

智能体时代的新基础设施

2026 爱分析·中国记忆市场研究报告

08. 2026

目录

报告综述 3

关键结论 3

1. 市场定义 7

1.1 市场认知的误区 7

1.2 记忆的定义 8

1.3 记忆的四项核心能力 9

1.4 记忆类型 9

2. 记忆市场的独立性 12

2.1 记忆不属于模型 12

2.2 记忆独特的数据门槛 13

3. 市场规模 16

3.1 记忆市场的驱动因素 16

3.2 记忆市场全景地图 17

3.3 市场规模预测 19

4. 竞争格局 22

4.1 五类厂商分类图谱 22

4.2 五类厂商分析 23

4.3 竞争象限 27

4.4 重点厂商评估 29

5. 未来趋势 34

5.1 从经验沉淀走向决策沉淀 34

5.2 记忆即服务 34

5.3 端侧与本地记忆、数据主权 35

5.4 记忆市场走向标准化 35

关于爱分析 37

研究与咨询服务 37

联系我们 37

法律声明 38

报告综述

2026年上半年，记忆从一个技术概念变成了一个行业赛道。

5月底，腾讯云发布Agent Memory，长期记忆准确率从48%拉升至76%。7月，阿里云在数据库产品线密集布局记忆能力。同时，火山引擎开源OpenViking，记忆驱动下Token消耗成本降低91%，GitHub Star突破2.1万。

众多厂商融资同步加速，记忆张量完成亿元级Pre-A轮融资，红熊AI完成数亿元A+轮融资，投后估值近30亿元，资本开始用真金白银投票记忆赛道。

市场热闹非凡，但对记忆的认知莫衷一是，仍处于早期混沌阶段，投资人难以判断边界，企业甲方难以评估价值。

因此，这个新生的行业需要一个正本清源的研究。

本报告范围内所提及的记忆，**特指智能体记忆，也就是Agent Memory，是智能体在长期交互中沉淀、动态更新并持续指导决策的经验资产，不再与知识库、上下文窗口等概念混用。**

爱分析通过本报告，试图回答以下三个问题，**记忆到底是什么，记忆为什么能独立成市场而不是模型的附属功能，记忆市场有多大、有谁在参与竞争。**

关键结论

记忆是智能体时代独立且高速增长的基础设施赛道。爱分析测算，中国记忆市场规模将从2025年的14.4亿元增长至2030年的642.5亿元，复合增长率超过110%。其中记忆操作系统到2030年规模达218.5亿元，约占记忆市场的34%，是增速最快的细分环节。

未来，记忆市场将沿四个方向继续演化。其一，从经验沉淀走向决策沉淀，记忆由记录工具升级为

决策资产。其二，记忆即服务，记忆能力从智能体内部模块变为可独立调用的服务。其三，端侧与本地记忆兴起，数据主权诉求推动记忆向端侧与本地迁移。其四，记忆市场走向标准化，接口、格式与评估体系的统一驱动记忆实现跨平台迁移。

竞争格局上，五类玩家会师记忆赛道，包括云大厂、模型厂商、独立记忆平台、企业记忆应用厂商、消费记忆产品。爱分析基于20余家厂商调研建立竞争象限，横轴衡量技术能力、纵轴衡量市场影响力，并将相关厂商落入四个象限。

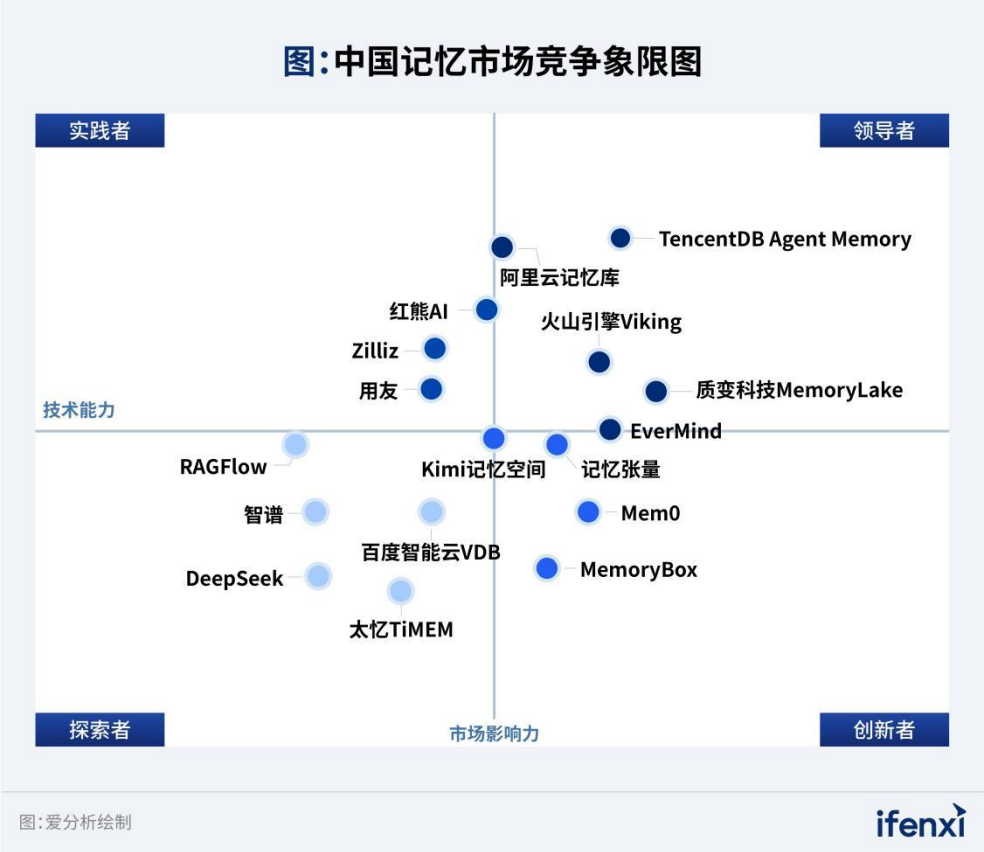


图1：中国记忆市场竞争象限图

领导者象限里，腾讯云Agent Memory与质变科技代表了两种发展路径。腾讯云依托云生态与企业级合规体系，记忆产品已成型且商业化能力强，是云大厂路线的代表。质变科技作为独立记忆平台，靠自进化能力与引用链溯源治理等挤入领导者，证明技术深度足以让独立厂商突破云大厂的生态壁垒。

创新者象限里，Kimi记忆空间凭长上下文基因切入，以记忆模型判断该记什么，体验透明可控。记忆张量则以MemOS记忆操作系统形态开路，在国央企落地并取得华为哈勃、荣耀战投的Pre-A轮支持。

实践者象限里，红熊AI是典型代表。其已有500余家客户、完成数亿元A+轮融资且投后估值近30亿元，商业化能力突出。

探索者象限里，百度智能云以向量数据库模块形式提供记忆能力，嵌在千帆平台内部场景，尚属云厂商对记忆的早期试探，产品独立性与市场落地都还在起步。

基于竞争象限图，可以梳理出记忆市场的竞争逻辑。领导者象限主要还是由云大厂主导，它们凭基础设施、客户与商业化能力占据头部地位；独立记忆平台也可以依靠技术深度也可以挤入头部领导者；模型厂商则主要聚集在创新者象限和探索者象限，其方向明确但产品化尚未成型。

CHAPTER

01

市场定义

1. 市场定义

在今天的AI行业里，记忆是被滥用最严重的一个词。有人认为记忆就是知识库，有人认为记忆是模型的超长上下文窗口，有人认为记忆是历史对话的临时缓存。同一个词指向完全不同的技术实现，投资人难以判断边界，企业甲方难以评估价值。

爱分析认为，市场混乱的根源，是记忆一直没有被认真定义。本章先纠正三个记忆的普遍误区，再给出记忆的明确定义，最后建立一套可用的分类框架。

1.1 市场认知的误区

市场对记忆的普遍认知，存在三个把记忆降格理解的误区。

第一个误区，**把记忆等同于知识库、RAG或超长上下文，这是最普遍的误读。**知识库存放的是通用事实，RAG解决的是检索增强，超长上下文延长的是单次对话的窗口，三者服务的都是让模型在当下这一次回答里调用更多信息。

记忆不止于此，记忆还要记录智能体在过去做过什么、结果如何、哪些做法有效，并让这些经验影响未来的决策。知识可以离线灌入，经验只能靠时间积累。把记忆当成知识存储，等于抹掉了记忆里最核心的时间维度。

第二个误区，**把记忆当成基础模型能力的一部分。**有一种看法认为，只要基础模型把上下文窗口拉得足够长，或者参数足够大，记忆问题就自动解决了。这背后是把记忆当成模型的一个附属功能。

爱分析认为这种看法站不住。模型记忆解决的是单次推理中信息的即时调取，会话结束即清空，但记忆要跨越会话、跨越任务、跨越版本长期存续。前者是算力问题，后者是系统架构问题。把记忆塞进模型能力，会掩盖一个关键事实，即记忆是一项需要独立设计、独立运维、独立演进的

基础设施能力。

第三个误区，**把企业级记忆与消费级记忆当成两个完全不同的赛道。**市场习惯把企业记忆描述为知识管理、流程沉淀、组织协同，把消费级记忆描述为个人助理的偏好与习惯。

爱分析认为，这二者在底层同源。组织沉淀的项目经验与个人累积的使用习惯，本质是同一类东西，都是智能体在长期交互中形成的经验资产。二者的差异只在经验资产的归属、权限与治理强度上，记忆本身的原理并无不同。把二者割裂，会让厂商错失同时在两个市场竞争的机会。

1.2 记忆的定义

记忆的本质，是智能体在成长过程中持续积累的经验资产。爱分析定义为，记忆是智能体在长期交互中形成、存储并动态更新的经验资产，支撑智能体跨会话、跨模型复用过往信息，并做出更有效的决策与行动。

要理解这个定义，有三件事要说清楚。

第一，知识与记忆的边界。知识是静态的事实，可以长期稳定的调用；记忆是只属于这个智能体或组织的经验，越用越多、越用越准。记忆记录的，是这个智能体、这个组织在具体情境里经历过什么、学到了什么。误把知识系统当记忆系统，正是误区一的来源。

第二，记忆的内涵远不止于模型记忆。模型记忆是单次推理的上下文，让模型在当下这一轮里想得更全，记忆是跨会话的经验积累，让智能体在更长的时间尺度上变得更强。模型记忆服务于思考，记忆服务于成长，前者只是后者在推理瞬间的一小片工作区。

第三，记忆的构成是复合的。它既包含个体经验，也包含组织经验；既包含事实性信息，也包含行为模式、决策逻辑与长期反思。记忆不是简单的数据留存，它同时记录了发生过的事实，以及智能体由此形成的判断习惯、行动方式和复盘结论。

1.3 记忆的四项核心能力

判断一套记忆系统是否成立，要看它能否同时做到理解、进化、可用、可信。

多模态能力，解决的是**理解**。记忆应能统一理解文本、图片、音频、视频以及业务系统数据。智能体面对的不是单一格式的输入，一份工业图纸、一段设备时序日志、一条语音指令，都可能成为需要沉淀的经验。记忆系统如果不能把这些异构信息统一编码进同一条经验，就只能在各种格式的孤岛里存碎片。

比如，制造业企业可以把多模态图纸和设备时序日志固化为可复用的工业记忆资产，这样跳出了文档检索的层面，还可以支撑设备故障预测、沉淀运维经验、让新员工快速上手。

自进化能力，对应的是**进化**。记忆应能自我更新、压缩、反思、遗忘与迁移。记忆不是只增不减的日志，随着经验积累，需要淘汰过时信息、压缩冗余细节、把多次失败的教训沉淀为反思，并把在一种任务上学到的经验迁移到新任务。

产品能力，解决的是**可用**。记忆应能持续读写、全生命周期管理、跨平台部署，并处理版本与冲突。一套记忆系统要被真实业务采用，只解决技术原理远远不够，它需要支持高频的持续读写，对记忆从产生到归档的全生命周期做管理，支持跨平台迁移，并处理多人协作下的版本与冲突。

治理能力，解决的是**可信**。记忆应具备权限管理、隔离、溯源与可信能力。当记忆承载了组织的经验资产，谁可以读、谁可以改，哪些经验属于哪个项目、某条结论从何而来，就都成了不可回避的问题。

1.4 记忆类型

弄清了记忆需要具备哪些能力，再看记忆本身有哪些类型。给记忆分类，是为了让市场有一个共同的坐标系。

首先，记忆可以按内容，分为**语义记忆、情景记忆和程序记忆**三种类型，这一分类出自认知科学领域的CoALA框架。

语义记忆，回答我知道什么，是客观知识与事实；情景记忆，回答我做过什么、结果如何，是经历与经验；程序记忆，回答我会怎么做、为何这样做，是决策逻辑与行为模式。

需要说明的是，这三种记忆是并行的三种类型，一个智能体可能同时拥有大量语义记忆、少量情景记忆和逐步成型的程序记忆，三者的区别只在内容属性，与成熟度无关。

其次，**记忆还可以按归属主体，分为个体记忆、项目记忆和组织记忆三层。**

个体记忆，属于单个智能体或单个用户，记录其独立形成的经验；项目记忆，属于一个具体项目或任务，记录项目执行过程中沉淀的集体经验；组织记忆，属于整个组织，记录组织层面长期积累、可跨项目复用的经验与做法。

目前，市场上已有记忆厂商推出对应的产品，映射这三级结构。

CHAPTER

02

记忆市场的独立性

2. 记忆市场的独立性

定义了记忆是什么，下一个问题是，记忆为什么能成为一个独立市场，为什么它不该只是基础模型的一个附属功能。

原因分两层：第一层，记忆承载的经验不属于任何模型，天然与多模型配合；第二层，做记忆有独立的数据能力门槛，和模型的数据治理完全不是一个事情。这两层原因合在一起，就是记忆市场得以成立的依据。

2.1 记忆不属于模型

模型负责智能体的思考，记忆负责智能体的成长。

模型决定一个智能体能做什么，记忆决定这个智能体能做多好。模型解决的是推理能力，同一个问题，模型越强，当下这一轮回答得越好；但智能体的价值不只在单次回答，而在长期交互中越用越懂一个组织、越用越懂一个用户。

想象两个智能体，用的是同一个模型，一个沉淀了这家企业过去两年的项目经验，一个没有，面对同一个新问题，两者的表现差距会非常大。造成这个差距的，是记忆。模型负责把当下这一轮的思考做好，记忆负责把过去所有的经验沉淀下来，让下一轮思考站在更高的起点上。

一个组织花了两年沉淀的项目经验，属于这个组织自己，不属于任何一家模型厂商。一个用户长期累积的使用偏好，属于这个用户自己，也不属于任何一家模型厂商。模型可以换，可以升级，可以今天用A家明天用B家，但这些经验资产必须跨模型、跨产品长期存续。如果记忆依附在某一个模型上，换一次模型，企业就等于把过去所有的经验重新清零一次，这没有一家企业承受得起。

跨模型、跨产品、长期存续，这正是一项基础设施该有的特征。基础设施被多方依赖，不随上层应用的更换而更换，数据库不会因为换了一家上层软件就换掉。记忆沉淀的是组织级、用户级的长期

经验，同样要求跨模型、跨产品长期存续。所以记忆的定位，只能是一项独立的AI基础设施。

爱分析判断，**企业最值钱的经验资产，不可能押在一家随时可能换掉的模型供应商身上，记忆为什么能成为一个独立的市场，根源就在这里。**这一判断，是记忆市场得以成立的第一块基石。

2.2 记忆独特的数据门槛

记忆能独立成市场的第二个原因，是它有**数据门槛**。记忆要处理的是真实业务里的脏数据，和模型训练时用的干净语料完全是两回事。这道数据门槛，模型训练的过程解决不了。

具体而言，记忆厂商需要解决三个核心的数据问题。

第一项，**数据预处理与解析**。记忆要沉淀的是真实业务里的异构数据，比如PDF里的嵌套表格、跨页表格，又如机械图纸中的复杂标注，基础模型在解析这些文件时精度会急剧下降。记忆厂商大多需要自研解析小模型，来专门处理数据。

第二项，**多模态数据管理**。记忆面对的是海量异构数据，文本、图片、音频、视频、业务系统数据混在一起。怎么存、怎么组织、怎么压缩、怎么治理，是一个完整的工程问题。

数据规模越大，这层能力的门槛越硬，而且这些数据要长期存续、随时可调，对存储结构和治理机制的要求，与模型推理时的超长上下文相比，是两个完全不同的工程体系。

第三项，**记忆计算**。这是记忆最独特的地方，召回、重排、更新、反思、遗忘，这些计算动作都以经验为单位，不以单条数据为单位。

举一个简单的例子，一次多跳检索要回答的，是一串结论之间的因果链，孤立的一句事实回答不了这类问题。记忆系统必须记住这条因果链，才能在下次被问到相关问题时，把前因后果一起调出来，不能只返回一句孤零零的结论。

这种以经验为单位的计算，模型算力再强也替代不了。算力解决的是算得快不快，记忆计算解决的

是记不记得住因果，两者是两回事。

以上三项能力叠在一起，构成了记忆厂商自己的护城河。谁能跨过数据门槛，谁才能在这个市场里站稳脚跟。

CHAPTER

03

市场规模

3. 市场规模

记忆作为一个独立市场，规模及增长空间是绕不开的问题。它为什么能长这么大，又能长到多大，决定了这条赛道的分量。

3.1 记忆市场的驱动因素

推动记忆市场高速增长的根本原因在于，**AI产业正在从知识驱动阶段进入经验驱动阶段。**

大模型时代的核心价值在于知识生产，模型通过海量训练数据学习世界知识，并通过推理能力完成知识调用。随着模型能力持续提升，知识获取成本正在快速下降，越来越多智能体将具备相似的知识水平和推理能力，模型能力差距也将逐步缩小。在这种背景下，经验开始成为决定智能体竞争力的关键因素。

知识决定能力下限，经验决定能力上限。

因此，经验正在成为智能体时代新的生产资料，而记忆则成为管理和利用这一生产资料的核心基础设施。

经验生产资料化，**首先体现在智能体能力演进过程中。**

当前，大部分智能体仍然依赖模型实时推理完成任务。记忆的价值在于将一次次任务执行过程转化为长期经验资产，任务目标、执行路径、决策逻辑、结果反馈以及优化策略，都能够被持续沉淀，并在后续任务中重复利用。

随着经验不断积累，智能体执行效率持续提升，错误率持续下降，并逐步形成自主优化能力，记忆因此成为智能体实现自主进化的重要基础。

值得一提的是，当智能体能够直接调用历史经验完成决策时，也能够显著减少推理token的消耗。

经验生产资料化，**同样体现在企业AI建设和消费级AI产品竞争过程中。**

企业最核心的竞争优势往往来自经验与Knowhow，业务经验、行业经验以及决策经验，长期掌握在少数核心员工手中，当员工流失时，经验也会随之流失。经验难以复制、难以传承、难以规模化利用，始终是企业经营中的重要问题。

随着数字员工逐渐承担更多工作任务，企业开始尝试将组织经验转化为可继承的数字资产，组织经验资产化正在成为企业智能化升级的重要方向。

经验生产资料化**还体现在消费级AI产品竞争逻辑的变化之中。**

随着模型能力趋于同质化，用户对于AI产品的评价标准正在发生变化，用户关注的不再只是模型是否聪明，而是产品是否真正理解自己。未来最具竞争力的AI产品未必拥有最强模型，但一定拥有最完整的用户记忆。

3.2 记忆市场全景地图

随着经验逐渐成为智能体时代的新生产资料，围绕经验资产的存储、管理和应用，一个完整的体系正在形成。

根据爱分析定义，记忆市场主要由**记忆硬件、记忆操作系统和记忆应用**三个层级构成。其中，记忆硬件负责经验存储，记忆应用负责经验消费，记忆操作系统则负责经验管理与调度。



图2：记忆市场全景地图

从市场规模角度看，三层市场都将受益于智能体普及带来的增长红利，但从价值分布角度看，记忆操作系统有望成为未来增长最快、价值占比最高的核心环节。

当前，市场对于记忆的理解仍然停留在存储层面。随着智能体开始承担越来越复杂的任务，记忆真正的挑战并非存储能力，而是管理、编排与调度等计算能力。

因此，记忆系统不仅需要存储能力，更需要计算能力，未来的记忆操作系统，需要具备记忆提取、压缩、关联、编排、反思等诸多能力。

爱分析认为，**记忆操作系统将位于价值链核心位置，一方面向下连接硬件基础设施，另一方面向上服务各类应用。**

3.3 市场规模预测

爱分析测算显示，**中国记忆市场规模将从2025年的14.4亿元，增长至2030年的642.5亿元，期间复合增长率超过110%。**



图3：2025-2030年中国记忆市场规模预测

按记忆的市场分布来看，记忆操作系统市场规模在2030年将达到218.5亿元，在记忆整体市场中占比34%，同时其市场规模复合增速也是最快的，记忆操作系统五年复合增速超过152%，而同期整体记忆市场的复合增速为114%。



图4：2025-2030年记忆操作系统市场规模

作为独立的基础设施市场，爱分析测算，到2030年，记忆在智能体基础设施中的占比将接近26%，成为仅次于模型的第二大独立市场。



图5：2025-2030年记忆在智能体基础设施中的占比

CHAPTER

04

竞争格局

4. 竞争格局

爱分析在调研20余家厂商后，将当前中国记忆市场的参与者归纳为五个类别。有意思的是，大家都挂着记忆的招牌，做的却是完全不同的生意。

4.1 五类厂商分类图谱

云大厂从数据库出发，想把记忆做成存储基础设施的一次AI升级，把熟悉的客户顺手迁过来。**模型厂商**则把记忆当成让自家模型更聪明、消耗更多Token的附属能力，记忆好不好用，关键看能不能喂饱模型。

独立记忆平台强调不站队任何模型，把记忆抬到与基础模型并列的独立数据主权资产，靠中立性撕开市场缺口。**企业记忆应用厂商**从业务效果倒推，把记忆当作推理的前置引擎，谁能帮营销多转化，谁就赢。**消费记忆产品**最激进，干脆让用户为记忆本身付费，把后台能力变成前台商品。

这五类玩家资源禀赋不同，下注的方向也不一样。



图6：记忆市场玩家分类图谱

4.2 五类厂商分析

4.2.1 云大厂

云大厂是记忆市场最为积极的玩家，特别是数据库团队。腾讯云Agent Memory由TencentDB数据库团队主导，阿里云也是在数据库产品线上寻求记忆产品的突破，而火山引擎年初开源的项目OpenViking，同样源自VikingDB向量数据库团队。云大厂的核心出发点是一致的，记忆是数据库能力的自然延伸。

腾讯云数据库副总经理罗云判断，未来的数据库需要管理智能体组织，主智能体与子智能体之间的权限授予、主备切换、数据多版本管理，这些过去是数据库的边角料需求，现在变成了核心命题。阿里云数据库资深产品专家葛双博的判断更加形象，模型决定AI多聪明，数据库决定AI能跑多远，AI的上限由模型决定，下限由数据库决定。

爱分析认为，**不仅是云大厂，现在的数据库厂商预计未来都会把自己定义为AI记忆公司，在数据库厂商视角中，记忆本质上是AI时代的数据存取系统。**

从竞争维度来看，云大厂在记忆赛道上的优势也是从数据库延伸而来，其底层向量数据库和海量数据存储经验，是其他厂商难以短期追赶的积累。同时，云大厂的数据库拥有成熟的企业级安全合规体系，降低了企业采购时的合规顾虑。

但是，云大厂做记忆也面临两个核心约束。第一，视角天然偏向存储与检索，关注的是记住什么、怎么存、怎么取，而对记忆的全生命周期管理，包括什么时候该记、该忘、该合并、该覆盖、如何迭代进化，关注度不足。第二，资源投入不足，有鸡肋嫌疑。云大厂的记忆团队通常是数据库部门的子项目，人力配置和战略优先级并不高，这与记忆在未来AI Infra的地位并不相符，极可能成为鸡肋。

云大厂是当前记忆市场上产品化程度最高、最容易被客户理解和采购的品类形态，但它解决的核心问题是存储基础设施的AI升级，而非推动智能体更好的积累经验以及自我进化。

4.2.2 模型厂商

模型厂商做记忆的出发点和云大厂完全不同，他们的出发点是为了让模型能力更强，Token消耗更多。更重要的是，模型厂商都将持续学习作为下一个跃迁能力点，而记忆就是实现持续学习的必备基础。

比如，智谱做GLM-5.2规划时，就将记忆作为要关键突破的技术。智谱创始人唐杰在2026年初的内部信中，明确将记忆列为下一代模型的核心命题，已经广泛使用近10年的Transformer架构显露出不足，包括超长上下文的计算开销、记忆机制、更新机制等。

模型厂商这条路径的核心逻辑，是把记忆定义为模型的附属增强能力，其记忆的架构设计

天然倾向于自身的模型生态，记忆的价值通过模型或者智能体效果提升来体现。因此，在模型厂商认知中，记忆要让自身的模型更好用，记忆本身不独立于模型而存在。

从竞争维度来看，模型厂商的优势来自于模型主战场的势能溢出，尤其突出的就是数据飞轮。用户调用模型过程中，产生了大量可以用于提升记忆能力的数据，这些数据会用于下一代的记忆架构调优。

但安全合规是模型厂商最明显的软肋。他们的治理体系要求用户数据必须闭环回流、服务于自家模型，这种把所有隐私数据攥在自己手里不放开的做法，天然踩在合规的灰色地带上。

爱分析判断，**模型厂商做记忆有一个解不开的结，记忆系统从一开始就被绑死在自己的模型生态上，想把它用到别家模型或别家应用里，几乎没有空间。**

4.2.3 独立记忆平台

独立记忆平台的出发点是另一个思路，他们不把记忆视为某个数据库或某个模型的附属功能，而是**将其定义为与基础模型并列的独立AI Infra。**这条路径的核心主张，可以归纳为三点。

第一，记忆的中立性本身就是最大的价值。模型的边界在无限扩张，不可抑制的吸收全人类的知识、个人的经验、企业的Knowhow，这对于商业世界造成巨大的破坏。记忆的中立性，使其成为AI时代最强的数据主权象征，避免个人与组织被模型蒸馏成为废弃物。MemVerge联合创始人范承工表示，无论是个人还是企业，都应该有自己的一套记忆系统。

第二，记忆是AI时代最先提出自进化的代表。质变科技CEO占超群在交流中表示，记忆是智能体进化的基础设施。记忆每天都在接收新的数据，通过持续的反思、验证、遗忘、冲

突管理等方式自进化。相比模型的数据飞轮，独立记忆平台的数据飞轮更加完美，因为其进化过程发生在终端的个人、组织一侧，是自发完成的，不需要人的参与。

第三，独立记忆平台的兼容性强，可以和任何模型、任何应用组合，这种架构自由度是其他类型玩家无法提供的。

从竞争维度来看，独立平台的优势在不同维度上各有体现。首先，独立记忆平台都在训练专注于记忆的小模型。参考人脑，小参数模型更适配记忆体系。一方面，推理调用记忆时响应速度必须快，大参数模型会导致延迟。另一方面，从数据主权角度出发，记忆未来的趋势是端侧或者本地化的模型，小参数模型算力需求低、更加契合需求。

其次，产品能力方面，独立平台覆盖了记忆全生命周期。以质变科技为例，其workspace、project、actor三层架构实现了组织记忆与个体记忆的协同。在企业内部，法务部门、研发部门、运营团队各自的数据沉淀在不同的project中，但在同一个workspace下协同演进。这套组织记忆堆叠体系，目前在其他记忆产品中几乎看不到同类设计。

4.2.4 企业记忆应用厂商

企业记忆应用厂商走了一条更直接的路，**从业务效果反向定义记忆。**他们的核心出发点是，更好的推理效果来自于记忆优化，由此可以彻底颠覆传统企业应用软件。通过先调记忆再调模型，执行效果大幅提升，Token消耗也显著降低。

红熊AI温德亮在访谈中表示，这个行业最后一定不是讲谁技术最好，而是讲谁商业化最好。这条路径的核心优势正是在于商业验证最直接。企业客户不关心记忆的技术架构，只关心记忆驱动的客服机器人回复效果是否有提升、营销转化率是否有提升，以业务效果为唯一检验标准。目前，客服、营销、教育等场景在率先落地企业记忆应用。

爱分析判断，这条路径的局限在于场景的泛化性。以垂直场景倒逼记忆能力建设，可能在单个场景做到极致，但跨场景的抽象度和可复用性不足。在客服场景中验证的记忆驱动推理逻辑，能否迁移到金融风控、工业运维、医疗诊断等场景，尚未被证明。

4.2.5 消费记忆产品

消费记忆产品走了一条更激进的路，不让记忆隐身为后台基础设施，而是**让它成为用户愿意直接付费的前台产品。**他们的核心出发点是一个直觉式的判断，一个人很大程度上就是其记忆堆积而成的，为记忆付费是消费者最自然的需求。

消费记忆产品主要有软件、硬件两种形态。软件形态的核心理念，是将各个AI应用中的对话和智能体轨迹统一回流到一个记忆中心，当用户发起需求时，基于记忆重写上下文、分发至合适的AI应用，最终交付最优结果。消费记忆产品的硬件形态，是always-on的音频或者视频随身设备。通过记录用户全天候的音视频数据，成为最懂用户的精神陪伴者。

显然，消费记忆应用的优势在于用户粘性。当用户将跨平台、跨模型的AI交互历史全部沉淀到一个记忆软件或者硬件中，迁移成本会随着时间指数级上升。

爱分析判断，消费记忆硬件产品在长期视角中成长空间更大，因为其记忆来自于物理世界而非数字世界，是当下最为稀缺的数据资产。但硬件产品不止依赖记忆技术本身的发展，也受制于硬件本身。比如，always-on视频设备的续航、端侧多模态数据处理能力等。

4.3 竞争象限

在完成市场参与者的类型划分之后，需要进一步评估各厂商的竞争态势。本章节采用技术能力与市场影响力的二维竞争象限，来逐一评估市场主要玩家。

该评估框架的构建，基于爱分析对中国记忆市场的持续跟踪研究。研究团队对20余家厂商

进行了调研，系统采集了其在技术架构、产品成熟度、场景落地与商业化等方面的数据，并据此确立评估维度与权重。

竞争象限的横轴为技术能力，由多模态(20%)、自进化(30%)、产品(25%)、治理(25%)四个子维度加权合成。纵轴为市场影响力，由场景(50%)与商业化(50%)两个子维度等权合成。

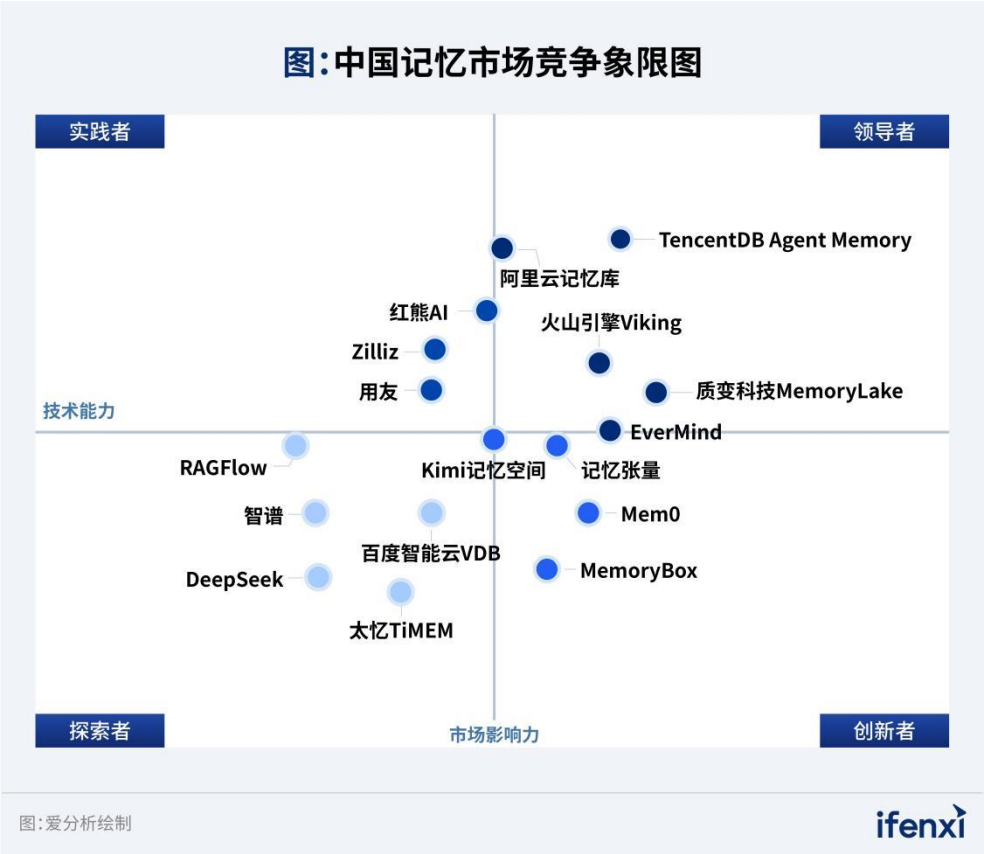


图7：中国记忆市场竞争象限图

依据两个维度的得分，厂商落入四个象限，分别对应不同的竞争态势。

领导者在两个维度上均居前列，技术能力领先且已实现规模化商业落地，构成当前市场的基准参照。

创新者技术能力突出但市场影响力有限，多在长上下文、自进化或记忆检索等特定方向形成优势，商业化与场景覆盖尚处早期，存在向领导者迁移的可能。

实践者市场影响力较强但技术能力相对薄弱，通常依托既有云或大模型业务的客户基础延展记忆能力，底层技术积累较浅或依赖外部方案，短期具备份额优势，中长期面临技术补强压力。

探索者两个维度均处早期，以初创团队与开源项目为主，产品化与商业化尚未完成，是市场未来格局的潜在变量。

爱分析认为，**记忆市场的终局，一定是与推理深度融合的。**这里的推理不光是逻辑推理，也包括情感等维度。比如，记忆加逻辑推理代表着更好的工作成果，记忆加情感推理代表着更好的生活状态。

因此，记忆的技术与产品的优劣，不能仅仅比拼记忆召回效果、Token消耗量，而应当围绕任务执行的成功率来衡量。当记忆市场的benchmark转换为任务效果比拼时，才是记忆竞争真正意义上的开局。

4.4 重点厂商评估

4.4.1 腾讯云，云大厂代表

腾讯云是云大厂这条路径的代表，**它的强项和短板，都来自同一个出身，数据库。**

产品优势层面，Agent Memory产品由腾讯云数据库团队从底层自研，基于腾讯云向量数据库构建，是一套独立的记忆管理底座，API和控制台完善，与CodeBuddy集成，企业可以开箱即用。

技术上，它构建了四层记忆架构，L0存原始会话，L1拆解为原子信息并建向量索引，L2聚合成场景化文件，L3沉淀为个性化画像。腾讯云数据库副总经理罗云这样解释，L3是强制披露直接植入上下文，L2采用渐进式披露按需读取，L0和L1通过向量与关键词检索。这套

分层设计让它在PersonaMem评测中把长期记忆准确率从48%拉到76%，用户事实召回率从不足30%提升到79%以上。

此外，依托腾讯多模态大模型的生态，腾讯云记忆产品的多模态基础解析能力强。治理合规上，腾讯云有成熟的企业级多租户体系，合规能力是现成的，降低了企业采购时的顾虑。场景上，C端WorkBuddy加B端企业智能体，多个场景都能铺开。

腾讯云记忆产品的劣势集中在优先级低、资源不足。记忆产品团队是数据库部门的子项目，人力配置和战略优先级都不高，有鸡肋嫌疑。对腾讯云来说，记忆只是一个增量产品，不是核心战略，资源投入的持续性和力度存疑。

4.4.2 Kimi，模型厂商代表

Kimi是模型厂商这条路径的代表，**它的优势来自长文本的技术底蕴和C端的产品体验，短板则体现在多模态和产品定位。**

先说优势。Kimi做记忆的底气，是其在长上下文上的长期积累。创始人杨植麟表示，上下文长度就是大模型的内存，是决定大模型应用的关键因素之一。Kimi既是全球首个支持20万汉字输入的智能助手，又开启了200万字无损上下文的内测。

产品层面，Kimi记忆空间体验好、透明可控，用户能清楚看到记忆里存了什么、为什么存、可以怎么改，这种透明度在记忆产品里并不多见。治理上，Kimi做了差异化，用户完全自主管理记忆，数据不用于模型训练，缓解了模型厂商最受诟病的合规软肋。自进化机制上，Kimi用单独的记忆模型判断该记什么，已经有了一套更新的逻辑。

再说劣势。多模态层面，Kimi的记忆目前还是纯文本，对图片、音频、视频等异构数据处理不了，限制了它能服务的场景。产品定位上，记忆不是独立产品，只是Kimi的一个功能，

以C端个人助手场景为主，没有独立商业化，依托会员体系和API变现。

4.4.3 质变科技，独立记忆平台代表

质变科技是独立记忆平台的代表性厂商，定位为中立的企业级生产力记忆基础设施，其能力可从五个维度理解。

一是产品理念。质变科技认为，记忆是有复利的。在生产力智能=思考力×记忆力这一框架中，基础模型提供通用思考力，而记忆提供可持续积累的经验、决策与状态，两者是乘法而非替代关系。

质变科技主张记忆的“复利”，让每一次任务的结果沉淀为可复用经验，从而缩短下一次任务的学习路径。这也内含了独立记忆层的中立性前提，即模型可迭代、可切换、可组合,但企业记忆应跨模型、跨Agent、跨周期复用，归属客户自己。这正是独立记忆平台区别于大厂自带记忆的根本。

二是组织记忆堆叠。大厂与开源方案大多不区分组织记忆与个人记忆，质变科技是少数把组织记忆治理做成产品的独立厂商。

质变的组织记忆架构围绕Workspace（项目容器）、Project（记忆资产）、Actor（归属主体）三层结构展开，让同一份知识按角色沉淀出不同记忆。比如，同一份产品文档，售后侧重故障排查，研发侧重版本迭代，管理层侧重GTM。不同角色各取所需，在公司层面形成记忆协同。

三是以“任务成功率”衡量记忆，而非“记住多少”。LoCoMo、LongMemEval等主流评测，考的是能不能记住，但记忆最终服务于推理，更该考察能不能靠记忆把任务做完。新的评测标准MemoryArena，将评测推进到子任务相互依赖的交互式场景。

在MemoryArena同一模型、同一评测脚本的受控对比中，质变科技在五个域的等权平均任务成功率为20.5%，高于最强对照组的13.5%，其中物理推理域优势最明显（60%vs45%）。

一个更说明问题的细节出现在渐进式多跳检索域，以相似度检索为主的Mem0子问答准确率更高（8.9%vs6.7%），但最终任务成功率却更低（15%vs20%）。这说明"记得多"不等于"做得成"。

四是可追溯、可治理的记忆，才符合企业治理。质变科技治理层的优势来自团队基因：核心团队出身阿里云数据库，长期服务B端客户。

开源项目普遍不碰治理，大厂记忆更偏C端隐私，质变科技的差异化来自B端积累。具体的治理能力体现在几个层面：一是引用链溯源，召回附带来源Project与片段级角标，决策轨迹全程可查，因为金融、合规等场景不仅要结论，更要能还原结论怎么来的；二是随身份、任务、上下文动态计算的权限模型，覆盖团队、空间、项目、角色到实体级多个维度；三是血缘级联删除，一条被删事实往往已派生出索引、摘要、规则、缓存等多个副本，质变按血缘追踪并级联清除，做到删源数据的同时删掉它造成的影响,这对合规审计尤为关键。

五是多模态数据解析，决定召回精度上限。多模态解析是召回准确率的前置条件，解析错了，后面的记忆和推理都会错。

质变科技的判断是，文档理解最难的不是OCR，而是把二维、层级化的版面还原成不丢失结构、关系、原始值与来源的记忆。它走自研解析模型路线，壁垒不在模型参数大小，而在长期积累的格式适配与数据工程。比如PDF嵌套表格、跨页表格、合并单元格等复杂版式下的处理，以及模态覆盖文本、图片、视频、表格、PDF、代码与数据库等企业知识资产的主要形态。

CHAPTER

05

未来趋势

5. 未来趋势

记忆与推理的融合是终局，未来几年，**记忆会沿四个方向演化，决策沉淀、记忆即服务、端侧数据主权、市场标准化。**

5.1 从经验沉淀走向决策沉淀

为了更好的与推理进行融合，记忆市场下一步不仅要沉淀经验，更要沉淀决策逻辑。

传统企业的信息体系，走的是“数据-知识-决策”这条链路，先把数据整理成知识，再用知识支撑决策。智能体时代，这条链路演化为“知识-上下文-记忆-执行”，记忆取代了中间的知识沉淀环节，直接服务执行。

更大的变化在后面。企业过去沉淀的是知识，现在沉淀的是经验，未来沉淀的是决策。决策逻辑、行为模式、组织经验，会取代知识成为企业最重要的资产。因为智能体不仅需要知道答案是什么，更需要理解为什么这样做、在什么情况下这样做。

比如，一个资深员工离职，带走的往往不是知识，而是做决策时的那套判断依据，为什么选这个方案、为什么拒绝那个客户。把决策沉淀下来，就是把这套判断依据留给组织，让它不随个人流失。

经验沉淀是当下正在发生的事，决策沉淀则是下一阶段才大规模到来的增量。等到记忆不仅能记住做过什么，还能记住为什么这么做，记忆就从记录工具升级成了决策资产。

5.2 记忆即服务

记忆正在从智能体内部的一个模块，变成一项独立的云服务，这个趋势可以称之为记忆即服务。

商业上，计费模式也在跟着变。过去的记忆成本被打包在模型推理的token消耗里，未来会转向记忆存储加检索次数这种独立计费方式。记忆从成本项变成了可以单独定价、单独收费的商品，这本身就是它独立成市场的信号。

技术上，跨智能体的统一记忆层正在成为方向。同一个用户在不同智能体之间共享同一份记忆，用户在这个应用里养成偏好，另一个应用能接着用，不用从头教起。记忆层一旦独立出来，就印证了本报告的核心判断，记忆是一个独立市场，而不是任何模型的附属。

5.3 端侧与本地记忆、数据主权

隐私法规在收紧，记忆正在往端侧和本地走。

欧盟的GDPR、中国的个人信息保护法，都在强化一个原则，个人数据要可管、可控、可删。记忆里存的是一个人或一个企业最核心的经验，这些数据一旦上了云，就存在被第三方处理和利用的风险。监管越严，企业越倾向于把记忆部署在本地，让数据不出设备、不出企业边界。

这个趋势在端侧已经开始落地，已有记忆厂商开始把记忆能力接入手机、家电等端侧设备，把记忆存在本地，只把必要的结果传出去。

端侧和本地记忆的背后，是数据主权。一个人或一个企业，对自己的经验拥有完整的控制权，不交给任何平台。记忆越往端侧走，独立记忆平台的中立价值就越高。

5.4 记忆市场走向标准化

记忆市场要真正长大，最后一步是标准化。

评测体系、记忆协议、记忆接口、行业基准，这些正在逐步形成。没有统一标准之前，每家厂商的记忆都有自己的格式和接口，企业一旦用了某一家，就被锁定住，换厂商的成本

极高。社区正在推动统一的记忆协议，让不同框架的记忆模块可以互操作，一个应用可以调不同厂商的记忆，一家厂商的记忆也能被多个应用复用。

标准一旦确立，记忆就不再是某一家厂商的私有技术，而是一层公共的基础设施。到那时，记忆有望成为继基础模型之后，AI时代新的基础设施层。这也意味着，谁在标准制定上占据先手，谁就能在记忆市场的下半场掌握主动权。

关于爱分析

爱分析是一家专注于AI市场的研究咨询机构，深耕AI与数字化十余年，为科技厂商提供从市场洞察到品牌营销的综合服务。

凭借深厚的行业积累与头部客户服务经验，爱分析助力科技厂商把握市场机会、塑造品牌影响力、高效触达目标客群，持续提升市场竞争力。

研究与咨询服务

技术研究

新兴技术研究，厂商能力调研，助力数字化最优决策

客户洞察

企业用户需求及实践调研，辅助制定业务与市场策略

行业研究

行业数字化趋势与实践研判，辅助业务与战略决策

商业研究

基于研究、数据和案例调研积累，辅助业务可靠落地

品牌&营销

权威背书，树立行业地位；教育市场，精准触达客户

投资研究

成熟方法论，一手数据，助力研判机会、稳健投资

联系我们

联系人：张静宜

邮箱：jingyi@ifenxi.com

手机/微信：186-1292-6091

法律声明

此报告为爱分析制作，报告中文字、图片、表格著作权为爱分析所有，部分文字、图片、表格采集于公开信息，著作权为原著者所有。未经爱分析事先书面明文批准，任何组织和个人不得更改或以任何方式传送、复印或派发此报告的材料、内容及其复印本予任何其它人。

此报告所载资料的来源及观点的出处皆被爱分析认为可靠，但爱分析不能担保其准确性或完整性，报告中的信息或所表达观点不构成投资建议，报告内容仅供参考。爱分析不对因使用此报告的材料而引致的损失而负上任何责任，除非法律法规有明确规定。客户并不能仅依靠此报告而取代行使独立判断。

北京爱分析科技有限公司2026版权所有。保留一切权利。

ifenxi

咨询/合作

微　信：ifenxi888

网　址：www.ifenxi.com

北京市朝阳区远洋国际中心E座1906

如欲了解更多爱分析精彩洞见
请关注我们的微信公众号

©北京爱分析科技有限公司2026版权所有


ifenxi
AI市场研究咨询机构