糊

↓ 人机协同筛选 ↓

"这条内容值得存储吗？"

记忆空间：结构化知识库

内容：仅存储经确认的信息

结构：带信任级别、版本号、时效条件

调用：在协作窗口中按需调用

模块 B：独立判断窗口

功能：不加载记忆空间

依赖：仅基础模型能力

用途：获取不受已有信息影响的独立判断

核心设计原则：两个模块使用同一个基础能力，差异在于是否加载记忆空间。加载了记忆空间的是持续协作系统，不加载的是独立判断系统。基础能力是公共的，记忆是私有的。

3.2 模块 A：协作记忆窗口

定义：一个持续存在的对话环境，加载记忆空间中的结构化知识，在交互中可以随时调用已存储的信息。

运作机制：

上下文随交互自然累积，短期内保持完整的连贯性

如同系统的短期上下文，长上下文终会丢失细节——这是技术的本质限制

窗口中的交互大多是"探讨"性质的——草稿、想法、未成熟的方向。它们不需要全部存储

当交互中产生了值得长期保留的结论时，进入筛选确认流程

在关键决策场景中，加载了历史记忆的系统可能会受到已有认知框架的约束，无法提供独立的判断视角。因此，系统需要一种"不加载记忆"的运行模式——这就是模块 B 的设计目的。

3.3 模块 B：独立判断窗口

定义：每次开启都不加载记忆空间的独立环境。它不知道用户的历史交互内容，不知道用户的偏好，不知道之前形成的任何结论。

为什么需要独立判断窗口？

当系统加载了全部历史记忆，它的判断会受到已有认知框架的影响。在绝大多数场景下这是优势——它更适配用户的认知模式。但在特定场景下，这变成了约束：

持续协作系统形成了一个方案，双方都觉得没问题。但这个"没问题"可能是因为双方共享了同样的前提假设。一个"不知道这些前提"的系统可以发现这个盲区。

在做重要决策时，需要一个不带任何历史立场的判断。

进入全新领域时，已有的经验可能成为干扰而非助力。

使用场景：

场景	为什么需要无历史记忆的系统
验证重要结论	发现协作系统和用户共同忽略的盲区
获取独立判断	不带立场的、基于公共知识的回应
交叉验证	用独立视角检验协作系统的结论
新领域探索	避免已有经验干扰新判断

大多数时候系统需要"了解你"。少数关键时候系统需要"不了解你"。两个模块的设计同时覆盖这两种需求。

3.4 记忆空间：结构化信息库

这是整个系统的核心。它不是对话记录的仓库，而是经过筛选、确认、结构化的信息库。

存储机制——人机协同确认：

交互中产生有价值的结论 ↓ 系统识别并结构化整理 ↓ 输出给用户确认 ("这条内容值得存储吗? ") ↓ 用户确认 → 存入记忆空间,成为可调用知识 用户标记 → 作为参考记录,不纳入基础调用 用户拒绝 → 不存,作为普通交互记录自然消退

3.5 双模块协作方式

两个模块最有效的用法是互相验证：

第一步：在模块 A（协作）中形成方案 → 基于记忆空间中的已有信息，得出一个方案 第二步：把方案提交给模块 B（独立）审视 → "不了解背景，仅从逻辑看，这个方案有什么问题？" 第三步：回到模块 A，综合两个视角调整 → 协作系统提供深度，独立系统提供客观 第四步：最终结论经确认后，存入记忆空间 → 信息持续积累

第四章：关键机制

4.1 存储结构

每条存储信息按以下字段组织，确保系统能"调用"而非"回忆"：

字段	必填	说明	示例
标题	✓	概括这条信息的内容	"客户沟通策略"
背景	✓	在什么场景下产生的	"私域用户触达的讨论"
结论	✓	核心观点或决策	"首次接触时不直接报价，先诊断需求"
信息类型	✓	规则型/感悟型/流程型/事实型	规则型
关键参数	选填	涉及的数字、标准、规则	"报价区间：599-999"
可复用内容	推荐	话术、模板、流程等可直接调用的内容	"标准话术：看需求……"

信任级别	✓	确认级/参考级/待验证	确认级
关联标签	选填	所属领域或项目	#客户沟通 #私域
时效条件	选填	什么情况下应被标记为过时	"定价策略调整时"
版本号	✓	同一主题的迭代记录	v1.0
关联信息 ID	选填	与此信息相关的其他信息 ID	"信息#023"
创建时间	✓	创建日期	2026-07-15
最后调用	系统自动	最后调用时间及累计次数	2026-07-20 / 累计 8 次

信息类型的意义：不同类型的信息，系统默认的调用权重和调用场景不同。规则型信息在高确定性决策中优先调用；感悟型信息在需要认知框架调整时调用；流程型信息在执行类任务中调用；事实型信息在知识问答场景中调用。

关联信息 ID 的意义：它把存储从平铺的表格升级为网络——信息之间是有关系的："客户 A"和"订单#123"是关联的，"原则 X"和"案例 Y"是引用的。

4.2 时间维度：信息的价值变化

信息的价值随时间变化，不同场景需要不同的排序方式：

排序方式	适用场景	说明
时间优先	近期信息关联度高	近期信息权重高，适合短期上下文
调用频率优先	高频调用的方法论	被调用越频繁，价值越高
信任级别优先	高确定性结论	确认级高于参考级

系统默认采用"信任级别优先+时间衰减"的混合排序。每条信息自带时间戳和调用频率记录。

4.3 冲突处理规则

随着存储信息的积累，必然会出现相互矛盾的内容。当系统检测到冲突时，按以下规则处理：

优先级	规则	说明
0	来源可靠性优先	同信任级下，来源为直接确认 > 系统提取 > 推测
1	用户指定	系统检测到冲突后主动询问用户，用户决定
2	时间优先	新信息覆盖旧信息，旧信息标记为"已替换，归档"
3	信任级别优先	确认级 > 参考级 > 待验证。同级别的按时间排序

核心原则：系统默认自动处理低风险冲突（如同一条信息的时间先后），高风险冲突上报用户处理。

4.4 认知代谢：信息的沉淀与消退

存储空间的价值不仅在于积累，还在于懂得消退。存储空间需要定期管理，否则会累积大量过时信息。

系统通过两种机制管理信息：

结晶——在交互的转折点，将高价值信息固化为结构化条目。确认的那一刻，探讨升级为知识。

修剪——对长期未被调用的信息逐步降权，使其回归背景噪声，不主动参与推理。

信息的四个状态：

状态	触发条件	处理方式
活跃	被定期调用，未过时效	正常参与推理
休眠	超过设定时间未被调用（默认 90 天）	不主动参与调用,但用户可手动调用
过时	被新信息替代，或时效条件触发	标记为"过时", 移入归档区
删除	用户主动删除	彻底移除

时效机制：用户可以为每条信息设置"时效"。有些信息是长期的（如核心方法论），有些是有时效的（如某次活动的具体安排）。时效到期后，系统会提醒用户续期、更新或归档。

认知代谢的目标不是"删除"，而是"保持存储空间的高信噪比"。
让无用的信息沉底，让核心的知识浮现。

4.5 调用方式

当前主流做法是"被动带入"——把存储信息全部塞进上下文，每次调用都带着。这导致两个问题：上下文被无关信息占用，以及判断被无关信息干扰。

本框架的做法：存储空间的内容按需调用，而非全量带入。

调用方式	说明
语义匹配	涉及某个领域时，系统自动调用该领域的相关条目
用户指定	用户明确要求："参考关于 XX 的内容"
关联触发	出现某个标签时，系统调用该标签下的相关内容
不调用	与当前话题无关的信息不参与上下文

4.6 主动提示机制

存储空间不仅是被动响应的仓库，还应具备主动提示的能力。

反向提示机制：系统定期扫描存储空间，检测用户当前行为与已确认原则的偏离。当偏离发生时，系统主动提示用户：

"三个月前确认了 X 原则，但最近的决策方向与之不同。需要重新评估吗？"

这个机制让存储从"冷存储"变成"活资产"——系统不仅被动响应，还会主动出现。

第五章：信息主权

5.1 三个核心原则

第一，用户决定存什么、不存什么。

确认机制确保了只有经过用户确认的内容才能进入存储空间。系统可以建议"这条值得存"，但最终决定权在用户手中。

第二，用户的信息用户可以带走。

用户可以随时查看、修改、删除存储空间中的任何条目。用户可以完整导出全部存储信息。存储信息不属于任何平台——它属于用户。

第三，用户的信息用户看得见。

存储空间中的每一条内容，用户都能看到它的来源、确认时间、信任级别、调用次数。没有"系统记住但用户不知道"的灰色地带。

用户对存储空间中的每一条内容拥有完整的控制权——可以查看、修改、删除。用户始终是信息的最终决定者。

5.2 为什么信息主权重要

用户越深度使用系统，产生的存储信息越多——如果存储信息不能自由迁移，用户就被平台锁定了。

信息主权不是附加功能，是这套系统的底层原则。

第六章：实施路径

6.1 阶段一：手动版（立即可行）

用现有工具模拟出这套系统 80%的效果：

系统组件	当前替代方案	效果
模块 A(协作)	固定的对话环境, 每次开工前把存储文档喂进来	约等于带存储的协作系统
存储空间	结构化文档（按本框架的字段格式存储）	手动维护，但结构化效果一致
模块 B(独立)	全新的对话环境，不喂任何资料	天然就是无存储的独立系统
确认机制	系统整理内容后输出，用户确认后存入文档	手动执行筛选
冲突检测	新存入前手动检查是否与已有条目矛盾	效率低但能保证质量

手动版的价值不是"凑合"，是"定义标准"。先用手动方式跑通逻辑，自动化才有方向。

6.2 阶段二：半自动版（近期目标）

- 系统在交互中主动识别高价值片段，生成结构化摘要卡片
- 推送给用户一键确认/修改/拒绝
- 确认后自动写入存储空间
- 新条目入库前自动检测与已有信息的冲突
- 对话开始时根据主题自动调用相关内容

- 时效到期自动提醒用户续期或归档
- 系统主动扫描存储空间，检测偏离并提示

6.3 阶段三：产品化版（远期方向）

- 存储空间内置在产品中，提供可视化管理界面
- 用户可浏览、搜索、编辑、删除、导出所有条目
- 每条信息标注来源、确认时间、调用次数
- 支持信息迁移——存储信息可以导出并在其他平台导入
- 双模块通过产品内切换实现
- 存储关系图谱——可视化展示信息之间的联系
- 主动提示机制——系统在检测到偏离时主动提示

6.4 演进原则

原则	说明
先定义标准，再做产品	先把"什么值得存"的规则跑通
先手动验证，再自动化	手动版跑通了，才知道自动化该自动什么
先少存，再多存	质量比数量重要
信息主权不可妥协	无论产品形态怎么变，用户对信息的控制权不动摇

结语

通用能力是所有人的工具——它强大、快速，但它不"记得"与你协作的历史。

具备存储空间的系统是持续的协作伙伴——它与你共同形成可复用的认知资产。

区别不在于基础能力本身。区别在于存储空间里的内容是与用户共同形成的——外部没有，别的系统没有。

一个不知道协作历史、不记得之前确认过什么、不了解已形成的方法论的系统，再强大也只是一个"能力很强的通用工具"。而一个记得、会调用、能积累的系统，才是持续的协作伙伴。

传统的软件用久了会折旧。但信息存储系统是越用越有价值的——使用时间越长，存储空间里的内容越丰富、越精准。系统越来越适配用户的需求。

记忆，是系统从工具走向伙伴的路径。

当与系统的交互不再是一次性的"见面即忘"，而是持续积累的"可调用信息"时，系统就不再是工具，而是伙伴。

附录 A：信息条目模板

标题：[一句话概括这条信息]

背景：[在什么场景下、什么时间产生的]

结论：[核心观点或决策，2—3 句话]

信息类型：[规则型 / 感悟型 / 流程型 / 事实型]

关键参数：（选填）

[涉及的数字、标准、规则]

可复用内容：（推荐填写）[话术/模板/流程/方法等]

信任级别：[确认级 / 参考级 / 待验证]

版本号：v[X.X]

时效条件：（选填）[何时应被标记为过时]

关联标签：（选填）#[标签 1]#[标签 2]

关联信息 ID：（选填）[与此信息相关的其他信息 ID]

创建时间：[YYYY-MM-DD]

最后调用：[系统自动记录]

附录 B：术语表

术语	定义
协作记忆窗口	持续的对话环境，上下文自然积累，但细节会随时间模糊。
记忆空间	经过确认的、结构化存储的永久信息库。
独立判断窗口	不加载记忆空间的独立对话环境，提供不受已有信息影响的独立判断。
确认性信息	经过用户确认、可作为系统推理基础的信息。与"探讨性信息"相对。
探讨性信息	未经确认、仅在当前对话环境中存在、不纳入长期存储的内容。
信任级别	确认级、参考级、待验证三级。
信息类型	规则型、感悟型、流程型、事实型。
人机协同确认	系统识别→结构化整理→用户确认→存入记忆空间。
冲突处理	按来源可靠性 > 用户指定 > 时间优先 > 信任级别优先处理。
认知代谢	通过结晶与修剪维持高信噪比。
信息主权	用户拥有完整的知情权、控制权、导出权和删除权。
主动提示	系统定期扫描存储空间，主动检测偏离并提示。

关于我们

东莞市意图共鸣科技有限公司

成立于 2026 年 3 月，位于东莞松山湖

致力于将 AI 能力封装为普惠的基础设施

联系方式：contact@xinirp.com

官方网站：www.xinirp.com

© 2026 东莞市意图共鸣科技有限公司。保留所有权利。

本白皮书所述认知架构、核心机制与技术方案已提交发明专利申请并获受理。

学术研究、媒体报道可自由引用，须注明出处。