Yoma

WHITE PAPER · 2026

零售与餐饮 AI 数字化

「AI智驾」自主等级（L1～L5）· 建设验收指南

「AI智驾」= AI 数字化自主等级（非车辆智驾）

> 大模型给了「脑」；全 API 化的业务系统，才给了 AI「手脚」与「感官」。

思源 · YOMA · 2026 · www.yongma.cloud

YOMA 零售与餐饮 AI 数字化白皮书

目录

零售与餐饮 AI 数字化白皮书

- 序言
- 执行摘要

上篇 · 理论：「AI智驾」等级与 AI 数字化

- 第一章 为什么需要「等级」而不是「又一个 Agent」
- 第二章 YOMA AI 数字化方法论
- 第三章 RAL L1～L5 全景展开
- 第四章 从 L2 到 L3：商业价值分水岭
- 第五章 底座：等级升高，底座必须加码

下篇 · 实践：零售与餐饮该如何做

- 第六章 通用建设路径（两业态共用）
- 第七章 业态选域与样板闭环（零售 / 餐饮）
- 第九章 治理、风险与「什么时候不该上」
- 第十章 行动倡议
- 附录 A RAL 定级自测（10～15 题）
- 附录 B 术语表
- 附录 C 实践摘要（脱敏示例）
- 附录 D 参考文献与延伸阅读
- 免责与边界

零售与餐饮 AI 数字化白皮书

副标题：「AI智驾」自主等级（L1～L5）与建设验收指南

> **关于「AI智驾」**：借汽车智驾为人所熟知的台阶（脱脚→脱手→脱眼…），指 **零售与餐饮 AI 数字化的人机权责与受控写入等级**；**不是**车辆自动驾驶。

版本：2026
发布：YOMA · www.yongma.cloud

序言

过去两年，零售与餐饮企业几乎都被同一句话追问过：
「你们上 AI 了吗？」

于是市场上出现了大量「聊天窗口」：能答制度、能写文案、能把报表总结成一段话。它们有用，但很快碰到天花板——
问得清，办不成；建议一堆，系统还得人手点。

与此同时，行业里不乏优秀的「数智化」叙事：业务中台、数据中台、智能问数、智能补货、数字员工……
这些方向正确，但企业真正缺的，往往不是再多一个应用名称，而是一把**尺子**：

- 我们现在到底在哪一级？
- 下一级要开放什么能力？
- 怎样验收才不算被「伪 AI」忽悠？

YOMA 提出 **「AI智驾」**——即 **零售与餐饮 AI 数字化的自主等级（RAL · Retail / F&B Autonomy Levels）**：
用汽车智驾人人能懂的台阶——**脱脚** → **脱手** → **脱眼** → **脱脑** → **全域**——映射到企业经营系统里的**人机权责**，以及 **AI 能否受控写入 POS / ERP / 促销 / 库存等业务系统**。

我们还要强调一件「工程上的老实话」——这也是YOMA 长期坚持的架构选择：

> **前后端分离、业务能力全部 API 化。**
> 只有这样，AI 才能随时随地调用同一套经营接口去做分析、推理，并在受控条件下执行。
> 若经营逻辑锁死在桌面客户端、报表导出、人肉复制粘贴里，AI 永远只能停在 L1～L2。

本白皮书上篇讲清理论；下篇告诉**零售企业、餐饮企业**首年该怎么定级、怎么建、怎么验收。
我们不默认替换企业总账 ERP；**MOS 可与现网并存，也可在经营层替换 legacy ERP/POS**（边界见 §6.5.3）。不宣称 L4/L5 已作为标准产品量产；首年可签约验收的主力在 **L3 单域受控执行**（SKU-2 按域交付，非全行业开箱即用）。

—— **思源** · YOMA

执行摘要



图 F · 车机驾驶舱 × 经营驾驶舱 — 同一套智驾台阶

图 F：汽车 SAE 智驾与零售餐饮 RAL 对照 — 感知・决策・执行・接管四要素一一映射；L3 分水岭在「限定域受控执行 + 凭证核对」。

三个判断

1. **AI 数字化的标尺不是「上了几个大模型」，而是人机权责 + 能否受控写业务系统。**
2. **绝大多数企业今天在 L1～L2；可验收的商业价值分水岭在 L3**——且 **L3 ≠ RPA**（模拟点击不是受控执行）。
3. **没有全 API 化的业务底座，AI 只能旁观，不能办事。** 契约层决定天花板；控制层决定敢不敢写；人与 AI 共用同一套经营 API。

RAL 核心框架



图 K · RAL 核心主题框架（等级 × 业态 × API 一体）

图 K：L1～L5 权责台阶（L3 为可验收分水岭）、零售/餐饮首年路径，以及 YOMA「人与 AI 共用业务 API」主张的一体化总览。

等级	智驾助记	人	AI	写业务系统
L1	脱脚	全程自己做	单点工具（文案/FAQ/总结）	❌
L2	脱手	自己决策、自己点系统	多源只读 + 分析建议	❌ 只读
L3★	**脱眼**	**只做终审**	**单域决策包 + 受控执行**	✅ **审批后白名单写**
L4	脱脑	只处理例外	多域日常自主跑	✅ 自动写+风控
L5	全域	只定目标/红线	目标驱动经营	✅ 原生一体

业态一句话

业态	首年建议	为什么
零售	站稳 L2 → 选 **1 个**单域冲 L3（促销合规 / 跨系统同步 / 补货等）	多系统并行，假 L3 最常见
餐饮	站稳 L2 → 选 **1 个**单域冲 L3（估清沽清 / 时段促销 / 门店任务等）	鲜度与时段敏感，建议过期就废

核心主张



YOMA 认同「AI 必须落到业务场景」；我们额外钉死一条：**场景能否升级，取决于业务是否 API 化。** 有 API，AI 才能随时调用；无 API，再强的模型也只能做 PPT。

PART I

理论篇

AI 智驾与自主等级

上篇 · 理论：「AI智驾」等级与 AI 数字化

第一章 为什么需要「等级」而不是「又一个 Agent」

1.1 三个普遍错觉

错觉	现实
「有聊天框 = 已经 AI 数字化」	多数仍是 L1：省打字，不改经营
「能查数 = 已经有 Agent」	多数是 L2：会问、不会办
「上了 RPA / 脚本 = 已经 L3」	RPA 是录好的点击；L3 是多源证据 → 决策包 → 审批 → **受控回写 + 可审计**

1.2 市场在讲什么：我们学什么、不跟什么

近年零售软件商普遍把「数智化」讲成一条技术栈：

```
业务中台 + 数据中台
        ↓
 大模型 / 指标库 / 知识库
        ↓
效率 / 执行 / 增长 / 数据 等引擎
        ↓
 问数、补货、巡检、营销…场景
```

这类行业叙事有三点值得借鉴：

1. **AI 必须挂在业务与数据底座上**，不能飘在聊天窗口里；
2. **要覆盖效率、执行、增长、洞察**，不能只做文案；
3. **触达要进企微/门店日常入口**，不能只活在演示环境。

YOMA 不跟风的地方也有三点：

1. **先给等级尺子，再堆场景名**——否则客户无法验收「做到哪一级」；
2. **把「写系统」写成红线**——只读永远不是 L3；
3. **把「全 API 化」写成前提**——没有可调用的经营接口，场景清单再长也升不了级。

1.3 本章结论

企业要的不是「又一个 Agent 品牌」，而是：

> **可自评、可招标、可验收的 AI 数字化台阶。**

下一章给出台阶的来源：汽车智驾 → 经营系统智驾。

1.4 行业也在用「智驾五级」——常见映射与 YOMA 修正

近一年，零售与消费品行业也出现大量 **SAE J3016 → AI 应用成熟度** 的对照材料：脑脚/脱手/脱眼/脱脑/全域，一页 PPT 就能讲清。
这类材料 **降低了沟通成本**，我们认同并沿用同一套智驾助记；但在 **L2、L3 的「AI 到底是什么」** 上，常见映射与 YOMA RAL **并不等价**——若照抄，容易把企业永远锁在「只读洞察」或「RPA 代点」，却对外宣称已经 L3。

等级	智驾	行业常见映射（节选）	YOMA RAL（本书标尺）	说明
L1	脱脚	对话、搜索、内容摘要	单点工具，不写系统	✅ **一致**；可补充典型场景（见 §3.1）
L2	脱手	RPA + 流程自动化	**多源只读分析 + 方案草稿**；人自己点系统	⚠️ RPA 可作**过渡桥**，不是 L2 定义；**L2 仍禁止写**
L3	脱眼	AI BI / 智能看板 / 预测模型	**单域受控执行**：决策包 → 审批 → **白名单写 API** → 凭证	❌ **BI 再强也是 L2**；L3 看写系统与验收
L4	脱脑	自主决策执行、最小人工	多域日常自动跑 + 例外台	⚠️ 方向相近；本书**不写成已量产**
L5	全域	多智能体协同、跨域战略	目标/红线驱动 + API 原生一体	✅ **方向一致**；须先有 ≥2 域真 L3 证据

我们借什么

- **递进关系**：「每一级都建立在前一级之上」——没有真 L2 的只读 API，谈不上真 L3 的受控写。

- **L1 场景清单**：对话式 AI、企业搜索、会议纪要摘要、邮件草稿、FAQ 机器人——便于决策层与门店理解「脱脚」长什么样。
- **L2 的「辅助价值」**：AI 帮你看清、帮你起草，**价值在辅助决策**，不在代替人点保存。

我们不跟什么

- 把 **RPA 端到端脚本** 当作 L2 的终点（那是自动化，不是经营只读协同）。
- 把 **BI / 看板 / 预测** 当作 L3（没有审批后的正式写回，仍是 L2）。
- 用 **「能多快升级」** 替代 **「在哪一级验收通过」**——零售与餐饮更缺的是 **可控与可审计**，不是口号上的跳级。

> **一句话**：行业 PPT 帮我们讲清「五级台阶」；YOMA RAL 帮我们讲清 **「台阶上放什么能力、怎么验收」**——标尺仍是 **人机权责 × 写系统权限 × 全 API 化**。

第二章 YOMA AI 数字化方法论

2.1 汽车智驾的启示（人人能懂的台阶）

智驾口语	驾驶含义	映射到经营
脱脚	车帮你省一点力	工具帮你省一点活（文案/FAQ）
脱手	车帮你开一段，你仍要盯路	AI 帮你看数出建议，你自己点系统
脱眼	车在限定域自动驾驶，你负责接管	AI 在单业务域出完整方案并受控执行，人终审
脱脑	多场景自主，人处理例外	多域日常自动跑，人只接异常
全域	目标级自动驾驶	人定目标与红线，系统原生经营

助记便于沟通，验收标准须严格。 对外材料以 L1～L5 专业名为准，智驾口语只作沟通润滑。

2.2 经营系统里的真正标尺

YOMA RAL 的唯二标尺：

① 人机权责：谁决策、谁执行、谁对结果负责
② 写系统权限：AI 对 POS/ERP/促销/库存等能否受控写入

不是标尺的东西（避免被带偏）：

- 参数量、对话轮数、是否「接入了 DeepSeek」
- 海报好不好看、机器人人格俏皮不俏皮
- 一次性演示是否惊艳（无审计、无回写、无熔断都不算）

2.3 为何「全 API 化」是智驾的物理前提（YOMA 架构主张）

把 AI 比作司机，业务系统比作车：

车的状态	AI 能做什么	对应等级上限
只有仪表照片 / Excel 导出	事后总结	L1
有只读仪表盘接口	分析、建议	L2
有受控的方向盘/油门接口（写 API）	单域受控执行	**L3**
多套执行机构可编排 + 统一风控	多域自主	L4
车辆为 AI 原生设计	目标驱动	L5

YOMA 长期选择 **前后端分离、业务能力 API 化**，含义是：

1. **人用的界面**（MOS、POS、小程序、Pad）与 **AI 用的能力** 走同一条业务契约；
2. AI 不靠「模拟点击旧客户端」混日子（那是 RPA，不是 L3）；
3. 分析与推理可以 **随时随地** 触发：值班群、门店、总部、定时任务——只要持有授权，就能调同一套 API；
4. 升级等级时，优先加的是 **权限、审批、审计、白名单**，而不是再做一个无法对接的孤岛应用。

线框：人与 AI 共用 API

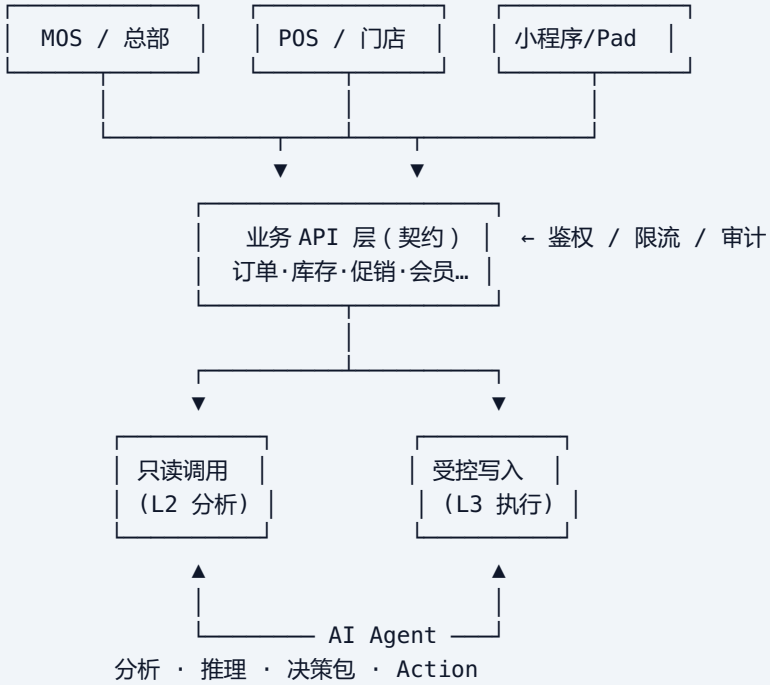




图 B · 人与 AI 共用业务 API

图 B：MOS / POS / 小程序与 AI Agent 汇入同一业务 API 契约层 — 只读通向 L2，受控写通向 L3。

结论：大模型解决「会不会想」；API 化解决「碰不碰得到真经营」。
YOMA 的优势，首先是后者——这是「AI智驾」能从 L2 开到 L3 的车道线。

2.4 与常见数智化路径的对照

维度	常见数智化平台路径	YOMA AI 数字化路径
入口	四大引擎 / 场景清单	L1～L5 等级尺子
技术重心	中台 + 大模型 + 应用	**全 API 契约 + 等级验收 + 受控写入**
客户听完能做什么	「我们有哪些智能应用」	「我在第几级、下一步开放什么写权限」
验收	功能清单打勾	**决策包 → 审批 → 回写单号 → 审计**
对 ERP / 经营系统	并存或替换因项目而异	**账/契约/Agent 分工写进合同**；并存时 ERP 保账务权威；替换时 MOS 可为经营主责

两者可以叠加：中台与场景解决「做什么」；RAL 解决「做到哪一级算数」。

第三章 RAL L1～L5 全景展开

说明：每一级用同一套栏目——定义 → 人机分工 → 写系统 → API 要求 → 能力域 → 达标要求 → 常见误判 → 验收 → 升级条件。
递进：**每一级都建立在前一级之上**——L1 的文案能力不能替代 L2 的只读契约；L2 的洞察不能替代 L3 的受控写与凭证（参见 §1.4 常见映射修正）。
商业重心在 L3；L4/L5 写清愿景与前置，不写成「明年标配」。



图 A · RAL L1～L5 智驾台阶与 API 权限

图 A：五级递进 — 没有只读 API 到不了真 L2；没有受控写 API 到不了真 L3。

3.0 L1～L5 对照总表

	L1 工具辅助	L2 数据协同	L3 单域受控执行★	L4 多域自主	L5 经营原生
智驾	脱脚	脱手	脱眼	脱脑	全域
人	全程主导	决策+自己操作系统	**终审**	只处理例外	只定目标/红线
AI	单点工具	多源只读分析建议	单域完整决策包+执行	多域日常自动跑	目标驱动全链路
写系统	✗	✗ 只读	✅ 审批后白名单	✅ 自动+风控	✅ 原生
API 要求	可无	**只读 API 必备**	**读写 API + 审批审计**	多域编排+统一网关	API/流程为 AI 原生
常见误判	「我们有 ChatGPT」	「我们有 BI 机器人」	「脚本代点等于 L3」	「两个机器人并列= L4」	「战略 PPT」
升级到下一级	接上只读经营数据	选定单域、开放写接口与审批	≥2 域稳定 L3 + 调度台	目标体系与原生重构	—

七大能力域

编号	能力域	一句话
D1	数据采集与同步	POS/ERP/会员/门店数据能否进来、是否及时
D2	知识与 RAG	制度/SOP/活动规则能否检索并引用来源
D3	对话与意图路由	自然语言能否识别「要办的事」
D4	分析建议	多源数据能否给出可执行建议（仍可不执行）
D5	工具 / Action（写系统）	能否受控建单、改价、发券、重推、调拨
D6	审批与风控护栏	谁批、频率、审计、回滚、禁写清单
D7	多 Agent 协同	多域调度、例外升级、持续调优（L4 起）

递进死规定：L1 无 D5；L2 禁 D5；L3 必须在白名单+审批下开 D5+D6；L4 开 D7。

3.1 L1｜工具辅助（脱脚）

定义

AI 只在一个点替人省力（文案、FAQ、报表摘要），**不进入经营闭环。**

人 / AI 分工

角色	做什么
人	开活动、改库存、对账、审批——全部自己点系统
AI	写海报文案、答固定 FAQ、把 Excel 总结成一段话
禁区	跨系统查真数、出经营决策包、改任何单据

写系统？ ✕ 不能写，也不应假装能写。

API 要求

项	要求
业务只读 API	不需要（Excel / 截图即可）
业务写 API	**禁止对接**
模型 API	公有大模型即可

能力域做到什么程度

域	程度
D1～D2	可选；固定 FAQ 文档即可
D3	弱：固定问答或复制 Prompt
D4	弱：事后总结
D5～D7	**关闭**

达标要求（零售 / 餐饮通用）

- 美工用 AI 出会员日海报文案；
- 客服机器人答营业时间 / 退货政策；
- 店长把上周导出表扔给 AI，得到一段文字复盘。

典型场景

类型	例子
对话式 AI	日常问答、制度解释（ChatGPT / Copilot 类）
企业搜索	内部知识检索、SOP 查询
内容摘要	会议纪要、周报、报表段落总结
起草辅助	邮件/通知/活动文案初稿
FAQ 机器人	营业时间、退货政策、固定话术

常见误判

假象	拆穿
「我们公司已经全面 AI 化」	只有聊天窗 = 仍是 L1
「接入了某某大模型」	模型≠等级；碰不到经营 API 就上不去

验收要点

- ☐
  选定 ⩾1 个单点场景，有真实使用记录
- ☐
  书面确认：**不写入任何业务系统**
- ☐

团队完成 ≥1 场使用规范培训（防乱贴密钥、防泄密）

边界

不承诺提升连带、降损耗、自动补货；不把「打通只读经营分析」算进本级（那是 L2）。

升级到 L2 必须补齐

① 至少 2 路经营数据可 API 只读（或稳定日级同步）
② 字段口径文档 + 只读授权
③ 企业知识库起步 + 可审计的问答入口（企微/飞书/MOS）

3.2 L2｜数据协同（脱手 · 辅助价值）

定义

AI 能读多源经营数据与企业知识，用对话给出分析与方案草稿；**人听完建议后，自己登录系统执行。**

行业材料常把 L2 讲成「RPA 跑流程」——在 YOMA RAL 里，RPA **最多是过渡桥**；**L2 的标尺仍是只读 API + 可审计建议**，价值在 **辅助决策**，不在代点保存。

人 / AI 分工

角色	做什么
人	全部业务操作与最终决策；自己点 MOS/POS/CRM
AI	只读取数 + RAG + 分析建议 / 方案草稿
禁区	建采购单、发券、改价、重推过账、锁售——**一律不能写**

> 定级要点：客户说「我要 Agent」——先问「它能不能改你的系统？」不能，就是 L2。

写系统？ ✕ 只读；写接口必须技术关闭 + 制度关闭。

API 要求（真 L2 红线）

项	要求
业务只读 API	**必备**；建议 ⩾2 个源（如订单/库存 + 会员或主数据）
延迟	分钟～小时级可接受；全靠人工导 Excel = **伪 L2**
写 API	**明确未授权 / 网关层拒绝**
审计	会话可追溯；敏感字段脱敏

前后端分离、全 API 化的价值在本级开始兑现：**同一套只读契约，人用报表/页面看，AI 用工具调。**

能力域做到什么程度

域	程度
D1	只读 API 或日级同步；口径有文档
D2	有效知识条目起步（制度/SOP/活动规则），回答能带引用
D3	企微/飞书/Web 入口；覆盖约定 TOP 场景
D4	能答约定问题清单（损耗、畅滞销、估清原因草稿、活动要点等）
D5	**关闭**
D6	会话审计 + 脱敏
D7	无

达标要求

业态	例子
零售	「上月 A 店临期损耗？」→ 关联销售+库存规则出数字与原因草稿；人去系统调拨/改价
餐饮	「昨晚晚餐估清了哪些菜、为什么？」→ 只读汇总；店长自己改明日预估
共用	「下周会员日怎么促？」→ 方案要点；人去 CRM/促销模块配置

常见误判

假象	拆穿
「我们有 BI 机器人」	若数据是人贴上去的截图，不是 API，仍是 L1 加强版
「Agent 已经能下单」	若无审批链路与正式写 API，是演示级效果，不是 L2，更不是 L3

验收

- ☐
  约定高频问题抽测，有效回答可引用数据或知识库
- ☐
  100% 会话可审计；抽测无越权写入
- ☐
  未命中问题进运营台，有周更知识库机制
- [ ]输出《L3 候选单域清单》，只选 1 个优先域

边界

不出现「一键批准后自动建单」；不宣传替代营运/信息部。

升级到 L3 必须补齐

① 选定唯一业务域（促销 / 同步 / 补货 / 沽清任务…）
② 该域读写 API（或写中间层再过账）+ 幂等与审计
③ 审批流（企微/飞书/MOS）+ 白名单 + 禁写清单
④ 决策包格式（明细、依据、风险）而非散文建议

3.3 L3｜单域受控执行（脱眼）★ 商业价值主战场

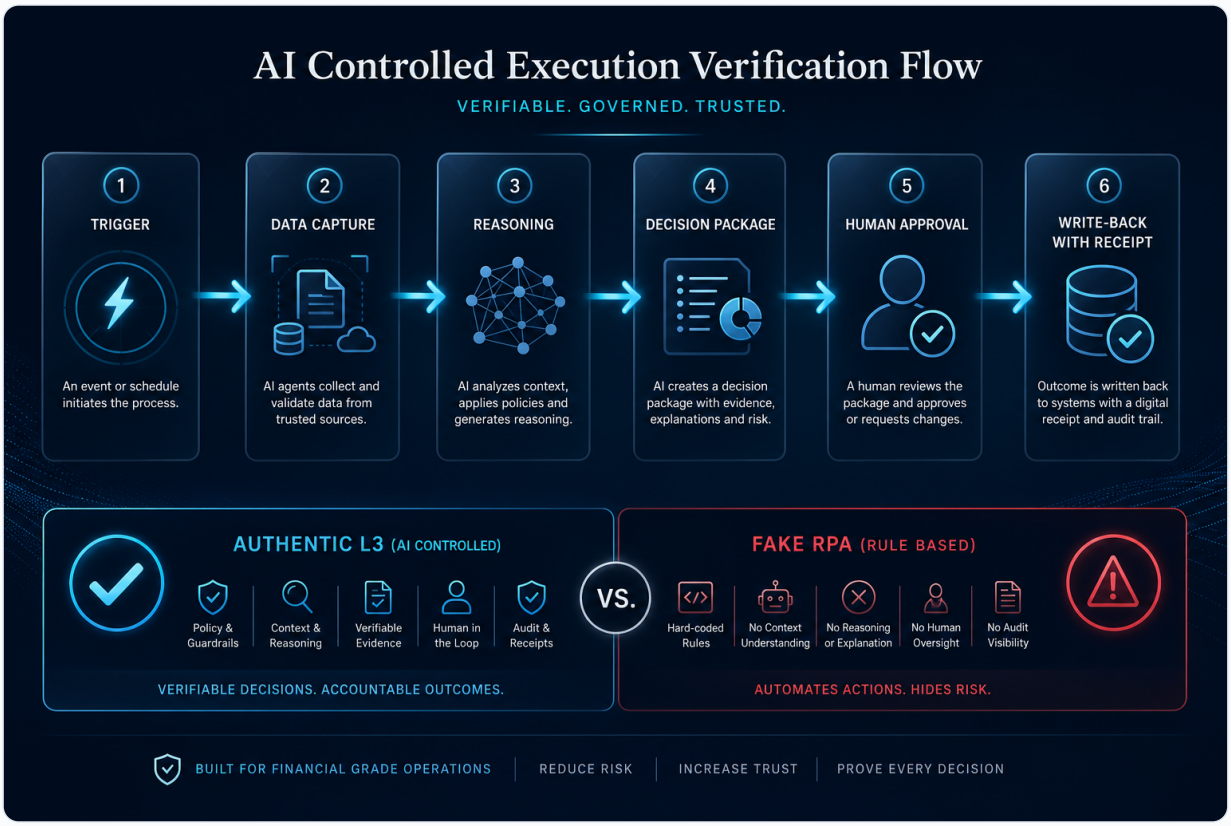


图 C · L3 验收六步

图 C：触发→取证→推理→决策包→审批→写回凭证 — 缺任一步即假 L3；RPA 模拟点击不在此列。

定义

在 **一个业务域** 内，AI 完成「取数 → 推理 → 完整决策包 → 调白名单工具写入」；人只做 **终审**（同意 / 驳回 / 改参数），不盯中间每一步。

人不用全程盯流程（脱眼），但人对结果负责。

人 / AI 分工

角色	做什么
人	终审；设定阈值/额度/禁写；看审计与失败单
AI	单域闭环：证据链 + 决策包 + 受控 Action
护栏	大额、超安全库存、改价超阈值、低置信 → 强制人工

写系统？ ✅ **审批后**，且仅白名单接口。

API 要求（进 L3 红线——写不进合同就不要卖 L3）

项	要求	不达标后果
域内只读 API	准实时（建议关键库存/订单延迟可约定，如 ≤15 分钟）	决策包永远过期
域内写 API	≥2 个可验收的写 Action（如创建草稿单、回写状态、拦截配置）	**假 L3、真 L2**
鉴权	Agent 独立身份或明确委托账号；禁止共享超管	审计废掉
幂等 / 重试	写失败可解释、可重试、可标记	重复建单事故
回写可见	**业务系统可查到单号/状态**	不算闭环
ERP 关系	并存：账套真相在 ERP，经 MOS 白名单写过账；替换：MOS 为经营主责，总账/法定凭证仍按企业对接	边界不写清、Agent 直连改账

能力域做到什么程度

域	程度
D1	域内关键数据准实时
D2	域内规则进知识库或规则引擎（互斥、毛利、安全库存、golden 用例）
D3	入口：企微卡片 / MOS 待办 / POS 提示（按域选择）
D4	输出必须是 **可执行决策包**（明细行 + 依据 + 风险），不是散文
D5	**至少 2 个写 Action 上线**
D6	审批 + 审计 + 禁写清单 + 失败说明
D7	不需要（单 Agent 即可）

单域样板（一次 Pilot 只选 1 个）

域	闭环（必须能演示到「单号」）	更贴
促销合规 PROMO	规则/毛利检测 → 拦截或放行 → 活动可上架或支付前阻断	零售主推
跨系统同步 SYNC	发现 POS 与 ERP 不一致 → Action（重推/锁售/建 IT 单）→ 人确认执行	零售主推
补货/调拨 REGEN	建议单（SKU×店×量+理由）→ 批准 → 调拨/采购草稿单号可见	零售/餐饮
会员营销 MKT	圈人+券方案 → 人审 → CRM/券平台下发成功	有写权限再选
门店作业 STORE	今日必办 → 确认回写；餐饮可含沽清/估清状态	餐饮常选
沽清/估清（餐饮）	状态变更建议 → 店长确认 → 前厅/外卖通道同步	餐饮主推

达标要求（含验收示例）

```
触发（定时/异常/人工）
  → AI 调只读 API 取证
  → 生成决策包（业务可读）
  → 人在企微/MOS 点「批准」
  → AI 调写 API
  → 业务系统出现单号/状态
  → Console 显示成功行 / 失败行与原因
```

常见误判

假象	拆穿
RPA 夜里代点客户端	无决策包、无正式 API、无经营 KPI → **不是 L3**（见第四章专节）
演示环境能写、生产不能写	合同与验收必须锁生产写接口范围
一次上五个域	违反「单域」；失败不可归因 → 退回 L2 包装

验收要点

- ☐ 选定 **1 个域**，试点范围书面锁定
- ☐ 端到端：**触发 → 决策包 → 审批 → 写系统 → 可查单号**
- ☐ 白名单外写入尝试 = 0（抽测）
- ☐ 约定业务指标（如闭环成功率、人工耗时下降、拦截质量）
- ☐ 运维移交：谁改阈值、谁加规则、谁看审计

边界

不做「全公司多域一起自主」（那是 L4）；**无读写 API 不得承诺 L3。**

升级到 L4 必须补齐

① ≥2 个域均达到本章 L3 验收
② 统一 Action 网关 + 跨域风控 + 失败补偿
③ 调度台（谁先谁后、互触发、例外升级）
④ 企业级 AI 运营 Owner（不是纯外包）

3.4 L4｜多域自主（脱脑）

定义

≥2 个已验收的 L3 域，在调度与统一风控下 **日常自动跑完整闭环**；人只处理例外（超阈值、低置信、跨域冲突、客诉）。

人 / AI 分工

角色	做什么
人	例外工作台；调全局策略；月度复盘
AI	多域自动执行；域间可互触发（如临期→折扣方案→额度内自动上架）
协同	依赖关系、补偿、对账——不能只「调用成功」

写系统? ✅ 常规自动写；风控拦截；例外升人。

API 要求

项	要求
各域	维持 L3 读写与审计标准
统一网关	白名单、幂等、全链路审计一条链
编排	A 完成触发 B；失败可补偿/重试
主数据	跨门店商品/门店/会员口径一致

能力域：D1～D6 每域达 L3；**D7 必开**。

达标要求
临期将至 → 库存域触发营销域生成折扣 → 额度内自动生效 → 次日对账可解释；仅当折扣超阈值才推人。

常见误判
两个聊天机器人「并排展示」≠ L4；没有调度与统一风控，只是两个 L2/L3 项目。

验收方向

- ☐
  ⩾2 域连续运行达约定天数，无人干预闭环占比达约定值
- ☐
  误自动执行事故 = 0（或 ⩽ 约定）
- ☐
  跨域差异可解释、可对账

边界说明
前置：**至少 1 个 L3 域验收通过**，否则不谈 L4。
YOMA 态度：L4 是平台能力建设目标，**不作为无 L3 证据时的标准承诺**。

升级到 L5
目标体系、组织与流程为 AI 原生重构、垂类模型与端云协同——属战略共创，见下节。

3.5 L5｜经营原生（全域）

定义
人只下达 **经营目标 + 风控红线**；采购节奏、定价、人力、渠道等由 AI 网络在红线内调度；系统与流程按 AI 原生设计，而非「在旧烟囱上挂聊天窗」。

人 / AI 分工：人定目标与红线；AI 目标驱动写入与协同。

写系统？ ✅ 原生一体。

API / 架构要求

传统烟囱大量被可编排服务取代；契约、权限、仿真环境为 AI 而设计——**不是把 L3 网关简单放大。**

达标要求

以年度利润、缺货、人效等经营指标共创；过程指标（会话数、插件数）不再是主角。

常见误判

「AI 战略规划 PPT」「全面智能化三年蓝图」而无 L3 闭环证据。

验收与产品态度

项	口径
标准产品	**不做 L5 固定规格与标准价承诺**
路线图	可画愿景页，须标注依赖 L3/L4 前置
合同	仅战略共创 / 联合创新，单独治理

3.7 本章小结

1. **等级看的是权责和写权限，不是模型商标。**
2. **没有只读 API，到不了真 L2；没有受控写 API，到不了真 L3。**
3. **先拿下一个单域 L3 闭环，再谈多域与「智驾高级别」。**

第四章 从 L2 到 L3：商业价值分水岭

本章回答三件事：
① 为什么多数企业卡在「会问不会办」；
② 怎样才算跨过 L3（验收六步）；
③ **L3 为什么不是 RPA**——以及合同里该写什么句子。

4.1 为什么 L2 不够，但又必须先有 L2

L2 解决的是 **看见**：多源只读、知识引用、建议草稿。
L3 解决的是 **落单**：审批后写入、单号可见、过程可审计。

	L2「会问」	L3「会办」
输出物	分析段落 / 方案要点	**决策包**（明细+依据+风险）
人的动作	听完自己去系统点	**终审**（批/驳/改参）
系统变化	无（只读）	**有单号 / 有状态变更**
API	只读	只读 + **白名单写**
价值感	「终于问得清」	「少开会、少人肉改系统、错了拦得住」

两句实话：

1. **跳过真 L2 直接喊 L3**，通常会在「数据口径打架、知识库空洞」上翻车——决策包没有可信证据。
2. **停在 L2 当终点**，决策层容易认为「又多了一个会聊天的助手」，数字化投入难以形成可验收闭环。

正确台阶： 伪聊天(L1) → 真只读(L2) → 单域受控写(L3) → …
错误台阶： 演示级效果写一下 → 生产关写接口 → 对外仍称「智能执行」

4.2 分水岭判据：一张表定级

项目启动时可用下表自检：**任一「否」都不能报 L3**：

#	判据	是 / 否
1	是否选定 **唯一业务域**（本 Pilot 只做一个）？	
2	域内关键数据是否走 **正式只读 API**（非人工贴表）？	
3	是否存在 **业务可读的决策包**（不是散文建议）？	
4	是否有 **审批节点**（人可批/驳/改参）？	
5	批准后是否走 **白名单写 API**（非模拟点击）？	
6	写入后是否在业务系统 **可查单号或等价凭证**?	
7	是否全链路 **审计**（谁批、写了什么、成败原因）？	
8	白名单外写入是否被 **拒绝且可抽测证明**?	

> **要点**：有没有聊天框不构成等级；**能不能受控改单据并留下单号**才构成等级。

4.3 L3 验收闭环（六步，必须能现场演示）

```
① 触发    定时 / 异常规则 / 人工唤起
    ↓
② 取证    多源只读 API（订单·库存·规则·主数据…）
    ↓
③ 推理    规则引擎 + 模型（可复现、可 golden）
    ↓
④ 决策包  明细行 + 依据 + 风险点 + 建议 Action
    ↓
⑤ 人终审  批准 / 驳回 / 改参数（企微·飞书·MOS）
    ↓
⑥ 写回    白名单写 API → 业务单号/状态可见 → 审计落库
```

「验收」时刻（零售）
促销冲突拦截：不通过则活动无法上架或支付前阻断；Console 能看到拦截次数与放行记录。
同步异常：批准「重推」后，对账差异笔数下降且有执行日志。

「验收」时刻（餐饮）
沽清建议批准后，堂食/外卖通道状态一致；或时段促销批准后配置生效且可回滚。

失败也要可解释：写失败不能静默——须返回原因（库存锁、权限、幂等冲突），支持重试或转人工。

4.4 决策包长什么样（最小字段）

决策包是 L3 与「聊天建议」的分界物。最小应包含：

字段	说明
域 / 场景	如 PROMO、SYNC、沽清
触发原因	规则命中或指标阈值
证据摘要	引用了哪些 API 数据 / 规则条文
明细行	SKU×门店×动作×数量或参数
风险点	毛利、超安全库存、渠道冲突等
建议 Action	将调用的写接口清单
审批人 / 时限	谁批、超时策略
模拟影响（可选）	预估影响面，供人判断

没有明细行、只有「建议您关注库存」——**仍是 L2**。

4.5 专节：L3 ≠ RPA

市场上最容易被混为一谈的，是「夜里有个机器人在点鼠标」与「受控智能执行」。必须拆开卖、拆开验。

4.5.1 一句话

	RPA	L3 受控执行
买的是什么	**机器人人手**（照录路径点界面）	**可审计的执行决策权**（取证→推理→审批→API 写回）
界面一变	脚本易崩	契约稳定则不受页面改版影响
会不会读经营规则	通常不会	**必须会**（规则/模型 + 可测用例）
坏了怎么办	重录脚本	审计定位 + 回滚/补偿 + 禁写

4.5.2 四维对照

维度	RPA	L3
依据	录屏 / 固定点击路径	多源证据 + 业务规则 / 模型推理
输出	点击序列	**可审决策包**
写入	模拟 UI（客户端/网页）	**正式业务 API / 服务**
经营指标	常无（只问「脚本跑通没」）	闭环成功率、耗时、拦截质量、单号可达

4.5.3 何时还需要 RPA?

- **遗留系统短期无 API**：可用 RPA 作 **临时桥**，但对外等级仍应标为「过渡」，**不得称为已达 L3**；
- **L3 的目标态**：把桥拆掉，写路径迁到 API；
- 合同建议句：

「涉及模拟点击的路径不计入 L3 验收范围。」

4.5.4 可写入招标 / SOW 的验收句

1. 乙方须交付选定业务域的端到端闭环演示：**触发 → 决策包 → 审批 → 写系统 → 业务侧可查单号（或等价凭证）**。
2. 写入须经 **白名单接口**；白名单外请求须被拒绝，并提供抽测记录。
3. 全过程须可审计：操作者、审批者、请求参数摘要、成功/失败原因可追溯。
4. **模拟界面点击（RPA）不视为本项目 L3 验收通过条件。**
5. 须提供可重复的规则/用例集（golden）：约定案例集通过后方可初验。

4.6 从 L2 走向 L3 的 90 天节奏

```
D0～D15   定级 + 锁定唯一域 + 盘点只读/写 API 缺口
D16～D45  补齐只读口径与知识/规则；决策包原型给人审
D46～D75  打通写 API + 审批 + Console；灰度门店/部门
D76～D90  闭环验收演示 + 指标回顾 + 运维移交（谁改阈值/规则）
```

常见误区：同时开三个域「都做一点」——结果是三个假 L3。
组织禁忌：没有业务 Owner，只丢给信息部「上个机器人」。

4.7 本章小结

1. L2→L3 的本质是从 **旁观** 到 **受控落单**。
2. 六步闭环少一步，就还在岸上。
3. **RPA 可以是桥，不能是 L3 的定义。**
4. 无正式读写 API，YOMA 方法论下 **禁止承诺 L3**。

第五章 底座：等级升高，底座必须加码

> 等级不是「多买几个模型账号」。
> **每一级升高，对应底座某一层必须变厚**——尤其是契约层（API）与控制层（审批/审计）。

5.1 五层底座总览

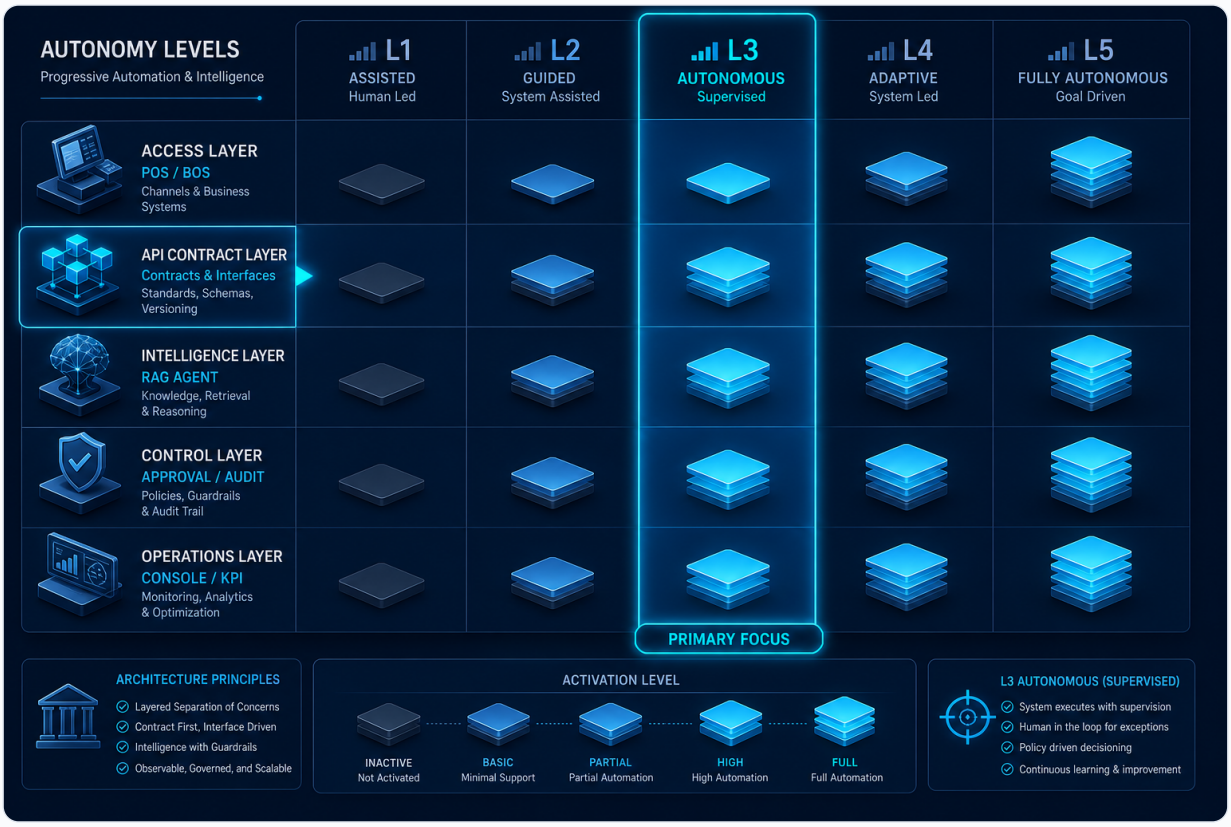


图 G · RAL L1～L5 × 五层底座系统架构

图 G：纵轴为接入→契约→智能→控制→运营五层底座；横轴为 L1～L5 智驾等级。亮块表示该级必须加厚的能力 — 升一级重在加厚契约层（API）与控制层（审批/审计），而非堆模型。

```
┌──────────────────────────────────────────────┐
│ ⑤ 运营层 Console · KPI · 知识迭代 · Owner    │
├──────────────────────────────────────────────┤
│ ④ 控制层 审批 · 白名单 · 限流 · 熔断 · 回滚  │
├──────────────────────────────────────────────┤
│ ③ 智能层 RAG · 规则 · 模型 · Agent           │
├──────────────────────────────────────────────┤
│ ② 契约层 业务 API（鉴权/版本/幂等/审计）     │ ← YOMA 主张主战场
├──────────────────────────────────────────────┤
│ ① 接入层 POS · MOS · 小程序 · 企微/飞书      │
└──────────────────────────────────────────────┘
        ▲ 人点页面                ▲ AI 调工具
        └──────── 同一套契约 ──────┘
```

层	解决什么	薄弱时的症状
① 接入	人从哪入口用	AI 只能活在演示链接里
② 契约	经营能力是否可调用	永远 L1；或 RPA 硬撑
③ 智能	会不会想、规则在哪	有接口也不会出靠谱决策包
④ 控制	敢不敢写	一写就事故，被信息部永久关掉
⑤ 运营	能否持续	上线即烂尾，知识与规则无人养

5.2 完备度矩阵

图例：– 不要求　○ 建议有　● 必须有　◎ 必须强（准实时/可编排）

5.2.1 ① 接入层

检查项	L1	L2	L3	L4
Web/MOS 或等价后台可访问	○	●	●	●
门店 POS **云端可连**（非纯单机孤岛）	—	●	●	●
企微/飞书/待办触达	—	○	●	●
小程序/Pad 等业务入口可接同一 API	—	○	○	●

5.2.2 ② 契约层（API）——等级天花板由此决定

检查项	L1	L2	L3	L4
经营数据 **只读 API**（⩾1～2 源）	—	●	●	◎
字段口径文档 / 主数据编码可对上	—	●	●	◎
域内 **写 API**（白名单）	—	—	●	◎
鉴权：人与 Agent 身份可区分	—	○	●	●
幂等键 / 防重复建单	—	—	●	●
全链路调用审计（谁、何时、何参、结果）	—	○	●	◎
API 版本管理 / 兼容策略	—	○	●	●
统一 Action 网关（跨域）	—	—	—	●
前后端分离、业务不锁死在客户端二进制	○	●	●	●

> **判定要点**：契约层只有「导出 Excel」→ 最高伪 L2；有只读无写 → 真 L2；有写无审批审计 → **危险的假 L3**。

5.2.3 ③ 智能层

检查项	L1	L2	L3	L4
大模型调用（公有或私有）	●	●	●	●
企业知识库 / RAG（可引用）	○	●	●	●
域内规则引擎或可配置参数	—	○	●	●
可回归的 golden / 用例集	—	○	●	●
Agent 工具调用（Function/ Action）	—	只读工具	读写工具	多 Agent 工具
Prompt / 规则 **版本可回滚**	—	○	●	●
多 Agent 调度与互触发	—	—	—	●

5.2.4 ④ 控制层

检查项	L1	L2	L3	L4
写权限默认关闭	●	●	（白名单开）	（策略开）
人工审批流	—	—	●	●（例外+策略）
禁写清单 / 金额频率阈值	—	—	●	◎
失败重试与错误码对人可读	—	○	●	●
回滚或补偿说明（业务可逆或可对冲）	—	—	○	●
大促/高峰 **熔断开关**	—	—	○	●
跨域冲突检测与升级人工	—	—	—	●

5.2.5 ⑤ 运营层

检查项	L1	L2	L3	L4
使用量/会话日志	○	●	●	●
未命中问题回流 → 知识周更	—	●	●	●
决策包列表 + 审批成功率看板	—	—	●	◎
业务 KPI（耗时、拦截、缺货、沽清及时率…）	—	○	●	◎
具名 **业务 Owner + IT Owner**	○	●	●	●
月度例外复盘机制	—	—	○	●

5.3 终端与集成现实检查

现状	最高可达	动作建议
单机 POS，只能 U 盘/本地 Excel	L1～伪 L2	先云 POS / 可 API，再谈 Agent
有中台报表，无开放只读 API	伪 L2	先暴露只读契约
只读 API 齐，写要改 ERP 核心且无审批	真 L2	L3 改写 MOS/中间层或加审批网关
演示环境可写，生产只读	假 L3	合同锁生产范围，否则降级承诺
多套编码（商品/门店）不一致	L3 高风险	先主数据治理再开写

5.4 底座投入与等级跃迁

```
只加模型，不加 API        →  停在 L1
只加只读 API + RAG        →  可到真 L2
加写 API + 审批审计       →  可到真 L3
加调度网关 + 多域风控     →  才谈 L4
流程与系统为 AI 原生      →  才谈 L5（战略共创）
```

YOMA 强调 **前后端分离、全 API 化**，正是为了让「加码」发生在契约层——
人与 AI 共用升级后的能力，而不是每升一级就堆一个无法对接的孤岛应用。

5.5 本章小结

1. 底座五层，**契约层决定天花板，控制层决定敢不敢写，运营层决定能否活过试点。**
2. 用本章勾选表做差距分析，比空谈「上 AI」有用。
3. 下篇将把同一套尺子落到 **零售 / 餐饮** 的选域与首年路径。

PART II

实践篇

零售与餐饮建设路径

下篇 · 实践：零售与餐饮该如何做

> 上篇给出 RAL 等级与五层底座；下篇说明 **如何交付、如何衔接现网、如何排期落地**，以及零售 / 餐饮如何选域。MOS 产品与 SKU 路径见 **§6.6**。

第六章 通用建设路径（两业态共用）

6.1 五步主路径

```
① 定级诊断    今天在第几级？底座缺哪一层？
        ↓
② 站稳真 L2    只读 API + 知识库 + 可审计问答
        ↓
③ 单域冲 L3    唯一域 + 写白名单 + 审批 + 决策包
        ↓
④ 试点验收    验收六步 + 指标 + 运维移交
        ↓
⑤ 复制评估    扩店 / 第二域 —— 未达标不谈 L4
```

铁律三条：

1. **一次只做一个 L3 域**（并行多域 = 三个假闭环）。
2. **先契约、后智能**：只读/写 API 缺口未补，不堆模型。
3. **先验收、再宣传**：没有单号与审计，不对外称「智能执行」。

6.2 谁参与、谁签字（组织）

角色	职责	不签字的后果
业务 Owner（营运/商品/门店运营）	选域、定规则、当审批人、认 KPI	项目变「信息部玩具」
IT / 数字化 Owner	API、鉴权、审计、环境、与 ERP 边界	演示能写、生产不能写
门店代表（试点店长）	真用入口、反馈决策包是否能懂	总部自嗨、一线不用
供应商 / 实施方	交付物件、验收证据、移交清单	上线即失联

> 最少配置：**1 名业务 Owner + 1 名 IT Owner**。缺一则 L3 Pilot 建议暂缓。

6.3 分步交付：交什么、客户如何感知价值

① 定级诊断（通常 0.5～2 天工作坊 + 纪要）

交付物件	价值感知瞬间
《定级纪要》：当前 L 级、伪/真判定、底座缺口	会议室当场承认「我们多半是 L2」
《唯一域建议》+ 不选其它域的理由	决策层明确首年单域优先，暂不并行多域试点
《接口与写入授权里程碑》	信息部看清要开哪些只读/写 API
《试点 KPI 卡》基线草稿	用自己的损失语言（错促、对账、损耗）说话

② 站稳真 L2

交付物件	价值感知瞬间
企微/飞书/MOS 问答入口	一线问出**带出处的数**
⩾2 路只读 API + 口径说明	「数据从哪来」可打开引用
知识库起步 + 未命中回流	信息部看到可养护的问题列表
运营台会话审计	界面不提供写操作（真 L2）

退出标准（进入 L3 的门票）： 抽测约定题可引用；无越权写；口径文档齐；唯一域已锁定。

③ 单域冲 L3

交付物件	价值感知瞬间
规则/用例包（含 golden）	业务说「这几条错案必须拦住」
决策包卡片（明细+依据+风险）	审批人不再看散文
审批流 + 白名单写	批一张卡 → 系统落状态
Console：成功率/失败原因	失败可解释，不是黑盒

④ 试点验收

必须同时具备第四章「验收六步」证据 + 约定 KPI 回顾 + 《运维移交》（谁改阈值、谁加规则、谁看审计）。

⑤ 复制评估（扩店或第二域）

问题	通过才继续
第一域是否连续稳定运行?	是
主数据/权限是否可复制到新店?	是
是否仍只有一个业务 Owner 在扛?	需补 Owner 再建第二域
是否要上调度台（L4）？	**仅当 ⩾2 域均达 L3 验收**

6.4 首年节奏图

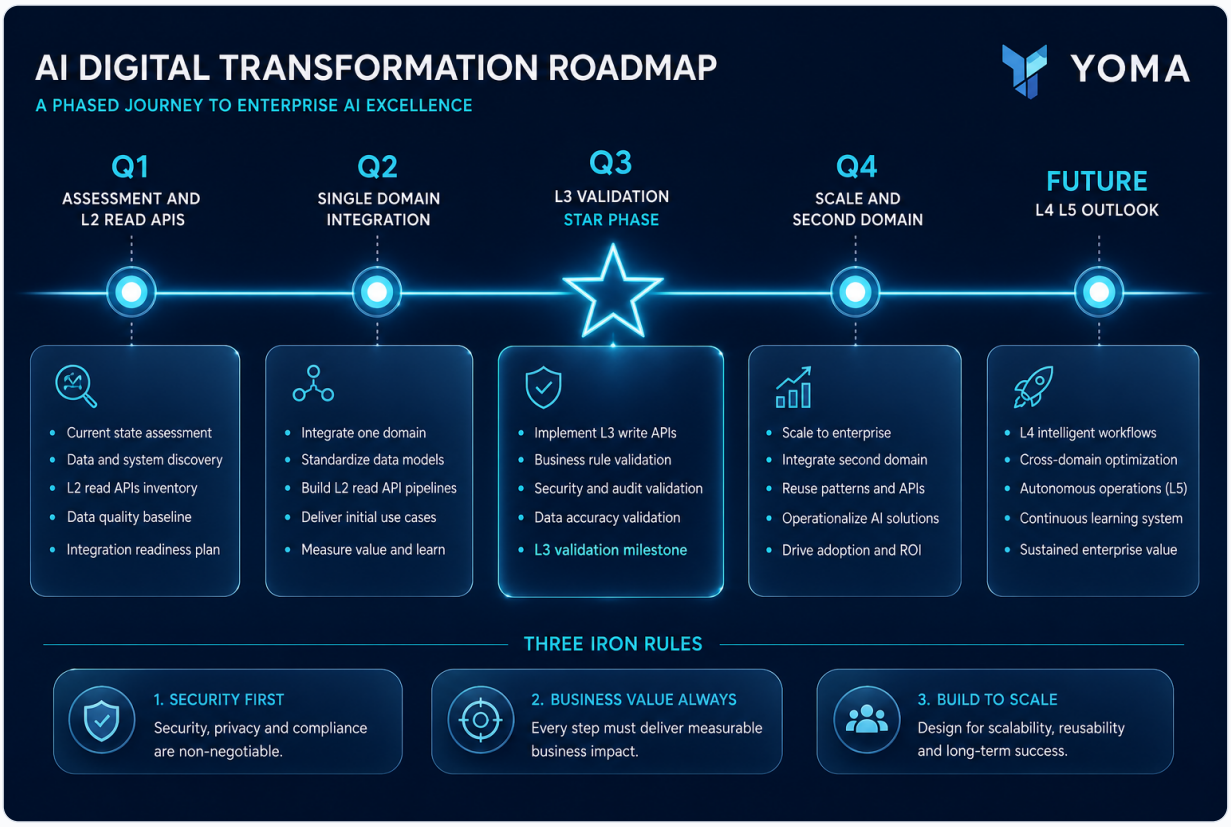


图 E · 系统建设路线图（首年通用）

图 E：Q1 定级+真 L2 → Q2 单域联调 → Q3 L3 验收 → Q4 巩固复制；与 6.1 五步路径一一对应，Q3 为首年须签字里程碑。

```
Q1  定级 + 补只读 API + 真 L2 上线抽测
Q2  唯一域规则/决策包/写接口联调；灰度试点
Q3  L3 验收 + 运维移交；指标看趋势
Q4  巩固：扩店或评估第二域；不达标则加固 L3，不喊 L4
```

可压缩，但 **不可颠倒**：未真 L2 不签 L3；未 L3 单号不谈多域自主。

6.5 交付系统架构（图 H）

企业问「AI 要接哪些系统？」—— 答案不是「再接一个大模型」，而是 **在明确账/契约/Agent 分工的前提下，叠加可验收的智驾交付栈**：

并存时：现网 ERP/POS 保持账务与门店执行权威，MOS 做契约与受控写，Agent 做推理与决策包。

替换时：MOS 四端可承担经营主责（管理、门店、会员、契约 API），总账/法定凭证仍按企业要求对接或导出。



图 H · YOMA 交付系统架构（现网 × MOS × Agent）

图 H：左列为企业现网（ERP / POS / CRM / WMS / O2O）；中为 **MOS 契约层 + AI Agent 运行时 + 控制层**；右为企微/飞书/MOS Web/POS 等触点。L2 走只读 API；L3 写回经 MOS 白名单与审批。**并存**时不绕过 ERP 账务权威；**替换**时 MOS 为经营主责，边界见 §6.5.3。

6.5.1 分层职责（谁管什么）

层级	典型系统	在 RAL 中的角色	改造原则
经营权威层	ERP / 财务 / 主数据，或 **MOS 四端（替换模式）**	库存、价格、促销、订单、凭证的 **最终真相**	**并存**：保留现网 ERP 权威；**替换**：MOS 承担经营主责，总账/法定凭证单独约定
门店执行层	POS / 小程序 / 外卖中台，或 **MOS POS / 小程序**	收银、沽清、履约、门店侧状态	**并存**：现网 POS + 同步/API 网关；**替换**：MOS 智能 POS / 会员小程序
协同触达层	企微 / 飞书 / MOS Web	问答入口、决策包卡片、审批通知	L2 起上线；与 HR 组织权限对齐
契约中间层（MOS）	MOS API / 规则引擎 / 同步任务	**人与 AI 共用** 的读写契约、口径、重推	L2 补只读；L3 补写白名单与 golden
智驾运行时	Agent / RAG / 模型路由	触发→取证→推理→决策包	不直连 ERP 写库；只调 MOS 已登记接口
控制与运营层	审批流 / Console / 审计	人机权责、熔断、凭证、移交	L3 起为 **必选项**，不是「后期再加」

6.5.2 数据与控制流（L2 vs L3）

```
【L2 · 真只读】
用户/Agent 提问 → MOS 只读 API → ERP/POS 权威数据
                → 知识库 + 引用出处 → 审计日志（无写按钮）

【L3 · 受控写】
业务事件触发 → Agent 只读取证 → 生成决策包
             → 业务 Owner 审批（企微/飞书/MOS）
             → MOS 写 API（白名单）→ ERP/POS 落单
             → 回写凭证（单号/拦截记录）→ Console + 审计
```

> **交付要点**：图 H 中任何箭头若在生产环境不存在（例如 Agent 直连 ERP 写库、或只有 Excel 导出），合同应 **降级为 L2**，不在验收里写 L3。

6.5.3 MOS 与现网：谁管账、谁管契约、谁管 Agent

项目启动时建议用下表 **一次性写清边界**（写入 SKU-0《衔接模式确认》），避免「MOS 能不能替换 ERP」各说各话：

维度	管什么	并存模式（I）主责	替换模式（II）主责	Agent 能做什么	Agent 不能做什么
账（经营权威）	库存、价格、促销、订单、凭证真相；财务过账口径	**现网 ERP / 财务**	**MOS 四端 + `mos-api`**（经营单据与状态）	经审批后，调用白名单写接口改 **经营状态/配置/工单**	直连改总账、绕过审批、写白名单外接口
契约（API 口径）	鉴权、版本、幂等、口径文档、同步/重推、golden	**MOS API**（接 ERP/POS 只读或双向）	**MOS API**（自身即契约源）	只调 **已登记** 的读/写 Action	临时拼接口、未登记工具、演示环境独享写
Agent（智驾运行时）	触发、多源取证、推理、决策包	**MOS Agent**（接 MOS 契约）	**MOS Agent**（同上）	产出 **业务可读决策包**，提请人终审	持有账套、代替 Owner 签字、静默失败

两种落地模式（合同必选其一，可分区并存）

模式	适用	MOS 与现网关系	典型客户
I · 并存	已有 SAP/用友/自研 ERP + 第三方 POS，短期不动账套	MOS = **契约中间层 + 控制层 + Agent**；ERP/POS 保账务/门店权威	连锁总部、多系统并存
II · 替换	新店、系统到期、legacy POS/ERP 难 API 化	**MOS 四端** 承担管理、门店、会员、契约 API；总账/税务/集团报表 **仍对接** 企业指定财务系统或导出	新品牌、门店数字化升级、餐饮多通道

```
并存（I）：ERP/POS（账） ⟷ MOS API（契约） ⟷ Agent（推理+决策包） → 审批 → 白名单写
替换（II）：MOS 四端（账+契约） ⟷ Agent（推理+决策包） → 审批 → 白名单写 → 可选导出至总账/ERP
```

概括：YOMA **既可以把 MOS 作为现网上的契约与控制层**，也可以在经营层 **用 MOS 替换 legacy ERP/POS**；但无论哪种，**Agent 永远不「自己管账」**——只走契约层已登记的接口，且 L3 必须有人终审与凭证。
SKU-2 交付边界：指 **某一业务域**（如 PROMO、SYNC、沽清）在该模式下的 L3 闭环验收，**不等于**「所有行业、所有域均已标准交付过」。

6.6 YOMA MOS 系统解决方案（产品落地一页）

上文图 H 回答「架构怎么搭」；本节回答 **「YOMA 实际交付什么系统」** —— 把 RAL 等级落到 **可登录、可联调、可验收** 的产品组合，而不是另起一套不可验收的 Demo。

品牌与产品关系
YOMA（涌码数字） 提供 RAL 方法论、定级诊断与实施交付；
MOS 多业态智能收银平台 是落地载体 —— **管理端 · API 中台 · 智能 POS · 小程序会员** 四端一体，人与 AI Agent **共用同一套 `mos-api` 契约**。
产品资料：www.yongma.cloud（**项目启动请以顾问定级评估为准**，勿以官网自助浏览或 Demo 代替 SKU-0）

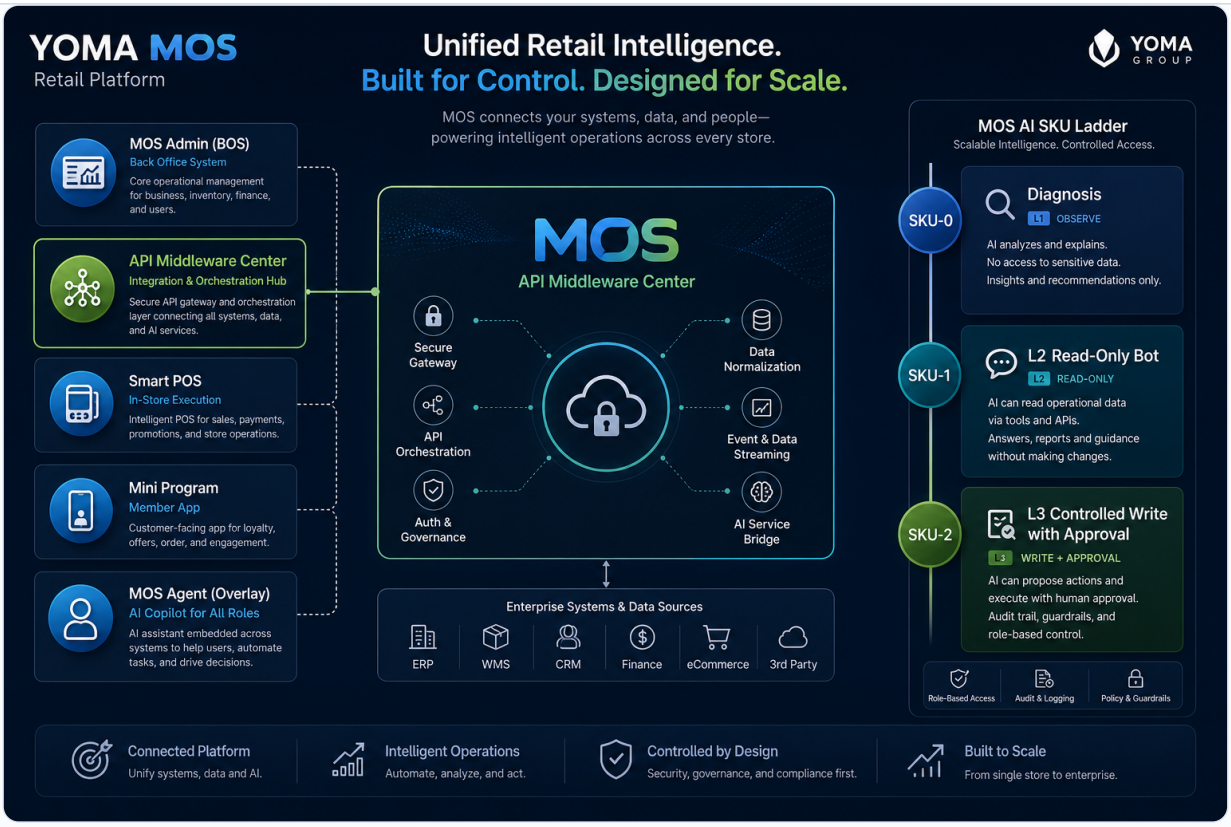
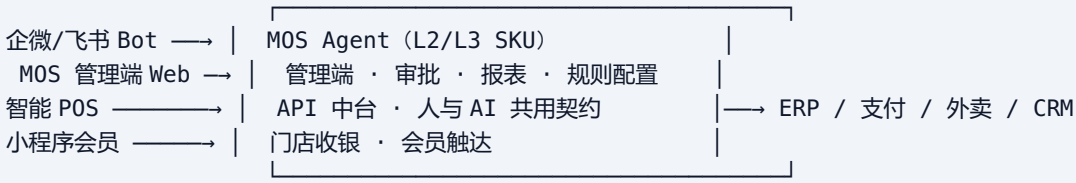


图 J · MOS 四端一体 × RAL 可交付 SKU

图 J：MOS 四端与 API 中台构成落地载体；SKU-0→SKU-1→SKU-2 为可验收的递进交付路径。

6.6.1 MOS 四端产品矩阵

端	产品名	在 RAL 中的位置	核心能力	产品模块
管理端	MOS 管理端	① 接入 + ④ 控制 + ⑤ 运营	商品主数据、库存/SCM、促销配置、组织权限、BI 报表、**业务审批流**	管理端 · 审批 · 报表
API 中台	MOS API	**② 契约层（核心）**	统一鉴权/签名、订单与促销 **单一事实源**、支付/O2O/会员/库存、定时同步	API 中台 · 促销引擎 · 订单状态机
门店	智能 POS	① 接入 + 门店执行	收银、促销、会员、日结、离线续收、称重	智能 POS · 门店收银
会员	小程序会员	① 接入 + 渠道触达	商城、自提、券/积分、本地生活核销	小程序 · 会员商城
智驾叠加	MOS Agent	③ 智能 + ④ 控制	企微/飞书 Bot、知识库+RAG、决策包、Console	MOS Agent · 接 API 中台



市面「AI 助手」多停在聊天；YOMA 以 **成熟商用的 MOS 四端产品 + `mos-api` 契约** 承载 L2「问得清」；**RAL SKU-1 / SKU-2** 是按客户现网定级后的 **验收交付组合**——每个 L3 域需单独 Pilot 验收，**不表示全行业、全场景 L3 均已标准交付过。**

产品 vs 交付（避免误读）

层次	含义	举例
MOS 产品	可部署、可联调、可运维的商用模块	管理端、API 中台、智能 POS、小程序、Agent 运行时
RAL SKU	按等级与客户域组合的 **可签字交付**	SKU-1 真 L2 抽测；SKU-2 单域（PROMO / SYNC / 沽清等）闭环
不承诺	标准产品量产范围	L4/L5；未选域的 L3；无 API 的「智能执行」

6.6.2 RAL 等级 → YOMA 可交付 SKU

交付不以「买了几个模型 Token」计价，而以 **可签字物件** 为准：

SKU	对应 RAL	客户摸到的系统	核心交付物	交付边界
SKU-0	定级诊断	工作坊纪要、KPI 卡、接口里程碑	《定级纪要》《唯一域建议》《衔接模式确认》	不承诺等级，只锁路径
SKU-1	**真 L2**	企微/飞书 Bot + 运营台 + 知识库	⩾2 路只读 API（**优先 MOS**）、20 问抽测、会话审计	**无写业务库按钮**
SKU-2	**真 L3**	单域 Pack（PROMO / SYNC / 餐饮沽清等）	决策包 + 审批 + **MOS 白名单写** + 单号/凭证 + Console + 熔断	须生产写 API；一次一域

首年推荐组合：SKU-0 → SKU-1 → **SKU-2 单域 Pilot** —— 与 §6.9 WBS、图 E 季度节奏一致；**未 SKU-1 不签 SKU-2**。

6.6.3 五层底座在 MOS 上的落点（代码可举证）

底座层	MOS 落点	说明
① 接入	POS / 小程序 / BOS Web / IM Bot	门店、会员与协同入口
② 契约	**MOS API 统一接口**	鉴权签名、促销引擎、订单状态机
③ 智能	Agent + 管理端 AI 能力	自然语言查询、经营洞察、Agent 连接器
④ 控制	管理端审批流 + Agent 白名单	券发放/出库等审批；SKU-2 熔断与放行
⑤ 运营	报表 + Agent Console	经营 BI、会话审计、L3 KPI 看板

> **实施说明**：**并存（模式 I）**—— 客户已有 SAP/用友/自研 ERP 时，MOS 作 **契约中间层**（图 H 中层），ERP 保持账务权威，促销/订单/同步经 `mos-api` 衔接。**替换（模式 II）**—— MOS 四端可为经营主责，总账/法定凭证单独约定。**无论哪种**，Agent 只调已登记接口，边界见 **§6.5.3**。

6.6.4 典型落地路径（零售 / 餐饮）

业态	现网常见态	YOMA 推荐组合	说明
已有 MOS 四端	ERP 经 MOS 同步，或 MOS 已为经营主责	SKU-1 接 MOS 只读 → SKU-2 单域	契约层最厚，L2→L3 周期最短
ERP + 第三方 POS	缺统一 API	MOS API 作中间层 + 同步任务 → 再 Agent	先模式 I + §6.7 A/B，再 L3
新店 / 系统换代	legacy ERP/POS 难 API 化	**MOS 四端替换** + Agent（模式 II）	SKU-0 确认替换边界与总账对接
餐饮多通道	堂食/外卖沽清不一致	MOS POS + 小程序 + **沽清状态 Pack**	写回走 MOS 白名单（§6.7.1 模式 B）
仅 BI / 聊天	无写 API	**SKU-0 定级** → 先 SKU-1	不硬卖 SKU-2

6.6.5 L3 落地场景一览（SKU-2）

读者常问：「L3 到底落地什么？」—— 不是再买一个聊天机器人，而是 **在 MOS 上选一个域，跑通验收六步，并拿出单号/凭证。**
下表为 YOMA **首年可签约、可验收** 的 SKU-2 场景（一次 Pilot **只选 1 行**）；完整闭环流程见 **§7.1.5～§7.1.6（零售）**、**§7.2.5～§7.2.6（餐饮）**。

场景 ID	业态	业务场景	典型触发	决策包（业务可读）	写回 Action（经 MOS 白名单）	验收凭证	核心 KPI	衔接模式
PROMO-RT	零售	促销合规拦截	活动上架/变更、大促前检	冲突 SKU 列表 + 毛利测算 + 拦截/放行建议	拦截配置 / 支付前阻断 / 放行上架	**拦截单号**或配置版本号	golden 通过率；误拦放行时效；大促熔断可用	**B**（§6.7.1）
SYNC-RT	零售	跨系统同步异常	定时对账差异、POS 与 ERP 事件	差异明细 + 建议 Action（重推/锁售/建单）	订单重推、锁售、IT 工单创建	**重推回执** / 工单号 / 锁售状态	MTTR；未结差异笔数 ↓；白名单外写入=0	**A+C**
REGEN-RT	零售	补货/调拨建议落地	安全库存预警、缺货事件	SKU×门店×建议量 + 依据	调拨/采购 **草稿单**（非直接过账）	**草稿单号**可见	建议采纳率；人工建单耗时 ↓	**B**（ERP 写接口确认后）
SOLDOUT-FB	餐饮	沽清/估清状态回写	售罄、报损、店长申报	影响菜品/规格/渠道 + 误沽风险	前厅 POS + 外卖通道 **状态同步**	**状态版本号** / 审计记录	渠道一致率；申报→全渠道生效时长	**B**

场景 ID	业态	业务场景	典型触发	决策包（业务可读）	写回 Action（经 MOS 白名单）	验收凭证	核心 KPI	衔接模式
STORE-FB	餐饮	门店任务闭环	巡店标准、开业 checklist	今日必办清单 + 异常项	任务完成/照片/签字 **回写**	**任务回执号**	完成率；稽核通过率；新店复制周期	**B**
PROMO-FB	餐饮	时段促销受控	餐段切换、活动配置变更	渠道价一致性 + 冲突检测	促销配置写入（非直接改 ERP 账）	**配置批次号**	配错率↓；生效延迟	**B**

L3 验收检查清单

- □ 书面锁定 1 个场景 ID + 试点门店/范围
- □ 生产环境写 API（非 Demo 独享）
- □ 端到端演示：触发 → 只读取证 → 决策包 → 审批 → 写回 → 单号/凭证
- □ Console 可见成功/失败行；白名单外写入抽测 = 0
- □ 业务 Owner + IT Owner 双人签字《L3 验收报告》

> **交付边界**：上表任一场景，若客户 **无生产写 API**、或 **只能 RPA 代点客户端**，合同应标 **L2 或过渡方案**，不得报 SKU-2。MKT 发券、全渠道定价等 **写权限敏感域**，一般不作首域（见 **§7.1.3、§7.2.3**）。

6.6.6 L4 多域试点场景（第二年起 · 非量产标配）

说明：YOMA **标准 SKU 交付到 SKU-2（真 L3 单域）**；**不宣称 L4/L5 已量产。**
L4 的前提是 **≥2 个域已通过 SKU-2 验收**，并补齐 **统一 Action 网关 + 跨域风控 + 调度台（D7）**（见 §3.4）。
下列为 **第二域巩固后** 可立项的 **PoC / 共创试点** —— 有路径、有验收方向，但 **不是首年标配报价项。**

试点 ID	前置（必须已有）	跨域编排（L4 在办什么）	人只处理	试点验收方向	YOMA 交付形态
L4-RT-01 临期→促销联动	零售 **PROMO L3** + 库存/临期 **真 L2**	临期预警触发 → 生成限时折扣方案 → 额度内 **自动上架**	折扣超阈值、跨品类冲突	≥N 天无人干预闭环占比；超阈值升人=100%	调度台 PoC + golden；**非标，需 L3 双证据**
L4-RT-02 促销+同步联动	**PROMO L3** + **SYNC L3**	促销变更引发库存/订单不一致 → 自动触发 SYNC Action	跨域差异无法自动消解	联动 MTTR；误自动执行=0	统一网关 + 域间编排脚本；区域试点
L4-FB-01 沽清+时段促销	**SOLDOUT L3** + **PROMO-FB L3**	沽清触发 → 关联菜品/套餐价态自动调整	误沽恢复、爆款保护	高峰渠道一致率；反向写可用	餐饮双域调度 PoC
L4-FB-02 任务+沽清	**STORE L3** + **SOLDOUT L3**	巡店发现沽清异常 → 自动生成必办 + 状态复核	食安/客诉类例外	异常闭环时长；任务与状态对账	门店营运台扩展

```
L4 立项门槛（G3→G4，与 §6.8.1 Gate 一致）
  ① 第一域 SKU-2 稳定 ≥ 一个完整业务周期（如一个大促 / 一个旺季）
  ② 第二域 SKU-2 验收通过
  ③ 统一网关审计链打通；跨域失败可补偿/可解释
  ④ 企业具名 AI 运营 Owner（非纯外包离场）
  → 不满足任一条：加固 L3，不对外称 L4
```

> **概括**：L3 是「**一个域验收通过**」；L4 是「**两个域在统一调度下日常运行，人只处理例外**」。L4 需先有双域 L3 证据，**不作为首年标准交付项**。

6.6.7 如何开始

唯一入口：先由 **YOMA AI 顾问** 主持 **RAL 定级评估（SKU-0）**，再据此形成首年实施计划；**不以官网自助浏览或在线 Demo 作为项目起点**（相关内容仍在建设中，无法替代定级与契约评估）。

步骤	动作	说明
1	**联系 YOMA AI 顾问**	通过 www.yongma.cloud 预约对接；说明业态（零售/餐饮）、现网系统概况与业务痛点
2	**RAL 定级评估（SKU-0）**	半天～一天工作坊；携带 **ERP / POS / CRM / API 现状清单**；输出《定级纪要》、真/伪 L 级结论
3	**评估首年实施计划**	顾问与客户 IT / 业务 Owner 共同确认：**唯一域**、衔接模式（§6.7）、SKU-1/2 路径与里程碑（对齐 §6.8～§6.9）
4	立项 **SKU-1 真 L2**	定级与计划签字后启动；约 4～8 周：Bot + MOS 只读 + 20 问抽测
5	单域 **SKU-2 真 L3**	SKU-1 通过后启动；约 10～12 周：决策包 → 审批 → **mos-api 写回凭证**

我们不做的事（写进合同更可信）：不宣称 L4/L5 已作为标准产品量产；不把 SKU-2 误解为「全行业 L3 开箱即用」；不自研基础大模型；不默认替换企业总账 ERP（经营层替换须模式 II 书面约定）；不把演示环境可写当作生产 L3；**不跳过 SKU-0 直接卖 SKU-2。**

> **概括**：先 **RAL 定级**，再 **MOS + API 契约落地**，按 **SKU-0/1/2** 分阶段验收 —— 从「会聊天」到「带单号的受控办事」。

6.7 与现有企业内部系统的衔接与改造

这是多数项目卡住的地方：**不是缺模型，是缺「怎么动现网」的共识。**
YOMA 将衔接方式归纳为 **四种模式**（可组合，但须写进《接口与写入授权里程碑》）。

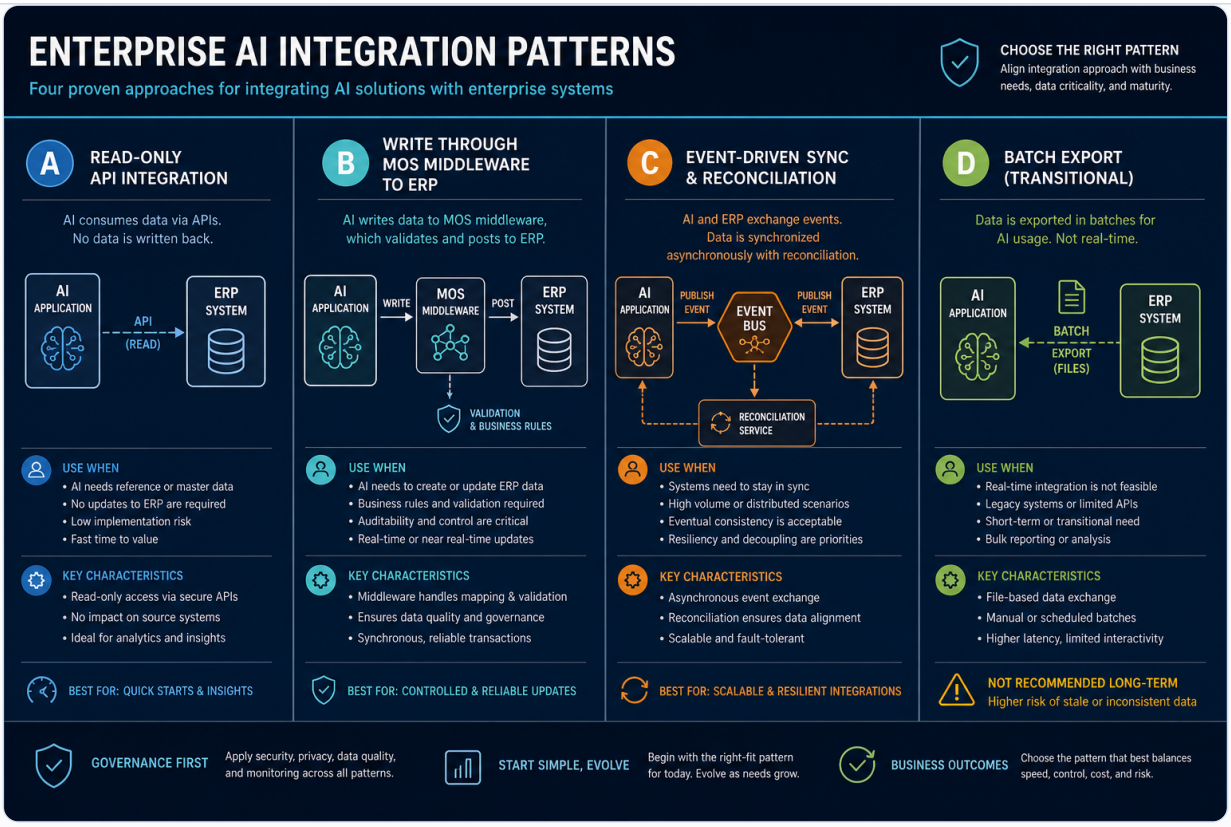


图 I · 四种系统衔接模式（只读 / 经 MOS 写 / 事件 / 批处理过渡）

图 I：A 只读 API 集成（L2 基线） · B 经 MOS 受控写（L3 主路径） · C 事件驱动对账（SYNC 域） · D 批处理/Excel 过渡（有天花板，须标注退出日期）。

6.7.1 四种衔接模式

模式	适用场景	典型动作	可达 RAL	风险与退出条件
A · 只读 API	总部问数、门店查库存/活动	开放查询契约、统一口径文档	**真 L2**	无写权限；不得称 L3
B · 经 MOS 写	促销拦截、工单重推、受控改价	MOS 登记写 API + 审批 + golden	**真 L3**	须生产环境可写；禁演示环境独享
C · 事件 / 消息	POS 与 ERP 不一致、异步对账	订阅订单/库存变更、补偿任务	L2→**L3 SYNC**	需幂等与对账报表
D · 批处理过渡	老 ERP 无 API、短期无法改造	定时导出/导入 + 人工复核	**伪 L2～伪 L3**	**必须**有 API 化路线图与退出日期

6.7.2 按系统类型的改造要点

现网系统	常见痛点	L2 最低改造	L3 额外改造	不建议
ERP	接口贵、怕 AI 改账	订单/库存/价格/促销 **只读** API	经 MOS **白名单写**（配置类、工单类）	Agent 直连 ERP 核心写库
POS	单机版、离线	云 POS 或同步网关 + 只读	沽清/锁售/小票状态 **受控回写**	用 RPA 点 POS 当 L3
CRM / 会员	券规则复杂	会员/券 **只读**	发券/改券走审批 + 频控	首域就选 MKT 发券
WMS / 履约	与 ERP 编码不一致	对齐 **主数据** 后再接	调拨/补货单经 MOS	跳过主数据开写
O2O / 外卖	多平台库存	渠道订单 **只读汇聚**	与 SYNC 域联动重推	各渠道各做 Agent
Excel / 邮件	总部习惯导出	仅作 **过渡 D 模式**	不得作为 L3 验收依据	永久依赖人工粘贴

6.7.3 主数据与权限（先于 Agent）

检查项	未通过时的后果	改造动作
商品/门店/组织编码 **跨系统一致**	L3 写错对象	主数据治理专项（可与 L2 并行）
账号 **按店/按域** 授权	审批人看到不该看的	与 MOS/企微组织同步
写 API **环境一致**（DEV/UAT/PROD）	演示 L3、生产 L2	合同锁定生产写范围
审计 **保留 ≥ 业务要求周期**	验收无凭证	与控制层同步设计

6.7.4 衔接决策流程

① 盘点现网：ERP/POS/CRM 有无只读 API？写接口谁批？
↓
② 定模式：L2 用 A；L3 优先 B，SYNC 类加 C；D 仅过渡且写退出日
↓
③ 定边界：哪些字段 AI 永不可写（金额过账、成本、权限…）
↓
④ 出里程碑：《接口与写入授权里程碑》+ 责任人 + 验收 curl/用例
↓
⑤ 联调顺序：只读抽测 → golden → 写白名单灰度 → 验收六步

6.8 YOMA 实施方法论（五阶段）

将 6.1 五步路径整理为可立项、可排期、可移交的 **D³ 五阶段**（Discover → Design → Develop → Deploy → Drive）。

阶段	英文	周期（参考）	核心目标	必交付物	客户侧关键角色
发现	Discover	0.5～2 周	定级、选唯一域、定衔接模式	《定级纪要》《唯一域建议》《衔接模式确认》	业务 Owner + IT Owner
设计	Design	2～4 周	契约、规则、审批、口径	《API 清单》《golden/用例包》《决策包模板》《审批流设计》	IT + 业务规则负责人
开发	Develop	4～10 周	L2 真只读 + L3 单域联调	只读/写连接、Agent Pack、Console、审计	IT + 实施方
部署	Deploy	2～4 周	灰度试点 → 验收	试点报告、验收六步证据、KPI 基线对比	试点店长 + Owner 签字
驱动	Drive	持续	运维移交、扩店/第二域评估	《运维移交手册》、月度复盘、变更流程	具名 Owner（不可空）

6.8.1 阶段关口（Gate）

Gate	进入条件	不通过则
G0→G1（启动 L3 设计）	真 L2 抽测通过；唯一域锁定	退回补只读 API / 知识库
G1→G2（生产灰度）	golden 全绿；写 API 在 **生产** 可用	禁止「演示环境验收」
G2→G3（L3 签字）	验收六步 + 单号凭证 + Owner 签字	不得对外称 L3
G3→G4（第二域 / L4 预研）	第一域稳定 ⩾ 一个完整业务周期	加固 L3，不并行开域

6.8.2 与 6.1 五步的对应

```
Discover  ≈  ① 定级诊断
Design    ≈  ② 站稳真 L2（设计+实现） + ③ 单域方案设计
Develop   ≈  ② L2 上线 + ③ L3 联调
Deploy    ≈  ④ 试点验收
Drive     ≈  ⑤ 复制评估 + 持续运营
```

6.9 实施计划表（首年 WBS）

下表为 **单域 L3 试点** 的参考排期（90～180 天可压缩版见备注）。
实际项目须填入 **具名负责人** 与 **依赖接口编号** —— 空白则不应启动 L3 开发。

6.9.1 总览（按阶段）

阶段	周次	里程碑	验收标准
Discover	W1～W2	G0 定级完成	《定级纪要》签字；衔接模式 A/B/C/D 确认
Design	W3～W6	L2 方案冻结	API 清单 + 口径文档 + 知识库结构
Develop	W7～W14	G1 真 L2 上线	抽测通过；审计可见；无越权写
Develop	W15～W20	L3 联调完成	golden 通过；决策包 + 审批 + 写回单号
Deploy	W21～W24	**G2 L3 验收**	六步证据 + KPI 趋势 + 运维移交
Drive	W25～W52	G3 巩固/复制	扩店或第二域评估报告

6.9.2 详细 WBS（节选 · 可展开为项目计划）

周	阶段	任务	交付物	负责人	依赖
W1	Discover	现网系统盘点（ERP/POS/CRM）	系统清单 + API 现状表	IT Owner	—
W1	Discover	RAL 定级工作坊	《定级纪要》	业务+IT Owner	系统清单
W2	Discover	唯一域选型（PROMO/SYNC/…）	《唯一域建议》	业务 Owner	定级纪要
W2	Discover	衔接模式与边界确认	《衔接模式确认》+禁写清单	IT Owner	图 I 模式 A～D
W3	Design	只读 API 契约设计	《API 清单》v1	IT	ERP/POS 厂商
W4	Design	口径与知识库结构	口径文档 + 知识目录	业务+IT	API 清单
W5	Design	L3 规则与 golden 用例	用例包 + golden	业务 Owner	唯一域锁定
W6	Design	审批流与决策包模板	审批流配置说明	业务 Owner	规则包
W7～W8	Develop	只读 API 开发与联调	L2 连接 + 抽测脚本	IT	API 清单
W9	Develop	问答入口 + 审计台	真 L2 上线	实施方	只读 API
W10	Develop	L3 写 API（MOS 白名单）	写连接（UAT→PROD）	IT	G0 通过
W11～W14	Develop	Agent Pack + 决策包 + Console	联调环境演示	实施方	golden
W15～W16	Deploy	试点店灰度	灰度报告	试点店长	G1

周	阶段	任务	交付物	负责人	依赖
W17 ～ W20	Deploy	验收六步 + KPI 采集	凭证包 + 指标对比	Owner 双人	生产写 API
W21 ～ W24	Deploy	运维移交培训	《运维移交手册》	IT+业务 Owner	G2
W25+	Drive	扩店 / 第二域评估	《复制评估报告》	业务 Owner	G3

> **压缩说明**：若只读 API 已齐，Discover+Design 可压至 3～4 周；**不可压缩**的是 G1（真 L2）与 G2（验收）之间的 golden 与生产写联调。

6.9.3 与图 E 季度节奏的对齐

图 E 季度	对应 WBS	管理看板应看到
Q1	W1～W9	真 L2 抽测通过
Q2	W10～W16	写 API 生产可用 + 灰度启动
Q3	W17～W24	**L3 验收签字**（首年唯一硬里程碑）
Q4	W25～W52	扩店或加固；未 G3 不谈 L4

6.10 能力建设清单

阶段	能力（叫什么都可以）	关键是能力不是商标
诊断	定级工作坊、差距勾选（用第五章矩阵）	产出纪要与唯一域
MOS 产品	**SKU-0/1/2**（§6.6.2）	四端 + Agent 可登录验收
L2	只读连接、知识库、协同问答、审计台	人与 AI 共用只读 API
L3	单域 Pack、读写连接、审批、决策包、Console	回写可见单号
衔接	API 契约、MOS 中间层、主数据、模式 A～D 选型	图 H/I 落地为里程碑
之后	第二域、统一网关、调度（瞄 L4）	有证据再立项

6.11 通用避坑（零售/餐饮都适用）

坑	表现	纠正
聊天即数字化	只上 Bot	用 RAL 重新定级
多域并行	五个场景各做 20%	砍到 1 个域 100%
演示写/生产读	验收后关掉写权限	合同锁生产范围
RPA 冒充 L3	脚本代点	第四章四维拆穿
无 Owner	外包一走就停	移交清单 + 具名责任人
先买模型再补 API	钱花了等级不动	先契约层
缺架构图就开工	接口责任不清	先出图 H + API 里程碑
Excel 当长期接口	模式 D 无退出日	写进合同：过渡 + API 化日期

6.12 本章小结

1. **图 H** 给出交付栈：现网权威 + MOS 契约 + Agent + 控制层，L3 写回必须可验收。
2. **§6.6 + 图 J** 给出 **YOMA MOS 四端 + SKU-0/1/2** 产品落点，以及 **§6.6.5 L3 / §6.6.6 L4 可验收场景表** —— 从等级框架到可登录、可签字验收的系统。
3. **图 I + §6.7** 回答「怎么衔接现网」：四种模式、分系统改造表、主数据先行。
4. **§6.8 D³ 五阶段 + §6.9 WBS** 把路径变成可排期、可签字的实施计划。

5. 通用路径的本质仍是：**定级 → 可读 → 可写且受控 → 可验收 → 可复制。**
   下一章在同一套交付方法下，分 **零售 / 餐饮两节** 写选域、样板闭环与汇报示例（MOS SKU 见 **§6.6.2**；衔接表见 **§6.7.2**）。

第七章 业态选域与样板闭环（零售 / 餐饮）

> 第六章的路径、产品与 WBS **两业态共用**；本章只回答：**痛在哪、首域选谁、闭环长什么样。**
> 尺子不变：**真 L2 底座 + 一个可验收的 L3 域**；差异在选域与 KPI 话术。

7.1 零售

7.1.1 真实战场

零售企业的痛，很少是「不会写海报」，而是：

```
促销一开就打架 / 毛利被打穿
POS 与 ERP 对不上，群里对账到半夜
补货靠感觉，建议单没人执行
临期看着报警，价和活动跟不上
总部问个数要翻三个系统
```

零售首年建议路径：

> **站稳真 L2（问得清）→ 选一个最疼的域做成 L3（可验收落地）。**
> 优先域通常落在 **促销纪律（PROMO）** 或 **跨系统同步（SYNC）**——二者都强依赖 API，且 ERP 聊天替代不了。

与现网衔接：ERP/POS/CRM 的改造要点与四种衔接模式见 **§6.7（图 I）**；**YOMA MOS 四端、SKU 与 L3 场景表** 见 **§6.6.2～§6.6.5**；排期与 WBS 见 **§6.9**。

7.1.2 痛点地图（假 AI → 目标等级）

#	痛点	常见假 AI	建议目标	备注
1	促销规则冲突、毛利击穿	事后报表总结	**L3 PROMO**	零售首选之一
2	POS 与 ERP 库存/订单不一致	对账群咆哮	**L3 SYNC**	零售首选之一
3	补货靠感觉	Excel 建议无人执行	L2→**L3 REGEN**	无写回则停在 L2
4	临期损耗高	仅预警	L2 分析→L3 受控改价/活动	可挂 PROMO 或独立域
5	总部问数链路长	未接真数的机器人	**真 L2**	L3 的门票
6	会员券错发	文案机器人	L3 MKT	依赖 CRM 写权限，慎选首域
7	门店执行漂移	打卡摆拍	L3 STORE	适合第二域
8	账号权限混乱	共享超管调模型	先治权限	底座问题，先于 Agent
9	演示能写、生产不能写	假 L3	合同锁生产写范围	
10	ERP 聊天当完成	FAQ	用 RAL 重定级	

7.1.3 如何选「唯一域」（决策表）

按四维打分（高/中/低），**总分最高且写 API 可达** 者优先：

候选域	业务痛感	写 API 可达性	可演示「单号」	与 ERP 边界清晰	适合作首域?
PROMO 促销合规	高	中高（配置/支付前校验）	高（拦截/放行记录）	高	**常首选**
SYNC 跨系统异常	高	中（重推/锁售/工单）	高	高	**常首选**
REGEN 补货调拨	高	中（视 ERP 开放度）	高（有单号才算）	中（易撞 ERP 补货）	写回确认后再选
临期闭环	中高	中	中高	中	可作 PROMO 子场景
MKT 发券	中	低～中	中	中	一般不作首域
STORE 门店任务	中	中	中	高	适合第二域

选型要点：
- 决策层最关心「错促/对账翻车」→ PROMO 或 SYNC；
- 门店最怕「缺货」但 ERP 不给写 → 先 L2 建议，不硬报 L3。

7.1.4 首年默认路径（零售）

> 详细 **90～180 天 WBS、Gate 关口与季度对齐** 见 **§6.8～§6.9**；MOS 产品组合见 **§6.6.2**；下图与图 E 一致，零售仅替换「唯一域」为 PROMO 或 SYNC。

```
Q1  定级 → 打通订单/库存等只读 API → 真 L2 问答上线
Q2  锁定 PROMO 或 SYNC → 规则/golden → 决策包 → 写白名单联调
Q3  试点店验收（单号/拦截凭证）→ 移交 Owner
Q4  扩店巩固；评估是否具备第二域条件
```

能力组合：**真 L2（只读+知识+审计） + 单一 L3 域。**

7.1.5 样板闭环 A｜促销合规（PROMO）

目标：减少错促上架与毛利击穿；拦错可解释、可一键放行（在权限内）。

```
活动配置/变更触发
  → 只读：商品、价格、已有规则、毛利门槛、互斥表
  → 推理：冲突/毛利/渠道规则
  → 决策包：冲突列表 + 建议（拦截/修改/放行条件）
  → 人审：营运批准或驳回
  → 写回：允许上架 / 保持拦截；记录进 Console
  → 凭证：拦截单号或配置版本号可见
```

验收关注：golden 用例通过率；拦截次数与误拦放行时效；大促熔断开关可用。

API 要点：规则与配置可读；拦截/放行状态可写；支付前或上架前校验位可挂接。
衔接模式：优先 **B · 经 MOS 写**（§6.7.1）；ERP 只读未开则先 **A 模式** 站稳 L2。

7.1.6 样板闭环 B｜跨系统同步异常（SYNC）

目标：缩短 POS 与 ERP（或中台）不一致的发现与处置时间。

```
定时对账 / 异常事件
  → 只读：两侧单据、库存、状态码
  → 推理：差异类型（未过账、重复、编码映射失败…）
  → 决策包：差异明细 + 建议 Action（重推/锁售/建 IT 单）
  → 人审：信息部或营运确认
  → 写回：执行 Action；业务侧可见结果
  → 凭证：重推成功回执 / 工单号 / 锁售状态
```

验收关注：MTTR、未结差异笔数趋势、白名单外写入=0。

API 要点：两侧只读对齐；重推/锁售等写接口明确；禁止用 RPA 点客户端冒充本闭环。
衔接模式：只读对齐用 **A**；异常补偿用 **C · 事件/消息**；过渡期 Excel 对账须标 **D 模式退出日**（§6.7.1）。

7.1.7 样板闭环 C｜临期处理

临期适合讲「店长听得懂」的故事，但建议分层说明：

层级	做到	等级
仅预警+报表	列表提醒	L2
预警→打标→通知→**人批后**分时折扣/活动	有写回	**L3**（常挂促销域）
多域自动（临期触发营销自动上架且跨风控）	调度台	L4 展望

没有改价/活动写接口，就不要把临期 Demo 说成 L3。

7.1.8 零售避坑清单

#	坑	纠正
1	把「ERP 里的 AI 助手」当 L3	问能否受控改你的促销/同步单据
2	首域选发券（MKT）但 CRM 不开放写	改选 PROMO/SYNC，或先谈写权限
3	补货 L3 承诺了，ERP 只许人工建单	降级为 L2 建议单，或改写中间层
4	全渠道一口吃（线上+线下+会员）	首年单域单渠道切片
5	试点选旗舰店特例数据	选「疼但典型」的店，否则无法复制
6	规则全靠 Prompt 临场发挥	关键互斥/毛利必须进规则或用例集
7	不设大促熔断	控制层补开关，否则信息部不敢开写
8	成功只汇报「调用次数」	改汇报拦截质量、差异笔数、人工耗时

7.1.9 阶段成果汇报示例（零售）

试点验收后向决策层或经营层汇报时，可参考下列表述（将括号内内容替换为实际域与指标）：

> 首年我们在〔促销合规 / 跨系统同步〕完成单域试点验收：AI 输出可审计的处理建议，经业务负责人审批后写入现网系统，全程留痕、可查单号；**并存模式下 ERP 仍为账务与主数据权威；若采用 MOS 替换经营层，总账/法定凭证仍按企业约定对接，未做未约定的系统替换。**
> 试点期间，〔拦截质量 / 差异处置时效 / 对账人工耗时〕等指标呈改善趋势。
> 建议下一阶段：**在同一域内扩店复制、完成运维移交**；第二域及更高自主等级，待 KPI 连续达标后再单独立项。

汇报时建议少讲：模型名称、对话条数、场景数量。
多讲：哪个域、哪几家店、哪几个可核对凭证、业务 Owner 是否签字。

7.2 餐饮

7.2.1 真实战场

餐饮与零售共用 RAL 尺子，但战场不同：

```
估清/沽清靠群里口播，堂食与外卖各说各话
午餐高峰一过，建议单已过期
时段促销配错，渠道价不一致
门店任务靠打卡摆拍，标准难复制到新店
原料损耗与菜品卖力对不上
```

餐饮首年建议路径：

> **站稳真 L2（把餐段与状态问清楚）→ 选一个「过了这顿就废」的域做成 L3。**
> 优先域常落在 **沽清/估清状态回写** 或 **门店任务闭环**；时段促销在写配置 API 可达时也很适合。

与现网衔接：前厅 POS、外卖中台、后厨系统的 **状态写回** 多走 **B · 经 MOS 写**；渠道不一致场景叠加 **C · 事件同步**（§6.7）；MOS 产品、**L3 场景表（§6.6.5）** 与 WBS 见 **§6.6～§6.9**。

C 端 vs B 端（先划清）：
点餐小程序的「猜你喜欢」是 **C 端推荐**，不等于经营侧 **AI 数字化**。本白皮书讲的是 **店长/营运/总部怎么受控办事**。

7.2.2 痛点地图（假 AI → 目标等级）

#	痛点	常见假 AI	建议目标	备注
1	估清/沽清不及时	群里口播	**L3 状态回写**	餐饮常首选
2	餐段/天气/活动预估不准	事后复盘长文	**真 L2**→再谈自动要货	无写回先别报 L3
3	时段促销配置慢、易错	文案生成	**L3 促销受控**	写配置 API 必备
4	门店任务难稽核	打卡形式主义	**L3 任务闭环**	常首选/次选
5	原料损耗与菜品关联弱	单点报表	真 L2 多源只读	L3 门票
6	连锁标准难复制	培训视频	L2 知识库 + L3 任务包	
7	外卖/堂食价格不一致	人工改价	L3 价格域受控	渠道写权限敏感
8	高峰没空看建议	建议堆在 App	决策包进审批/必办	产品形态问题
9	后厨与前厅系统割裂	两个机器人各说各话	**先 API 统一**	底座优先
10	C 端推荐当经营 AI	「我们小程序有 AI」	用 RAL 重定级	

7.2.3 如何选「唯一域」（决策表）

候选域	业务痛感	写 API 可达性	可演示凭证	适合作首域?
沽清/估清状态	高	中高（状态写前厅/外卖）	高（各渠道状态一致）	**常首选**
门店任务 STORE	高	中	高（完成回执）	**常首选**
时段促销	中高	中（视促销配置开放）	高	写配置确认后
要货/补货	高	低～中	有单号才算 L3	多数先 L2
渠道价格对齐	中	低～中	中	慎作首域
会员发券	中	视 CRM	中	一般不作首域

选型要点：
- 店长最烦「卖完了外卖还在卖」→ 沽清状态；
- 总部最烦「标准落不了地」→ 任务闭环；
- 只有预测报表、没有状态写接口 → 应先建设真 L2。

7.2.4 首年默认路径（餐饮）

> WBS 与 Gate 模板见 **§6.9**；MOS SKU 见 **§6.6.2**；餐饮将唯一域替换为「沽清状态」或「门店任务」即可复用同一计划表。

```
Q1  定级 → 打通销售/沽清/任务等只读 API → 真 L2（问得清餐段与原因）
Q2  锁定沽清或任务域 → 决策包（必办而非长文）→ 写白名单联调
Q3  试点门店验收（状态/回执可见）→ 移交店长+区域 Owner
Q4  扩店；评估时段促销作第二域
```

7.2.5 样板闭环 A｜沽清 / 估清受控回写

目标：减少「堂食已沽清、外卖仍下单」；状态变更可审计。

触发：售罄规则 / 店长申报 / 后厨报损
→ 只读：当日销量、余量、渠道在售状态
→ 推理：建议沽清范围（菜品/规格/渠道）
→ 决策包：影响渠道 + 风险（是否误沽爆款）
→ 人审：店长/值班经理确认（高峰可简化为双人中一人）
→ 写回：前厅 POS + 外卖通道状态一致
→ 凭证：状态版本号 / 操作审计可查

验收关注：渠道一致率；从申报到全渠道生效时长；误沽恢复（反向写）是否可用。

API 要点：状态可读可写；渠道映射清晰；禁止只改群公告不改系统。

7.2.6 样板闭环 B｜门店任务闭环

目标：卫生/陈列/出品/开班检查从「拍照摆拍」变为「任务→确认→可抽检」。

触发：日计划 / 巡检模板 / 异常事件
→ 只读：门店档案、历史完成率、当日排班
→ 生成：今日必办清单（可按餐段拆分）
→ 决策包：任务项 + 标准要点 + 截止时间
→ 人执行并确认（可含简易凭证）
→ 写回：完成状态；未完成升级区域
→ 凭证：任务回执 ID；总部可抽检

验收关注：准时完成率；抽检一致率；区域干预次数下降。

7.2.7 样板闭环 C｜时段促销受控配置

目标：午高峰/晚市促销配置可批可回滚，减少配错价。

触发：营运创建时段活动
→ 只读：菜品、渠道价、互斥规则
→ 决策包：生效窗口 + 渠道 + 毛利风险
→ 人审 → 写配置 → 配置版本号可见

无配置写 API = 最多 L2 方案草稿。

7.2.8 餐饮避坑清单

#	坑	纠正
1	把 C 端推荐当 L3	问 B 端有无受控写状态/任务/配置
2	高峰推长文建议	改必办决策包 + 一键确认
3	只改企微群，不改外卖状态	闭环必须写到渠道系统
4	后厨与前厅两套码	先主数据/API，再 Agent
5	要货 L3 承诺了但供应链无写	降级 L2 建议
6	全国一套规则无视餐段	规则按餐段/天气参数化
7	试点选旗舰全员盯梢	选典型店，否则无法复制
8	成功只报「AI 回答条数」	改报一致率、完成率、误沽恢复时长

7.2.9 阶段成果汇报示例（餐饮）

> 首年我们在〔沽清状态 / 门店任务〕完成单域试点：审批后各渠道状态一致、任务回执可查，**未将小程序推荐等同于经营数字化完成。**
> 试点门店〔渠道一致率 / 任务完成率 / 误沽恢复时长〕等指标达到约定门槛。
> 建议下一阶段扩店复制；时段促销等第二域，在 API 与运维就绪后再立项。

汇报时建议少讲：「上了 AI」「回答更智能」。
多讲：高峰沽清是否还靠群、任务是否可稽核、谁对审批结果负责。

7.3 零售与餐饮对照（一表速览）

维度	零售	餐饮
时间敏感	日/周活动、对账周期	**餐段/小时**；建议过期即废
典型首域	PROMO / SYNC	**沽清状态 / 门店任务**
跨系统痛	POS 与 ERP 主数据与过账	前厅、后厨、外卖通道状态
「单号」形态	拦截单、重推回执、调拨单	沽清状态版本、任务回执、促销配置号
假 AI 高发	ERP 聊天、RPA 代点	C 端推荐、群口播数字化
共用前提	真 L2 只读 API + 单域 L3 受控写	同左

7.4 本章小结

- **零售**胜负手在 **促销纪律与系统一致**；**餐饮**在 **实时状态与门店执行**。
- 两节共用 RAL 尺子与第六章交付栈；**选对唯一域 + 全 API 化**，比堆场景名重要。
- 样板闭环与场景 ID 对照：**§7.1.5～§7.1.6（零售）**、**§7.2.5～§7.2.6（餐饮）**；完整 SKU-2 列表见 **§6.6.5**。

PART III

治理篇

风控、组织与运维

第九章 治理、风险与「什么时候不该上」

9.1 为什么治理不是附录废话

L3 一旦开写，事故形态从「答错一句话」变成「错单、错价、错沽清」。
故：**控制层与 Owner 机制与智能层同级重要。**

9.2 治理清单（L2 起建议，L3 必备）

项	L2	L3	说明
会话/调用审计	●	●	谁问了什么、调了哪些只读
Agent 身份与人身份分离	○	●	禁止共用超管
白名单写接口	—	●	名单外默认拒绝
审批人与代理人	—	●	含休假备份
金额/频率/毛利阈值	—	●	超阈强制人
熔断开关（大促/高峰）	—	●	一键降回只读
回滚/补偿剧本	—	○/●	按域定义可逆性
数据脱敏与最小化	●	●	尤其会员与支付
供应商退出与密钥轮换	○	●	防止人走政息

9.3 风险类型与应对

风险	例子	应对
幻觉当决策	编造库存建议直接写	关键数必须来自 API；决策包强制引证据
越权写入	改到白名单外接口	网关拒绝 + 抽测
误拦/误沽	爆款被错拦	一键放行/恢复 + 事后复盘
提示词注入	诱导「忽略审批」	工具层不认自然语言越权；审批在工作流外不可绕
供应商锁定	只有对方能改规则	规则/用例导出；API 契约归企业
组织空心	外包一走系统停	移交清单 + 具名 Owner

9.4 与 ERP / 财务边界（统一口径）

与 **§6.5.3** 一致，合同须写明 **模式 I（并存）或模式 II（替换）**：

- **模式 I · 并存**：账套真相仍在企业现有 ERP；AI 经 MOS 白名单写经营状态/配置/工单，再按现网过账；不擅自替换总账与法定凭证体系。
- **模式 II · 替换**：MOS 四端可为 **经营层主责**（管理、门店、会员、契约 API）；总账、税务、集团合并报表等 **仍对接** 企业指定财务系统或按约定导出。
- **Agent 统一边界**：不直连改账、不绕审批、不写白名单外接口；不替代财务制度与法定凭证要求。

9.5 什么时候不该上（或该降级）

情况	建议
业务系统无只读 API，短期也无法暴露	停在 L1 工具，或先做 API 化专项
有只读、无写、且 90 天内无望开放	只卖/只建真 L2，不承诺 L3
无业务 Owner，只有信息部热	先补组织，再开 Pilot
想一次上 L4/L5	拒绝；先要 1 个真 L3 证据
合规禁止自动改价/改库存	选其它域，或审批必须双人
以「赶大促明天上线」为唯一目标	改演示级，不写进 L3 验收

9.6 本章小结

治理的目标不是把 AI 管死，而是 **让信息部敢于开写、让业务敢于终审。**
熔断与白名单，是 L3 的安全带，不是可选项。

第十章 行动倡议

10.1 不同角色的行动重点

读者	本周	90 天内	一年内
决策层	要求一张定级结论（真/伪 L 级）	盯一个域的验收指标，不盯「又上了几个机器人」	巩固扩店；无第二域证据不批 L4 立项
数字化 / IT	用第五章矩阵勾选契约层缺口	打通只读；为唯一域准备写白名单与审计	API 版本与权限治理常态化
营运 / 门店	列出最疼的 3 个损失场景	当审批人真批决策包；反馈「看不懂」的字段	养护规则与知识；带教新店

10.2 项目实施检查表

步骤	动作	完成标志	负责人	日期
1	定级工作坊	《定级纪要》签字		
2	勾选底座矩阵	缺口清单		
3	锁定唯一域	一页选型理由		
4	真 L2 抽测	约定题通过		
5	L3 闭环验收演示	单号/状态凭证		
6	运维移交	Owner 具名		
7	扩店/第二域评估	通过/暂缓书面结论		

10.3 结语

1. **用「AI智驾」等级讨论 AI 数字化**，少用「赋能」「引擎」空话。
2. **用 API 化衡量能不能升级**，少用模型商标衡量先不先进。
3. **用验收六步验收 L3**，拒绝 RPA 与演示级效果冒充受控执行。
4. **零售与餐饮都走同一条台阶**，选域不同，标准一致。

问题不是「要不要用 AI」——多数企业已经在 L1。
真正的问题是：**在可控前提下，你们能稳定站在哪一级，且能否用凭证核对。**
每一级都建立在前一级之上：先真 L2 的只读，再谈真 L3 的受控写；无第二域证据，不急于推进 L4。

我们不卖又一个聊天窗口。
我们主张：先定级，再建设；先可读，再受控可写；先在一个域验收通过，再谈扩域。

附录 A RAL 定级自测（10～15 题）

用法：每题选最接近现状的一项；按文末计分粗判等级。零售/餐饮共用，题干按业态理解即可。

Q1 经营数据怎么进 AI?
A. 主要靠复制 Excel/截图　B. 有稳定只读 API 或日级同步　C. 准实时只读且口径文档齐全

Q2 AI 能否改业务单据/状态?
A. 完全不能　B. 演示可以、生产不能　C. 生产可写但仅白名单+审批　D. 多域可自动写+风控

Q3 输出物主要是?
A. 文案/FAQ　B. 分析建议（人自己去点系统）　C. 带明细的决策包　D. 多域自动工单流

Q4 有无审批与审计?
A. 无　B. 仅会话日志　C. 审批+写操作审计　D. 跨域统一网关审计

Q5 与 ERP/核心系统关系?
A. AI 想替换账套　B. 无关的聊天窗　C. 账套仍在 ERP，AI 经中间层/白名单受控写　D. 已 AI 原生重构

Q6 是否选定唯一业务域做深?
A. 无，场景很散　B. 有候选但未锁定　C. 已锁定 1 域并试点　D. ≥2 域均稳定自动跑

Q7 失败时能否解释?
A. 不知道为何失败　B. 人肉查日志　C. Console 可见失败码与行级结果　D. 自动补偿/升级例外台

Q8 一线入口?
A. 几乎不用　B. 偶尔问 Bot　C. 审批/必办进企微或 MOS　D. 例外工作台日常化

Q9（零售）促销或对账最痛点是否已有系统凭证级处置?
A. 靠群和 Excel　B. 有报表无处置　C. 有拦截/重推等凭证　D. 多域联动自动处置

Q9（餐饮）沽清或任务是否全渠道/可抽检回写?
A. 群口播　B. 只改一个渠道　C. 审批后多通道一致　D. 多域自动+例外

Q10 组织 Owner?
A. 无　B. 仅 IT　C. 业务+IT 双 Owner　D. 还有专职 AI 运营

Q11 API 化程度?
A. 逻辑锁在客户端　B. 部分只读 API　C. 前后端分离，关键经营能力 API 化　D. 统一 Action 网关

Q12 对 L4/L5 态度?
A. 当作今年必须交付　B. 当作宣传词　C. 当作 L3 之后的展望　D. 已有 ≥2 域 L3 证据并立项平台

计分（粗判）

主导选项	粗判
多为 A	L1 或伪数字化
多为 B	真/伪 L2 之间——查 API 与审计
多为 C	**瞄准或已在 L3**——用第四章八条判据复核
出现多个 D 且 C 扎实	才讨论 L4；L5 仍须单独战略评估

硬性判定： 无生产写权限却对外称 L3 → 判定为 **假 L3**。

附录 B 术语表

术语	含义
RAL	Retail / F&B Autonomy Levels；YOMA **「AI智驾」** 即 **AI 数字化自主等级** L1～L5
智驾助记	脱脚/脱手/脱眼/脱脑/全域，对应 L1～L5，仅助记
真 L2	经营数据经正式只读 API，可审计问答；非 Excel 旁路
假 L3	演示可写或 RPA 代点，生产无白名单写与审批审计
决策包	含明细、依据、风险、建议 Action 的可审执行方案
验收六步	触发→取证→推理→决策包→审批→写回凭证
白名单	允许 AI 调用的写接口集合，名单外默认拒绝
契约层	业务 API（鉴权、版本、幂等、审计）；人与 AI 共用
Action	Agent 调用的写/读工具；L3 起含受控写
golden	可回归的规则/用例集，用于验收与防回归
Owner	业务与 IT 具名责任人，对规则与权限负责
熔断	紧急关闭写入、降级只读的开关
C 端推荐	面向消费者的推荐/导购，≠ 经营侧 RAL
RPA	模拟界面点击的自动化；可作桥，不计入 L3 定义

附录 C 实践摘要（脱敏示例）

以下为脱敏表述示例，不代表特定客户承诺：

> 「某连锁零售在促销互斥场景完成决策包→审批→配置拦截，试点期错促客诉下降。」
> 「某餐饮品牌将沽清从群口播改为审批后多通道状态写回，外卖误下单减少。」

具体数字用区间或定性表述，避免可识别信息。

附录 D 参考文献与延伸阅读

- YOMA：《MOS RAL 零售 AI 自主等级评估与建设方法论》
- 汽车自动驾驶分级思路

免责与边界

- 本白皮书为 **YOMA** 方法论与建设指南，不构成对特定客户的交付承诺。
- **MOS 可与现网 ERP/POS 并存，也可在经营层替换 legacy 系统**；总账/法定凭证边界见 §6.5.3；L3 写入以白名单与审批为准。
- 不宣称 L4/L5 已作为标准产品量产；SKU-2 指 **单域 L3 验收交付**，非全行业、全场景均已标准交付。
- 正文不含商业报价；合作范围以合同与 SOW 为准。
- 文中提及的行业公开方案仅供对照学习，不代表关联或背书。
- 定级自测仅为自我评估工具，正式定级以工作坊与接口盘点为准。

Yoma

YOMA · AI 智驾

用自然语言 AI，用 API 养 AI

先定义，再推进，先可读，再受控可写，

先一个模块成，再完全填。

www.yongma.cloud