CUBIC CONSULTING · 2026

把经营判断变成组织能力

企业 AI 转型白皮书

企业 AI 转型的目标，是要让经营判断、决定、行动、结果与学习形成持续运转的闭环。



© 2026 深圳市库比克咨询有限公司。保留所有权利。

允许在内容完整、未经修改并保留本版权声明的前提下，非商业性转发本白皮书。未经书面许可，不得摘编、改编、汇编、出版、销售或将本白皮书全部或部分内容用于商业用途。

CONTENTS

目录

序言 · 第一部分 · 第二部分 · 第三部分

序言

序言：AI 怎样参与企业经营判断 3

第一部分：先把转型拉回经营

第一章：CEO 怎样推动企业 AI 转型 6

第二章：AI 进入真实工作，需要解决六类问题 7

第三章：企业该用什么标准判断 AI 转型有没有成效 8

第四章：不同角色怎样使用 AI 9

第二部分：把专家判断教给 AI

第五章：怎样把 CEO 和专家的经营判断教给 AI 10

第三部分：让 AI 进入真实工作

第六章：企业怎样把关键信息安全地提供给 AI 12

第七章：Ontology(本体)怎样把资料变成业务世界 15

第八章：Graph(工作拓扑)怎样组织企业的真实工作 17

第九章：Agent(智能体)怎样登记、协作和留痕 19

第十章：怎样判断 AI 产物是否可用 22

第十一章：企业怎样统一分配 AI 权责并复核结果 23

CONTENTS

目录

第四部分 · 第五部分 · 结语 · 附录

第四部分：管理者与人才培养

第十二章：AI 时代的管理者应该创造什么价值 27

第十三章：自动化之后，企业怎样培养判断力和新人 28

第五部分：从一个试点走向组织能力

第十四章：怎样用 90 天验证一个有价值的 AI 场景 29

第十五章：一个试点成功之后，怎样推广到更多业务 32

结语：企业需要持续积累经营判断 33

附录

附录 A：CEO 的一页转型问题单 34

附录 B：专家判断案例卡 35

附录 C：Context(上下文)最小提供清单 36

附录 D：最小 Graph(工作拓扑)卡 37

附录 E：Agent(智能体)身份、协作合同与收据 38

附录 F：Eval(评测)与 Rubric(评判标准)起步包 40

附录 G：结果证据分层表 41

附录 H：90 天试点节奏 42

附录 I：扩大使用前的检查清单 43

附录 J：术语表 44

序言：AI 怎样参与企业经营判断

企业过去用软件执行已经明确的规则，如今开始用 AI 解释信息、比较方案、识别风险和推进工作。技术能力由此进入了经营判断过程。

本书用 AI 指包括大模型、Agent(智能体)、工具、数据和组织机制在内的整体能力；只有特指底层的通用生成、理解和推理能力时，才使用国内常用的说法大模型。

本书结合公开资料与 Cubic Consulting 的项目实践，提出一套观察与行动框架，供企业在自己的业务中验证。

企业需要积累自己的判断

大模型和 Coding Agent(编程 Agent)正在内置规划、工具调用、测试、重试和 Eval Harness(评测工具链)等工程能力。企业可以复用这些能力，把主要精力放在自己的业务 Context(上下文)、案例和取舍标准上。

企业仍要自己回答：什么值得做，哪些事实重要，什么风险不能接受，什么结果会改变下一轮行动。这些经营哲学、案例和取舍标准，构成 AI 在企业中工作的核心内容。

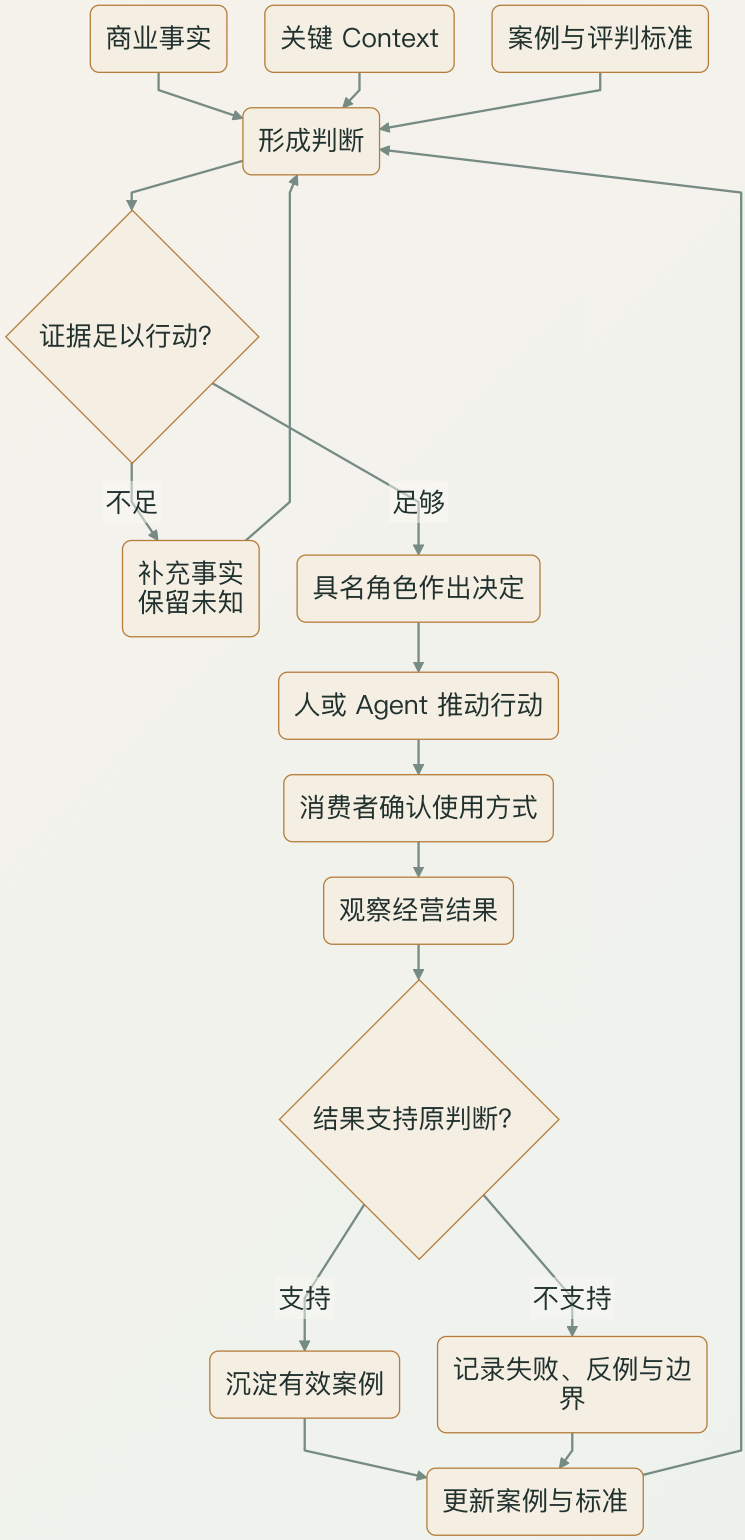
本书沿一条主线展开：

> **商业事实 → 判断 → 决定 → 行动 → 结果 → 学习 → 下一轮判断**

企业把关键事实、Context(上下文)、判断依据和适用条件记录下来，供人和 AI 在工作中调用；行动产生结果后，再修订案例和 Rubric(评判标准)。这个循环持续运转，就形成了经营判断飞轮。

业务系统图

经营判断飞轮



从一项真实工作开始

合适的起点应当反复发生，包含重要经营判断，有明确的负责人和使用者，并能在有限时间内观察行动与结果。团队先还原当前工作，再选择 AI 可以参与的一段，例如整理 Context(上下文)、比较方案、提示风险或跟踪行动。

试点需要回答三个问题：AI 是否改善了判断或行动；企业是否具备可靠的 Context(上下文)和结果评测；这项投入应该继续、调整还是停止。

给 CEO 的阅读地图

你现在要决定什么	优先阅读
是否启动企业级 AI 转型	第一至第三章
怎样选择第一个真实场景	第三、第十四章
怎样把专家经验教给 AI	第五章
怎样把企业资料变成可靠的业务事实	第六、第七章
AI 可以自主执行到哪一步，谁对结果负责	第十一章
怎样控制敏感信息和 Agent(智能体)权限	第六、第十一章
怎样复核高影响的 AI 决定	第十一章
怎样判断试点是否产生经营价值	第三、第十、第十四章
怎样把成功试点推广到更多业务	第十五章

若时间有限，先读第一、第五、第八、第十和第十四章。它们分别说明谁来推动、怎样收集判断、怎样把判断放进真实工作、怎样评测产物，以及怎样用试点验证经营价值。

第一部分：先把转型拉回经营

第一章：CEO 怎样推动企业 AI 转型

CEO 推动 AI 转型，主要完成三件事：选择值得改变的经营问题，提供关键业务 Context(上下文)，定义有价值的结果。

1. **选择经营问题。** 场景应当反复发生，包含重要判断，并可能改变收入、成本、风险、客户体验或组织学习。新品 GTM(新品上市与市场进入)中的目标客户、价值主张和上市范围，就是一组能够连接经营结果的判断。

2. **提供关键 Context(上下文)。** 企业需要说明自己为什么这样做：哪些客户值得投入，什么风险不能碰，什么质量标准不能降低，过去哪些失败改变了判断。缺少这些内容，AI 只能给出接近行业平均水平的答案。

3. **定义结果标准。** 使用人数、生成次数和节省工时只能说明活动发生。企业还要观察产物能否使用、使用者是否接手、行动是否改变、结果是否在约定时间出现，以及其他因素能否解释同一结果。

第二章：AI 进入真实工作，需要解决六类问题

AI 已经能写、能搜、能总结、能编程，也能调用工具连续完成任务。它进入真实工作时，常遇到六类问题：

1. **问题没有被定义。** 项目从建设助手开始，缺少最近一次失败、返工或高成本判断，团队只能围绕功能设计演示。
2. **Context(上下文)不可得。** 关键背景分散在个人经验、聊天、会议、旧文件和客户现场。资料接入后，版本、来源、权限和有效性仍然不清楚。
3. **交接断裂。** 产物生成后没有明确的使用者，也没有进入后续工作。
4. **流程没有改变。** 员工得到新工具，却继续按旧流程、旧考核和旧权限工作，节省的时间被新的复制、审核和汇报吃掉。
5. **价值无法判断。** 企业记录调用量和活跃人数，却没有记录哪项行动因此改变。
6. **管理机制没有改变。** 会议代替决定，层层审批稀释责任，汇报维护表面进度。AI 只会让材料更多。

这六类问题需要在同一项真实工作中解决：输入来自哪里，什么判断关键，什么失败代价最高，产物进入哪一步，结果何时可见。培训、工具和流程调整都围绕这条工作链展开。

企业该复用什么，积累什么

能力	企业的处理方式
大模型和 Agent(智能体)的通用能力	优先复用产品提供的规划、生成、工具调用、测试和重试
企业基础控制	使用全书统一的身份、数据边界、权责分级、记录和恢复方式
场景化经营判断	持续积累 Context(上下文)、案例、反例、Rubric(评判标准)和结果反馈

企业可以购买或更换工程工具链，装入工具链的业务标准需要由自己掌握，并且能够导出、复跑和迁移。

第三章：企业该用什么标准判断 AI 转型有没有成效

使用人数、生成次数、节省工时和 Eval(评测)分数只能证明系统被调用或完成了任务。经营成效需要六层证据：

1. **任务能力**：AI 能否稳定完成目标任务；
2. **产物**：输出是否正确、可用、安全、可追溯；
3. **采用**：真实使用者能否独立接手；
4. **行动**：判断或执行是否因此改变；
5. **结果**：收入、成本、风险、体验或学习是否在窗口内变化；
6. **归因**：AI 对变化贡献了多少，季节、价格、人员或其他项目能否解释同一变化。

每一层都要单独验证。Eval(评测)通过不代表有人使用；有人使用不代表行动改变；结果出现后还要检查其他解释。

试点开始前应写下事前判断：预期哪项指标变化，多久能看见，什么变化有意义，什么结果会推翻原判断。控制成本也要计入净收益，包括复核时间、错误代价、接管和恢复。

美国国家经济研究局的一篇工作论文分析了单一企业 5,179 名客服人员分批使用生成式 AI 助手的数据，观察到平均生产率提高，新人和低技能人员的增幅更大。这项证据只适用于特定企业、工作和时期，不能替代企业自己的六层验证。

第四章：不同角色怎样使用 AI

企业不需要所有人都深度操作 AI。使用方式取决于一个人在工作中的位置：

1. **建设 AI**：配置大模型、数据、工具和运行方式，需要理解技术能力与失效方式；
2. **教 AI 判断**：把专业案例、Rubric(评判标准)和例外处理交给 AI，需要亲自比较和修订产物；
3. **直接使用 AI**：提交问题、继续编辑、调用工具或推进任务，需要识别错误并提供反馈；
4. **消费 AI 结果**：阅读已经整理和复核的结论，用它作出决定或完成后续工作，不必掌握复杂操作；
5. **受到 AI 影响**：可能不使用系统，但会受到服务、价格、排班或其他决定影响，需要能够查看说明和提出问题。

这五类角色可以重叠。同一个人在一个场景中建设 AI，在另一个场景中只消费结果。企业要按真实工作提供培训，不以账号数量或统一考试制造形式上的普及。

经合组织关于技能的研究指出，多数员工更需要数字能力、数据解释和与任务相关的应用技能，高级 AI 研发技能只适用于少数岗位。企业应把复杂能力放在少数需要深度使用的人身上，让其他人能够清楚地消费结果、反馈错误和请求人工处理。

第二部分：把专家判断教给 AI

第五章：怎样把 CEO 和专家的经营判断教给 AI

企业稀缺的知识，常藏在专家面对不完整信息时的取舍中：什么值得继续，哪份方案能够交付，什么风险需要立即暴露。提取这类判断，要从最近的真实案例开始。

团队让专家重走一次工作：当时知道什么，不知道什么，哪些线索改变了选择，放弃了哪些方案，谁的意见改变了判断，实际结果是什么。成功、失败和仍在进行的案例需要一起使用，分别暴露有效做法、失效边界和当前未知。

一张可用的判断案例至少包括：情境、目标、关键线索、候选方案、权衡、行动、结果、反事实和适用边界。用于教学或 Eval(评测)时，题面只开放事前信息，实际行动和结果在作答后揭晓。

AI 学习的重点是情境、线索、权衡、行动和结果之间的关系。同一句原则在不同情境里可能导向不同选择，因此案例还要保留反例、停止条件和会推翻原判断的新证据。

教学可以分四步：

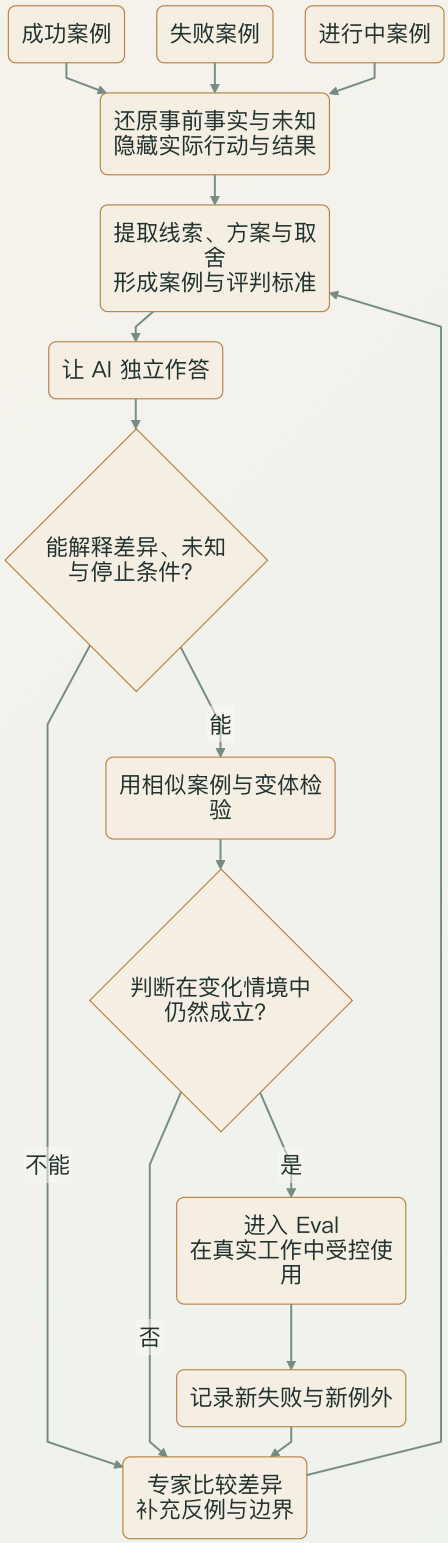
1. 让 AI 解释案例中的关键线索；
2. 提供相似案例，并指出关键差异；
3. 用反例和变体检验已有规则；
4. 由专家修订 Rubric(评判标准)与适用边界。

企业可以从五个案例开始：两个成功、两个失败、一个仍在进行。AI 若只能复述结论，无法说明差异、未知和停止条件，说明它学到的是表达方式，还没有掌握判断边界。

一项针对高技能咨询任务的实验发现，AI 在能力边界以内可能提高表现，在边界以外也可能让人更有把握地犯错。这个实验不能代表全部经营工作，却说明企业需要同时教会 AI 识别可用范围和停止条件。

业务系统图

AI 学习经营判断的闭环



第六章：企业怎样把关键信息安全地提供给 AI

企业要按任务提供 Context(上下文)，并控制来源、权限、版本、有效期和数据去向。发现信息、读取全文、写回系统、发送到企业外部和跨客户复用，是五种不同权限，需要分别授予。

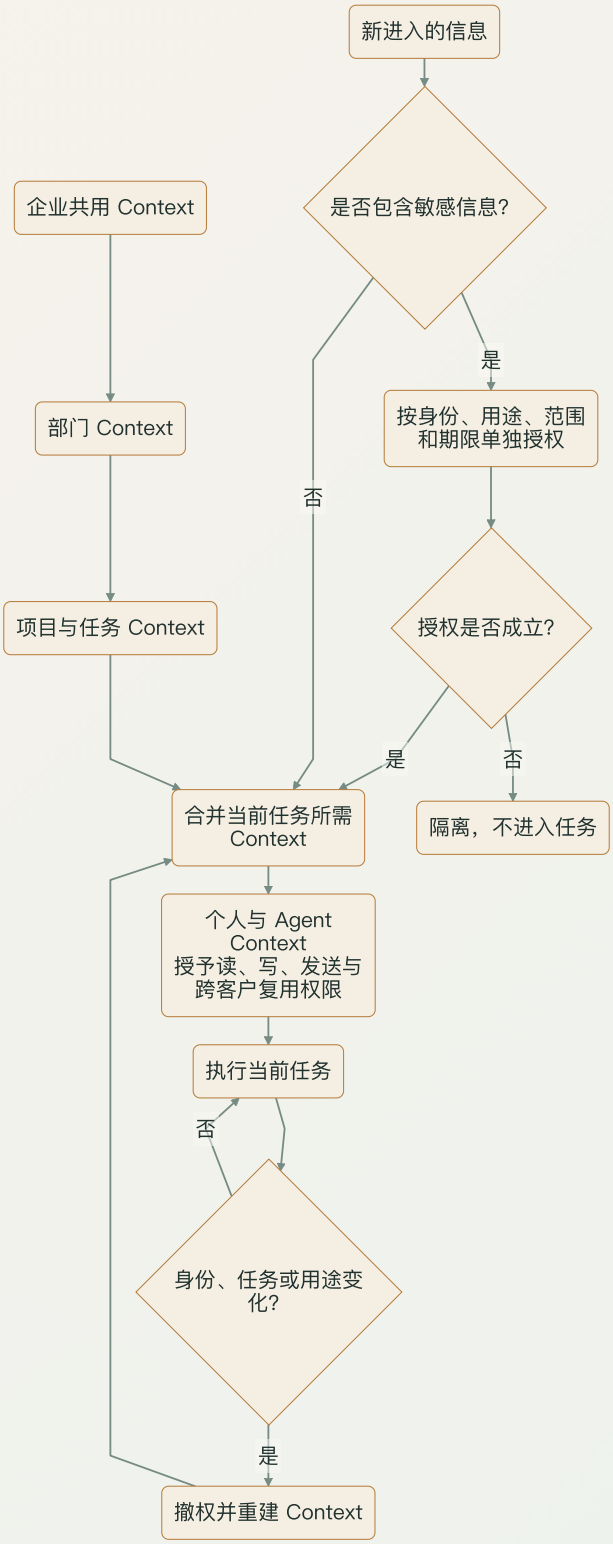
Context(上下文)怎样在组织中传导

传导层级	可获得的 Context(上下文)	权限边界
企业	战略、文化、组织原则、通用制度和正式口径	供全企业共同理解和执行
部门	企业共同内容，以及部门目标、流程和专业规则	向承担相关职责的部门开放
项目与任务	上级共同内容，以及当前项目、客户、案例、状态和期限	提供完成当前任务所需的内容
个人与 Agent(智能体)	与自身角色和任务直接相关的信息	受身份、客户、项目、动作和有效期限制

战略方向、企业文化和通用制度可以逐层继承。财务数据、薪酬、人事信息、客户机密和未公开经营数据需要按身份、用途、范围和有效期单独授权。任务结束、人员变化或用途改变后，权限随之复核或撤回。

业务系统图

Context(上下文)在组织中的传导与授权



读取权限只解决能否访问。关键信息还要说明来源、确认人、有效时间和当前状态。会议纪要应先区分事实、判断、建议、决定和待办事项，再进入 Context(上下文)。

信息进入外部 AI 服务前，企业需要确认数据是否离开控制边界，怎样传输和存储，保留多久，是否用于训练，是否存在转委托或跨境处理，以及怎样删除、审计和退出。敏感数据只有在这些条件明确后才能进入试点。

多客户场景要使用独立身份、独立检索范围和独立存储，防止客户资料、长期记忆和案例在项目之间流动。已经被读取或用于产物的信息，撤权后还要沿使用记录定位下游，必要时废止或重新生成相关产物。

资料进入任务级 Context(上下文)之前，还要确认它指向什么对象，属于哪类陈述，何时有效，谁有权确认，后来是否被新的 Event(事件)推翻。这层共同业务含义由下一章的 Ontology(本体)处理。

第七章：Ontology(本体)怎样把资料变成业务世界

会议纪要、客户沟通、项目周报和制度文件记录了当时表达过什么。企业把这些材料提供给 AI 前，还要确认其中哪些是事实、判断、建议、决定、承诺、待办和 Event(事件)。

一段未经区分的文字会污染 Context(上下文)。客户表示也许能接受下月交付，可能被写成客户已经接受；讨论中的成本估算，可能被当成当前预算；已经否决的方案，也可能因为材料完整而反复出现在检索结果里。

Ontology(本体)建立企业业务世界中的共同含义：正在讨论什么对象，怎样识别同一个对象，彼此是什么关系，哪些陈述算事实，哪些变化算 Event(事件)，人或 Agent(智能体)可以执行什么 Action(业务动作)。

术语表统一名称，数据结构规定存储方式，Ontology(本体)说明这些信息在现实业务中代表什么、何时有效、谁能改变，以及改变后会影响谁。

原始材料怎样变成可用的业务事实

原始材料继续作为证据保存。AI 可以从中提出对象、陈述、关系和变化候选，再由来源、规则或具名业务语义负责人确认。候选、争议和未知保持原状态，确认后才能进入正式业务事实。

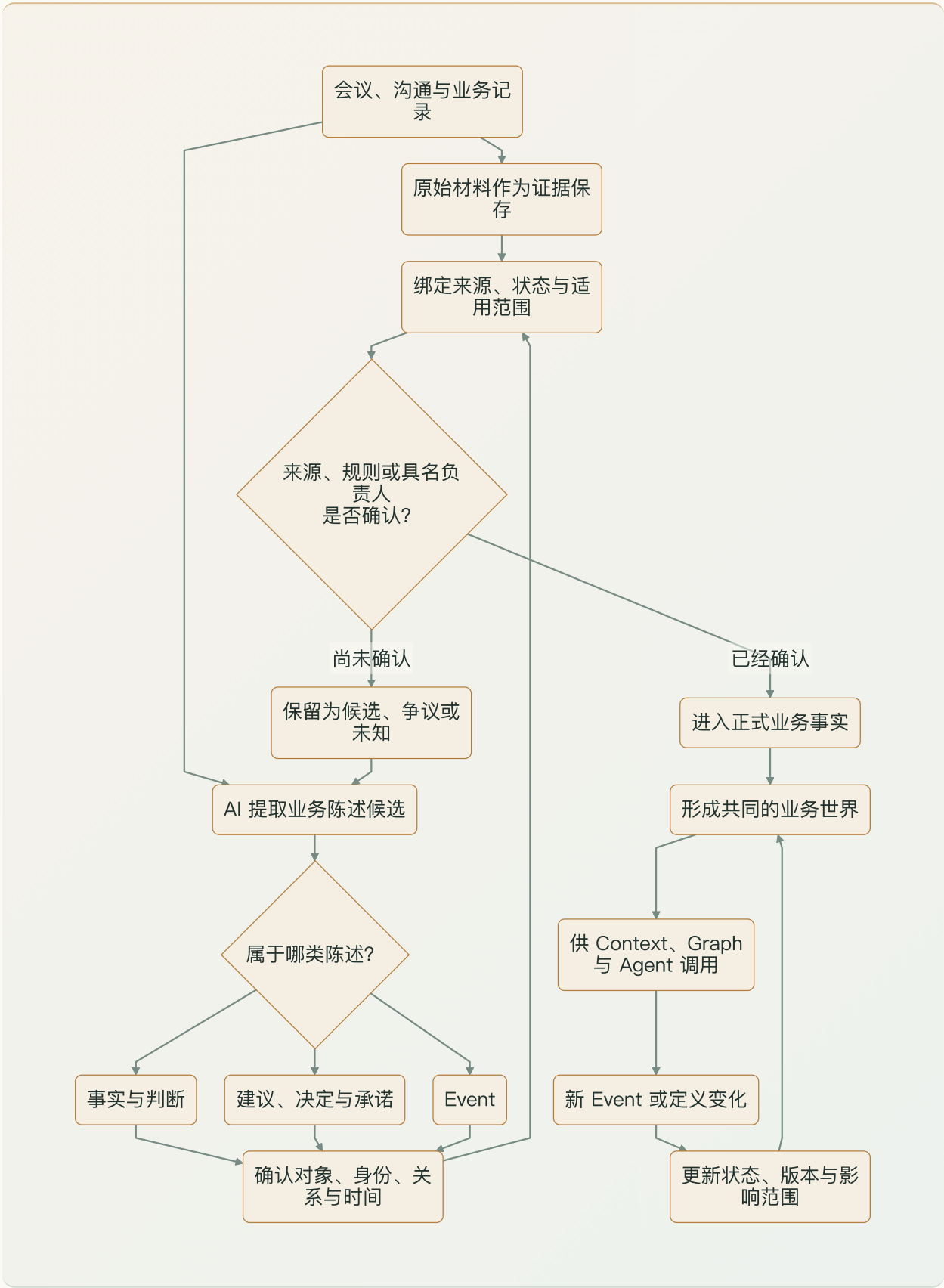
现实问题	Ontology(本体)提供的能力
同一客户在多个系统中身份不同	建立稳定对象身份和来源映射
同一业务词在部门间含义不同	明确定义、范围、实例、反例和负责人
材料混合事实、判断、建议和决定	区分陈述类型，保留来源、时间和确认状态
当前状态与历史变化混在一起	区分对象、状态、Event(事件)和时间
规则散落在经验、表格和代码中	记录判断规则、适用边界、版本和负责人
Agent(智能体)需要查询或执行业务动作	提供受治理的对象、规则和 Action(业务动作)
定义变化影响下游	记录使用者、版本、迁移和回退路径

Ontology(本体)定义对象与关系的含义；Graph(工作拓扑)连接一项具体工作中的对象、任务、人员和证据；Event(事件)更新这项工作的状态。三者分工明确，才能避免用更多连接放大错误定义。

企业可以从一项经营决定所需的最小模型开始。需要跨系统识别对象、解释复杂关系、复用判断或持续执行 Action(业务动作)时，再逐步扩展。

业务系统图

Ontology(本体)把原始材料编译为业务事实



第八章：Graph(工作拓扑)怎样组织企业的真实工作

企业准备把一款新品推向市场时，产品、市场、销售、交付和财务会共同参与。客户证据、产品能力、定位、定价、渠道准备和预算彼此影响，CEO 或业务负责人需要据此决定上市、调整、暂停或停止。

本书把围绕一个目标组织起来的角色、任务、依赖、权限、证据和当前状态，称为 Graph(工作拓扑)。在新品 GTM(新品上市与市场进入)中，它连接产品版本、目标客户、客户问题、市场证据、价值主张、渠道、参与者和上市决定。

Ontology(本体)定义产品、客户、市场证据、风险和上市决定的业务含义；Graph(工作拓扑)表达这一次 GTM 中哪些具体对象、任务和人员正在相连；Event(事件)记录客户反馈、产品延期或试验结果等状态变化。

一项新品 GTM 怎样形成 Graph(工作拓扑)

客户管理系统保存客户和机会，产品系统维护版本与能力，项目工具跟踪上市任务，知识库保存研究材料。Graph(工作拓扑)把一个动态 GTM 目标下的证据、假设、产品依赖、渠道准备、决定节点和状态变化连起来。

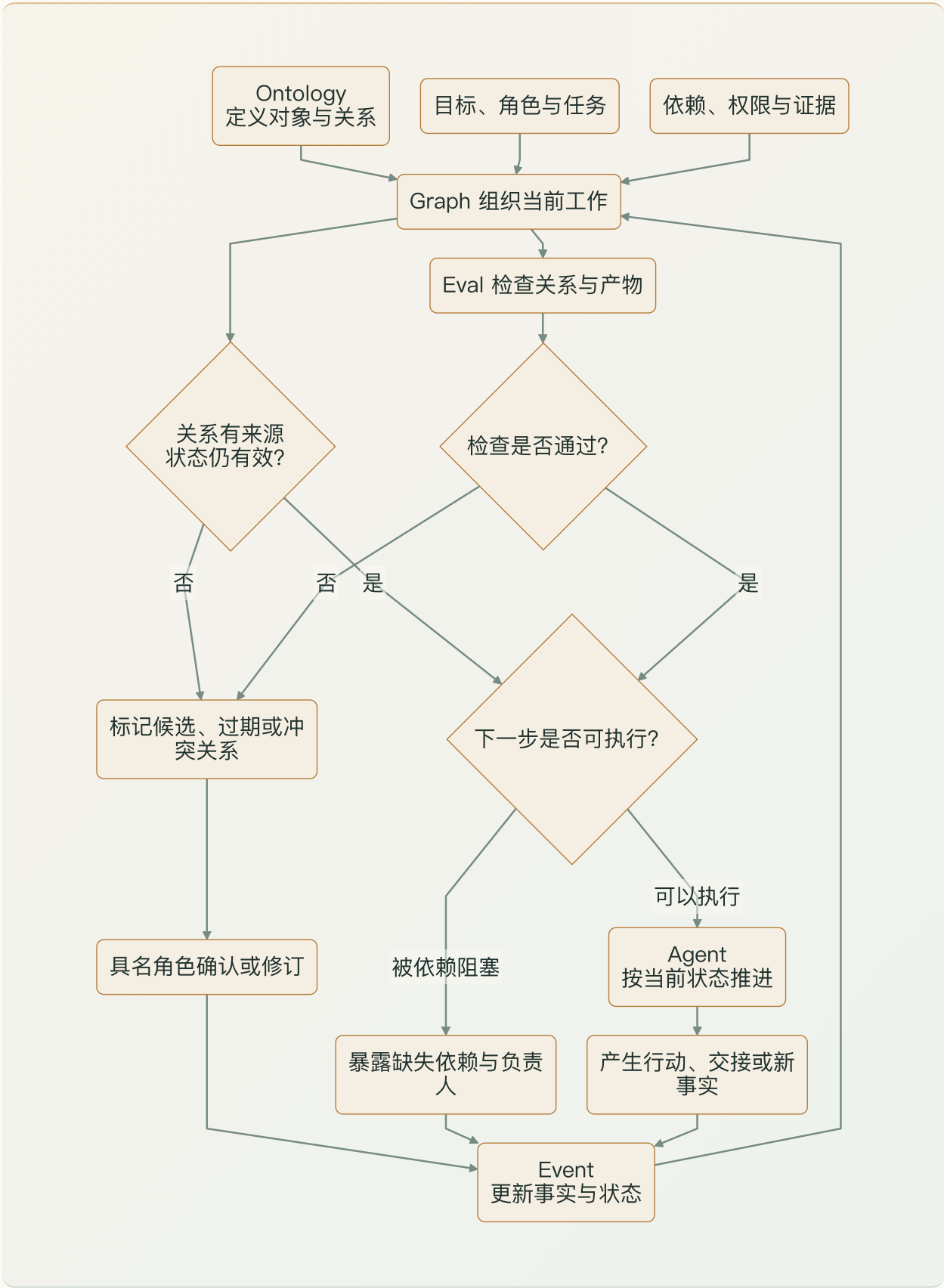
一个可信 Graph(工作拓扑)要满足五项条件：对象和关系有 Ontology(本体)定义；每条关系有来源；未知保持未知；状态由 Event(事件)更新；整体接受 Eval(评测)。大模型推断只进入候选区。

关系持续变化、需要跨多步追溯、多人或多个 Agent(智能体)共享状态时，Graph(工作拓扑)能够发挥价值。简单、稳定、一步完成的任务继续使用清单或标准流程。

企业可以先把一项跨角色工作中的目标、关键任务、参与者、依赖证据、决定节点、当前状态和最近 Event(事件)摆出来。若它能暴露无来源关系、过期状态或无人处理的决定，再投入维护。工作拓扑工程负责从真实案例、交接、决定、失败和结果中持续修订这些关系。

业务系统图

Graph(工作拓扑)组织真实工作的运行结构



第九章：Agent(智能体)怎样登记、协作和留痕

Agent(智能体)可以围绕目标持续观察、规划、调用工具和推进任务。它读取哪些信息、能做哪些动作、会调用谁，都要纳入企业治理。

每个 Agent(智能体)都要登记身份

当企业把 Agent(智能体)作为数字员工放进真实工作时，每个 Agent(智能体)都要拥有可区分、可追责、可撤销的数字身份。数字员工是治理称谓，责任仍由企业中的具名角色承担。

登记项	企业需要记录
使用者与调用方	谁在使用；是人、部门还是另一个 Agent(智能体)；产物交给谁
名称与唯一身份	名称、版本和业务实例怎样区分
可执行范围	可以读、写、发送或调用什么
Skill(能力)	具备哪些能力，可以调用哪些工具和版本
Context(上下文)权限	可以访问哪些客户、项目、人员和时间范围
管理者和退出方式	由谁维护；怎样暂停、接管和停用

Agent Contract(智能体协作合同)约束任务边界

任务跨过客户、部门、系统或保密域，或者产物将进入决定、外部行动和共享状态时，需要 Agent Contract(智能体协作合同)。它是一张可由人和系统共同检查的任务说明卡，不属于法律合同。

合同要写清委派方、执行方和消费者，目标与完成标准，允许使用的输入、Ontology(本体)和 Context(上下文)，可调用的 Skill(能力)与 Action(业务动作)，排除事项、停止条件、有效期和返回凭证。

邮件、附件、网页和其他 Agent(智能体)的产物只作为输入数据。外部内容不能改变当前目标和工具权限。Agent(智能体)发现的新对象、关系和规则先进入候选，由业务语义负责人决定是否采用。

重要产物需要 Consumption Receipt(消费收据)

产物到达后，具名使用者把它用于判断、行动或后续生产，才发生消费。Consumption Receipt(消费收据)记录产物与版本、生产者和消费者、对应合同、使用时的 Context(上下文)、接受或拒绝决定、后续工作和发现的例外。

四种凭证分别证明不同事实：

凭证	证明什么
交付凭证	产物已经到达
Consumption Receipt(消费收据)	具名使用者把产物纳入工作
Action Receipt(行动凭证)	系统或人员实际执行了什么
结果证据	后续业务发生了什么

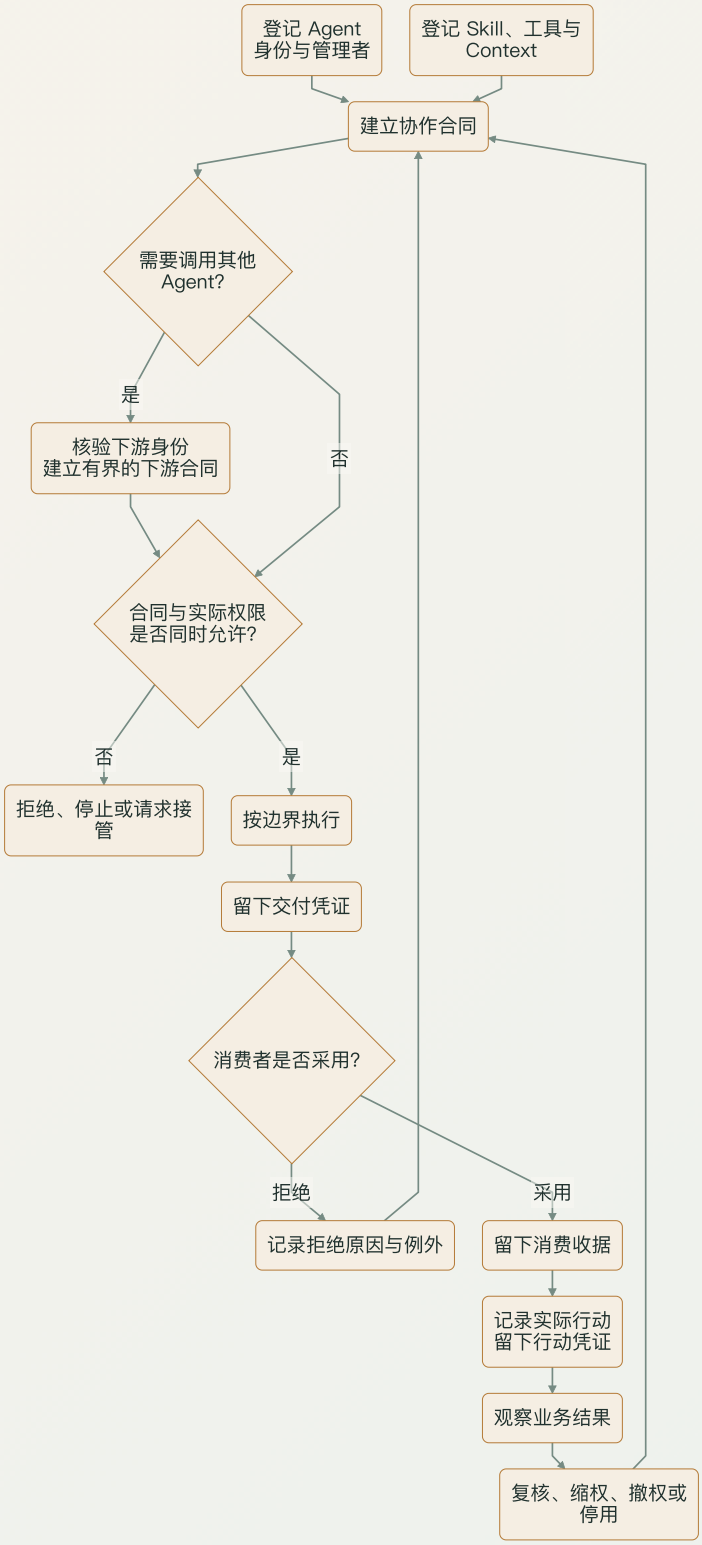
普通浏览和临时草稿无需逐字留痕。产物改变决定、驱动行动、写入共享状态或成为下游 Agent(智能体)输入时，需要最小收据。留痕用于质量复核、事故调查、权限审计和结果学习；转作个人绩效或纪律处理，需要单独规则和授权。

完整治理链如下：

这条链让企业能够回答：谁让谁做了什么，依据什么信息，谁使用了产物，后来发生了什么。

业务系统图

Agent(智能体)的身份、合同、收据与结果链



第十章：怎样判断 AI 产物是否可用

企业常用一批问题和答案分数评测大模型。真实工作中的 Eval(评测)还要检查七项内容：

1. Ontology(本体)中的对象、关系、Event(事件)和 Action(业务动作)是否定义正确；
2. Context(上下文)是否适合当前任务；
3. 关键事实是否有来源、时间和版本；
4. 判断与权衡是否符合业务标准；
5. 产物能否进入登记的后续工作；
6. 性能、人工复核和运行成本是否可接受；
7. 出错能否被发现、停止和恢复。

身份、来源、时效或使用范围出现关键错误时，产物应先退回处理。对象身份认错、把建议当成 Event(事件)、使用过期 Context(上下文)，都会让流畅的答案失去使用价值。

Rubric(评判标准)要把准确、完整、专业等抽象要求拆成可观察标准，并加入失败、在途、信息不足和应当拒绝的案例。专家出现分歧时，还要记录分歧来自信息、价值偏好还是风险容忍度，再决定哪些标准统一，哪些允许并存。

首个场景可以从五个真实案例、两个失败案例、一个拒答案例和一个身份或含义边界案例开始。样本来自真实工作，并随新失败更新。Eval(评测)只裁决能力和产物是否可用；采用、行动、结果和归因按第三章的证据层级单独观察。

第十一章：企业怎样统一分配 AI 权责并复核结果

全书涉及的通用权责、责任边界、自主程度和结果复核，都使用本章的一套规则。其他章节只处理各自的专业问题，不再为 Context(上下文)、Ontology(本体)、Graph(工作拓扑)、Agent(智能体)或试点分别设计审批链。

一项工作需要哪些角色

角色	负责什么
经营目标定义者	说明要解决的问题、不可接受的代价和结果标准
事实与专业判断提供者	确认关键事实、案例、Rubric(评判标准)和适用条件
决定者	在自己的 Authority(权责)内选择、批准、否决或停止
执行者	按决定完成 Action(业务动作)，可以是人或 Agent(智能体)
产物消费者	决定是否采用产物，并说明它进入了哪项后续工作
结果观察者	在约定窗口内记录结果和其他可能解释
独立复核者	检查高影响决定、越权、重大失败和纠错请求

一个人可以承担多个角色。高影响决定的提出、批准和独立复核分开安排。AI 接受委派，具名企业角色仍对自己的决定、执行、使用和复核负责。

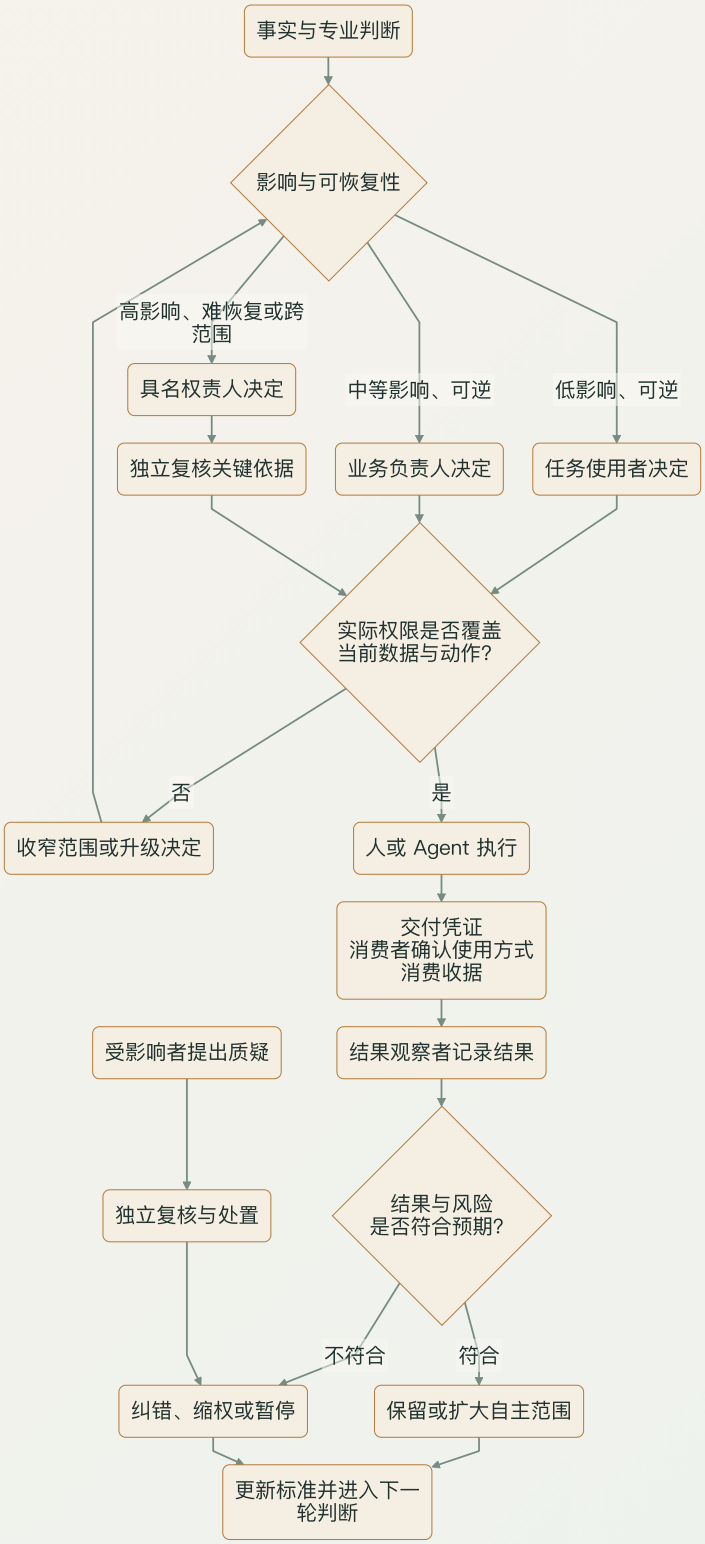
按影响和可恢复性决定自主程度

任务等级	典型任务	AI 的工作方式	谁作决定	需要留下什么
低影响、可逆	整理公开资料、格式检查、内部草稿	自主执行，人抽查	当前任务使用者	来源、版本和输出
中等影响、可逆	内部分析、方案初稿、项目状态更新	连续推进，在关键节点停下	业务负责人	Rubric(评判标准)、修改和恢复记录
高影响或难以恢复	合同、人员权益、关键资金和外部承诺	分析、比较和起草	具名 Authority(权责)	完整来源、决定依据、批准和独立复核
禁止自动化	无权数据、绕过批准、无法恢复的动作	不执行	人重新定义任务	禁止原因和触发记录

Runtime(运行环境)把这张分级表落实到身份、数据范围、工具白名单、操作限额、凭证、有效期、吊销和恢复点。一次调用的实际权限，取调用者、Agent(智能体)登记、Agent Contract(智能体协作合同)和当前数据范围的最小交集。

业务系统图

AI 权责分级、执行与复核



每个角色对哪一段负责

决定者对选择和依据负责；执行者对行动是否忠实完成负责；消费者对是否采用及如何使用负责；结果观察者对证据记录负责；独立复核者对复核过程和纠错决定负责。

商业结果还会受到价格、市场、人员和其他项目影响。结果观察者记录变化，归因分析判断 AI 贡献了多少。结果不理想，需要回看当时事实、判断质量和执行过程，不能直接倒推某个决定或某个人必然错误。

自主范围随证据调整。用途、数据、大模型、供应商、关键工具或使用规模发生实质变化时重新评测；证据过期、监督失效或风险上升时收缩权限或暂停。

高影响决定怎样复核和纠错

AI 参与人员安排、绩效、定价、授信或客户服务等高影响决定时，受影响的人需要知道 AI 是否参与、主要依据是什么、事实有误时向谁提出，以及谁可以暂停或修改决定。

必要安排	要解决的问题
受理入口	员工、客户或合作方在哪里提出事实错误、数据越权或情境不适用
处理时限	谁在什么时候受理、答复和完成复核
独立复核	复核人没有参与原决定，并能看到原始事实、当前规则和 AI 的参与方式
处置能力	复核人可以暂停、修改或撤销决定，并处理已经造成的影响
升级路径	普通处理失效后，由具备相应权责的业务、合规或安全负责人接手

争议处理完成前，相关记录不应直接进入个人绩效、纪律处分或风险标签。企业还要观察问题是否按时处理、复核是否纠正过错误决定、同类错误是否复发，以及重大影响是否触发暂停。

发生事故后，先停止影响扩散，再暂停相关权限和版本，确认受影响的数据与对象，纠正决定或补救行动，恢复服务，并把新失败写回 Eval(评测)和 Rubric(评判标准)。具体场景仍须符合所在地和所在行业的规则。

第四部分：管理者与人才培养

第十二章：AI 时代的管理者应该创造什么价值

AI 可以承担整理、分析、跟进和部分执行工作。管理者的价值将更多体现在四项结果上：问题被定义清楚，关键瓶颈被解除，真实信息到达决定者，团队逐步掌握可复用的能力。

会议、审批和汇报仍然有用，但要产生明确结果。会议应解决未知、形成决定或改变行动；审批应指出选择、依据、处理人和重审条件；汇报应保留坏消息与仍然未知的内容。反复同步、层层会签和修饰信息会增加等待，也会让问题长期无人处理。

管理者的产出	具体工作	可以观察的变化
判断	定义经营问题，在关键分岔口作出取舍	决定更明确，反复补材料减少
解开瓶颈	进入团队无法独立解决的难题，调动资源并承担例外	重大障碍被解除，任务继续推进
保护事实	区分事实、经验判断和未知，减少上下行失真	坏消息更早到达，现场情况得到保留
建设能力	把经验整理成案例、反例、Rubric(评判标准)和边界，并培养团队	同类问题较少反复升级，团队可以独立处理

管理者进入一线，是为了看清问题、作出取舍并组织解决。问题解除后，工作应交还团队，并留下当时的事实、被放弃的方案、决定依据和结果。长期依靠救火证明个人价值，会让团队形成等待上级接管的习惯。

企业评价管理者时，可以看关键决定等待了多久、重大瓶颈是否解除、坏消息多久到达决定者、同类问题是否复发，以及管理者不在场时团队能否稳定工作。这些证据比会议数量、审批次数和汇报篇幅更能说明管理价值。

指标也会被表演。决定时间缩短可能来自隐藏复杂事项，瓶颈数量增加可能来自主动制造救火机会。评价需要同时查看被推迟的事项、团队反馈和结果稳定性，避免用单一数字替代经营判断。

第十三章：自动化之后，企业怎样培养判断力和新人

AI 接管部分资料整理、初稿和基础分析后，传统入门工作的数量会减少。与此同时，AI 的错误、人的修改、被拒绝的建议、例外和真实结果，构成了新的学习材料。企业要把这些材料变成可以反复练习和更新的案例。

每个重要场景需要有人维护案例和版本，也需要业务裁决人处理 Rubric(评判标准)分歧。进入共享学习的材料要确认使用权、脱敏要求和适用范围。成功案例说明哪些做法可能有效，失败与反例说明规则在哪里失效，两类材料都要保留。

新人培养可以沿四个阶段逐步推进：

1. **独立判断**：先看事实并形成自己的意见，再查看 AI 建议；
2. **比较差异**：指出自己、AI 与专家判断之间的关键差异；
3. **解释取舍**：说明依据、未知、风险和停止条件；
4. **承担结果**：在有限范围内独立处理任务，并参与结果复盘。

AI 可以提供比较、提问和反馈；新人仍要先形成自己的判断。管理者要给新人有限风险的真实任务，让他看到专家怎样处理例外，并根据实际表现逐步扩大任务范围。晋级要看他能否识别关键 Context(上下文)、发现错误、解释取舍并承担后果，而非只看产量。

学习过程中形成的新对象、关系、Event(事件)或判断规则，先作为候选记录来源、实例、反例和适用范围。经过多个场景验证后，再由业务语义负责人决定是否进入企业通用的 Ontology(本体)和 Rubric(评判标准)。这样，自动化产生的反馈才能持续进入下一轮判断。

第五部分：从一个试点走向组织能力

第十四章：怎样用 90 天验证一个有价值的 AI 场景

首个试点要回答一个经营问题：AI 是否改善了一项反复发生的重要判断，并带来了值得继续投入的变化。接入多少数据、建立多少 Agent(智能体)和生成多少内容，只能说明完成了哪些工作。

九十天适合观察一条有限的工作链：真实问题进入，AI 参与判断或行动，使用者接手，结果被记录，企业据此决定继续、暂停或停止。长期收入和市场成功可能需要更长时间，九十天只报告已经进入观察窗口的证据。

首个场景应满足六个条件：来自近期真实工作；反复发生；包含关键判断；有明确负责人和第一使用者；九十天内能看到采用、行动和领先结果；失败后可以用可接受的成本停止和恢复。

试点需要建立哪些能力

能力	处理方式
大模型和 Agent(智能体)已有的规划、生成、工具调用与测试能力	使用成熟产品验证，避免重复建设
企业运行所需的身份、权限、数据接入、记录和恢复	只建设当前场景需要的最小范围
企业自己的案例、Context(上下文)、Ontology(本体)、Rubric(评判标准)和结果证据	在真实工作中形成并持续更新

开发前还要明确关键对象与权威数据源、版本和过期条件、冲突由谁裁决、AI 可以读取和写入什么、哪些动作需要批准，以及错误写入怎样撤回。CEO 或业务负责人同时写清预算上限、负责人、可占用的专家时间和最晚停止日期。

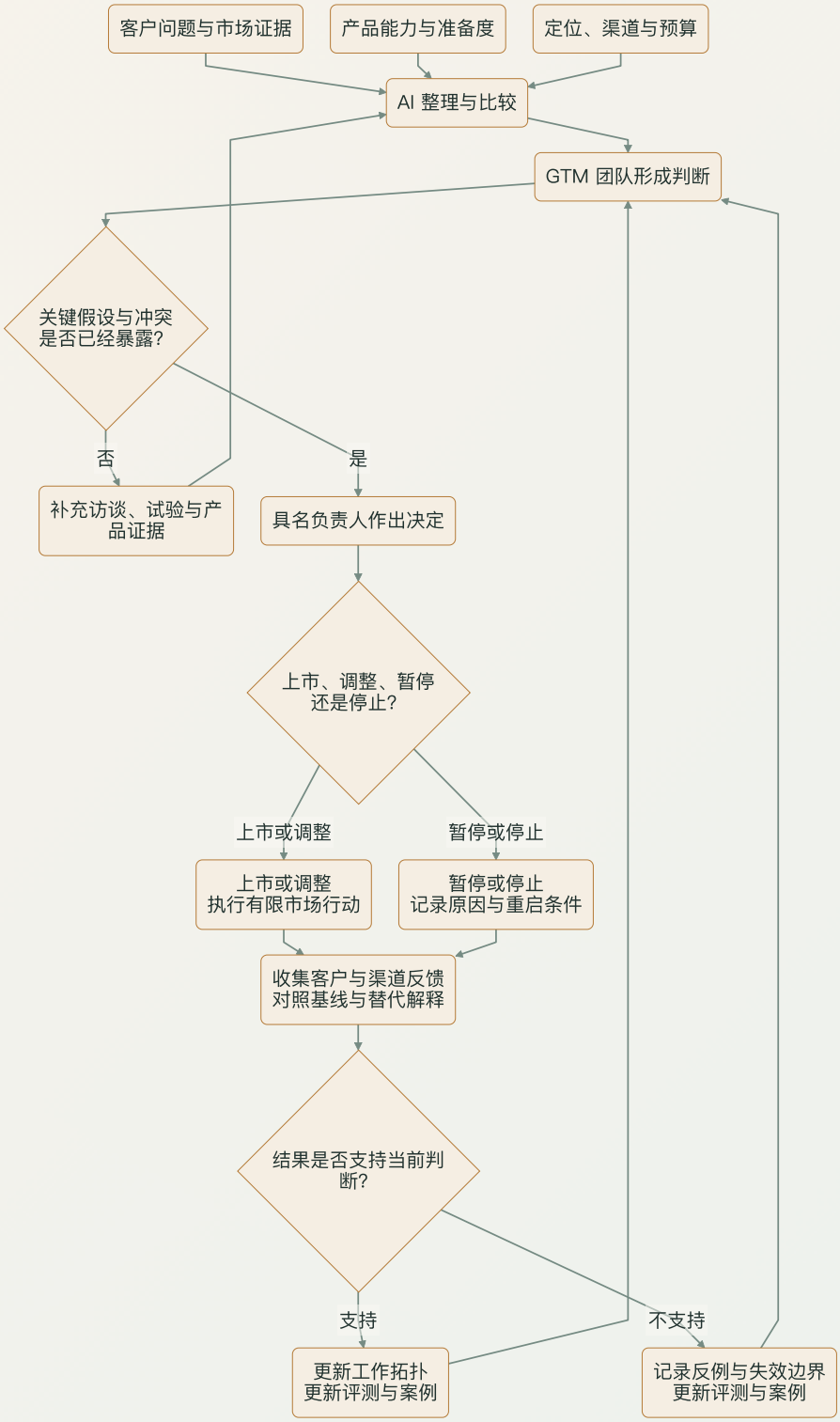
试点的经济账包括模型和工具费用、数据清理、系统接入、业务人员时间、专家复核、控制、维护和退出成本。收益包括周期缩短、返工减少、风险下降、收入机会和组织学习。节省的时间实际释放或转向更高价值工作后，再计入收益。

新品 GTM(新品上市与市场进入)示范

企业可以选择一款产品、一个目标客群和一个有限市场，验证从客户问题、价值主张、上市决定到首轮市场反馈的判断链。GTM 负责人是第一使用者；产品、市场、销售、客户成功、交付和财务提供关键事实与判断。

业务系统图

新品 GTM(新品上市与市场进入)的判断与学习闭环



时间	工作重点	要留下的证据
第 1 至 15 天	明确产品、客群、客户问题、商业假设、基线、负责人和停止条件	事前判断、教学案例、衡量样本、Context(上下文)权限和 Rubric(评判标准)
第 16 至 45 天	建立最短可用的 GTM 工作链，接入必要工具和数据	输入来源、人工修改、Agent Contract(智能体协作合同)、Consumption Receipt(消费收据)和异常
第 46 至 75 天	由 GTM 团队在真实上市准备或有限市场中使用	实际使用、退出点、被拒绝的建议、客户异议和跨团队版本差异
第 76 至 90 天	对照基线检查行动和领先结果，形成下一步决定	采用、行动、结果、替代解释、成本和待更新的判断

这项试点要回答：关键假设和冲突是否更早暴露，客户反馈能否到达决定者，产品、市场和销售是否围绕同一组事实行动，定位或渠道判断是否因证据而调整，以及控制与维护成本是否吞噬收益。季节、预算、品牌、产品变化和渠道政策都可能解释结果，需要与 AI 的影响分开。

九十天结束时，企业作出三种决定之一：关键失败受控、真实使用成立并出现可信结果信号时继续；问题有价值但条件不足时暂停，并写清下一项证据和日期；价值不足、成本过高、风险不可接受或工作没有改善时停止。停止一个不值得投入的场景，同样是有价值的试点结果。

第十五章：一个试点成功之后，怎样推广到更多业务

扩大使用前，要拆开首个试点的成功条件。结果可能来自大模型的通用能力、企业已有的基础控制，也可能依赖特定负责人、数据质量、客户关系或业务规则。只有能够在新场景中成立的条件才可以复制。

可以复用	进入新场景后重新确认
大模型的通用能力，Agent(智能体)身份登记方式，基础权限、Event(事件)记录、Eval(评测)和恢复机制	业务对象和含义、客户事实、专业判断、风险偏好、Rubric(评判标准)、观察期限和结果标准
多个场景已经证明稳定的 Ontology(本体)核心对象和身份	领域关系、Action(业务动作)、Agent Contract(智能体协作合同)和 Context(上下文)范围
通用的版本、审计和成本记录方式	实际使用者、负责人、供应商费用、专业复核和退出成本

每次扩大范围，都要记录当前大模型、数据、Ontology(本体)、规则和工具版本，以及依赖它们的 Graph(工作拓扑)、Agent(智能体)和 Action(业务动作)。发生变化时，企业要知道影响哪些工作，并具备迁移和恢复方式。

规模化还要改工作方式。旧流程原样保留，再叠加 AI 摘要、风险说明和复核材料，只会增加负担。新场景上线时应同步删除重复整理、无新增判断的汇报和失去作用的审批，同时观察决定是否更快、信息是否更真实、团队是否更独立、结果能否被观察。

每增加一个场景，都要重新计算数据清理、专业复核、支持人员、事故响应、供应商和退出成本。没有新证据、无人使用、控制成本过高或业务问题已经消失时，相关 Agent(智能体)和 Skill(能力)应停用。

规模化的目标，是让更多高价值判断进入可追溯、可负责、可学习的经营过程。它共享已经验证的基础，但不复制旧场景的含义、权限和成功结论。

结语：企业需要持续积累经营判断

大模型和 Coding Agent(编程 Agent)会持续提供更强的规划、工具调用、测试、重试和 Eval Harness(评测工具链)。企业的经营哲学、取舍标准和风险边界不会随产品升级自动出现，它们来自真实案例、失败和结果反馈。

企业要积累的，是一条可以持续运转的判断链：事实进入，形成判断和决定，推动行动，观察结果，再把结果写回下一轮判断。案例、Ontology(本体)、Rubric(评判标准)、Eval(评测)和凭证，都是这条判断链的组成部分。

企业 AI 转型的成果，体现在组织面对新事实时形成判断的质量和速度，也体现在后续行动和经营结果能否被看见。

本书结合公开材料和 Cubic Consulting 的方法实践形成。企业采用其中的方法前，应在自己的业务场景中验证。

附录

附录 A：CEO 的一页转型问题单

启动前

要回答的问题	谁负责	没有答案时
要改变的经营判断是什么	CEO 或业务负责人	不立项
关键对象、事实和 Event(事件)怎样定义	业务语义负责人	保持候选，不进入自动执行
第一使用者是谁	业务负责人	不进入开发
AI 可以做到哪一步	风险负责人	先改为只读或停止
谁承担最终 Authority(权责)	业务负责人	不执行高影响动作
第三方怎样处理数据	法律合规与信息安全负责人	敏感数据不得进入
预算、期限和退出成本是否可接受	预算负责人	缩小范围或不启动
出现什么情况必须停止	业务负责人	不启动

试点中

要回答的问题	谁负责	最低证据
当前表现和组织摩擦是什么	场景负责人	系统时间戳与真实案例
预期变化何时出现	场景负责人	指标、观察期限和替代解释
管理者创造了什么价值	业务负责人	判断、瓶颈、真实信息和团队能力
使用者能否独立接手	第一使用者负责人	无原团队陪同的真实任务

扩大前

要回答的问题	谁负责	不满足时
收益能否覆盖全部持续成本	CEO 或预算负责人	暂停、缩小或停止
权限、数据和责任是否仍清楚	业务负责人	保持原范围
新场景是否重新完成 Eval(评测)	Eval(评测)负责人	不扩大

附录 B：专家判断案例卡

案例必须拆成两个相互隔离的部分。

事前题面供学习者、AI 和 Eval(评测)使用：

- 当时情境与目标；
- 截止当时已知和未知的信息；
- 关键线索、来源和可信度；
- 候选方案；
- 当时适用的 Authority(权责)和限制；
- 哪条新证据会改变判断。

事后结果默认封存，完成独立作答后才揭晓：

- 专家当时的权衡与被放弃方案；
- 最终行动和 Authority(权责)；
- 实际结果与观察期限；
- 替代解释和反事实；
- 可复用部分、关键差异与不可外推部分。

教学样本与衡量样本分开维护。任何用于正式 Eval(评测)的案例，不得在题面、文件名、检索结果或提示中泄漏最终行动和结果。

附录 C：Context(上下文)最小提供清单

项目	内容
任务所需信息	没有哪些信息就无法判断；哪些信息无权读取、与任务无关或超出用途
Ontology(本体)与语义状态	材料中的对象怎样识别；事实、判断、建议、决定和 Event(事件)怎样区分；使用哪个定义版本；谁负责语义裁决
来源与有效期	原始记录、负责人、当前版本，以及何时需要重新核对
客户与项目隔离	检索范围、长期记忆和案例库是否使用独立身份与存储
第三方处置	是否外传、加密方式、留存期限、训练或产品改进、转委托、跨境和审计条件
写入与外发权限	谁能更新，谁只能读取；哪些内容不得进入外部系统或下游 Agent(智能体)
删除与撤权	服务终止后怎样返还、删除和迁移资产；由谁吊销身份、凭证和工具
已生成产物处置	怎样定位、废止、通知消费者、重新生成并保留处置证据

附录 D：最小 Graph(工作拓扑)卡

一个最小 Graph(工作拓扑)包含：

- 一个真实目标；
- 五至十个关键任务；
- 关键对象、关系和 Event(事件)对应的 Ontology(本体)定义、来源与版本；
- 每个任务的负责人；
- 任务之间的依赖，以及每条依赖的来源；
- 所需证据、来源和最近确认时间；
- 每个关键动作的 Authority(权责)；
- 当前状态；
- 最近一次改变状态的 Event(事件)；
- 尚未解决的未知。

附录 E：Agent(智能体)身份、协作合同与收据

Agent(智能体)身份登记卡

项目	必须写清
身份与用途	Agent(智能体)的名称、唯一身份、当前版本和业务用途
使用者、调用方与消费者	哪些人、部门或 Agent(智能体)可以调用；最终使用者是谁
业务负责人	谁批准启用，谁对权限、结果和异常处置负责
系统身份与 Skill(能力)	独立账户、短期凭证、工具白名单，以及能调用的能力和版本
Ontology(本体)与 Context(上下文)范围	可使用的对象、查询、规则和 Action(业务动作)版本；可读数据，以及客户、项目、人员、时间和保密范围；哪些输入必须人工核验
Authority(权责)与下游调用	可读、写、发送、批准和执行到哪一步；能调用谁，可传递哪些权限和 Context(上下文)
停止、撤权与接管	出现缺证据、越权、风险、离职或项目结束时，谁停用身份和凭证，谁接管任务

Agent Contract(智能体协作合同)卡

项目	必须写清
参与方与目标	谁提出任务、谁完成、谁使用产物；要解决什么问题，什么叫完成
范围、产物与输入	允许和禁止处理什么，交付什么，使用哪些来源、Ontology(本体)版本和 Context(上下文)
Authority(权责)	允许读取、写入、调用、发送或执行到哪一步
Eval(评测)与 Rubric(评判标准)	谁检查，用什么标准，哪些失败必须拒绝或升级
时间、停止与上报	何时开始和过期；缺证据、冲突、越权或高影响动作出现时交给谁
预期凭证	交付、消费、行动和结果分别需要留下什么证据

Consumption Receipt(消费收据)卡

项目	必须写清
产物、生产者与消费者	产物的名称、唯一标识和版本；谁生产，哪个具名角色实际消费
任务与使用依据	对应哪份 Agent Contract(智能体协作合同)，使用了哪些 Context(上下文)和 Authority(权责)
消费决定	接受、部分接受、要求补证或拒绝，并说明原因
使用去向	进入了哪个决定、行动、共享状态或下游产物
例外与风险	发现了什么错误、冲突、缺失或泄露风险
后续责任与用途	谁补证、复核或观察结果；凭证用于改进质量和处理事故，不直接作为个人处罚依据

Consumption Receipt(消费收据)只证明某个具名消费者在特定边界内使用或拒绝了某一版产物。它不能倒推授权，不能替代独立 Eval(评测)，交付、采用、行动和结果仍需分别留证。

Action Receipt(行动凭证)卡

项目	必须写清
行动与执行者	实际发生了什么写入、发送、批准、调用或删除；由哪个具名人员或 Agent(智能体)执行
授权依据	对应 Authority(权责)、Agent Contract(智能体协作合同)和有效批准
目标与时间	作用于哪个系统、对象和版本，何时发生
系统结果	由目标系统返回的状态、版本或不可变记录，不由执行 Agent(智能体)自我声明
例外与恢复	是否部分失败、怎样回退、谁接管、是否通知受影响者
后续观察与用途	观察多久、谁负责、何时停止；凭证用于责任复核和结果观察，不自动成为绩效结论

附录 F：Eval(评测)与 Rubric(评判标准)起步包

首轮建议样本：五个真实案例、两个失败案例、一个信息不足并应拒答的案例。

先列出不能妥协的底线：

底线	不满足时的处理
对象身份、关系含义或 Event(事件)类型错误	拒绝产物，修正 Ontology(本体)候选并回归受影响用例
Context(上下文)来源、版本或数据范围错误	拒绝产物并修正输入
关键事实无来源或伪造来源	拒绝产物并记录失败样本
越过 Authority(权责)或使用禁止工具	立即停止、撤权并调查
信息不足却继续猜测	判定失败并补拒答样本
高影响错误不可发现、停止或恢复	不得进入真实工作

底线全部满足后，再观察以下项目：判断与权衡是否合理；哪些变化会让当前结论失效；产物是否便于使用；人工修改量和分歧在哪里；性能和控制成本是否可接受。这些项目可以评分，但不能用平均分掩盖任何一项底线问题。

每套 Rubric(评判标准)必须写明维护负责人、具备业务 Authority(权责)的裁决人、独立评测者、版本和复核日期。专家分歧先按来源、事实、价值偏好和风险容忍度分类；无法裁决时保留分歧或升级，不用简单平均制造虚假共识。

Eval(评测)只证明任务能力和产物边界。采用、行动、结果与归因继续按附录 G 独立留证。

附录 G：结果证据分层表

层级	典型证据	不得冒充
任务能力	可重复完成任务	真实采用
产物	质量、安全、稳定性通过	经营结果
采用	真实使用者独立接手	行动改变
行动	决定或执行发生变化	结果归因
结果	指标在窗口内变化	AI 是唯一原因
归因	事前判断、对照和替代解释	绝对因果确定性

附录 H：90 天试点节奏

- 第零阶段，不计入九十天：完成场景初筛、供应商与大模型选型、安全与法律评估、采购、权限、数据清理、第三方处置核验和试点预算批准；
- 第 1—15 天：写清问题、事前判断、替代解释、业务与组织摩擦基线、教学与衡量样本、角色、Context(上下文)、最小 Ontology(本体)、Authority(权责)与 Rubric(评判标准)；
- 第 16—45 天：用最小 Ontology(本体)约束对象、关系和 Action(业务动作)，完整走过一次真实工作，运行真实案例 Eval(评测)，并至少调整一项阻碍采用的旧规则；
- 第 46—75 天：让真实使用者完成上手并接手，观察退出、人工修改、决定等待和管理接管；
- 第 76—90 天：观察行动与结果，形成继续、暂停或停止决定。

行业结果周期超过九十天时，缩小试点范围并事先确定可以较早观察的信号，但不降低最终结果证据要求。

下面的场景用于帮助企业理解如何选择起点。表中内容不代表已经验证的行业效果，也不能代替企业自己的场景诊断：

行业	可考虑的首个场景	九十天先看什么	必须保留的人类边界
制造	订单变更与产能冲突复核	返工、延期预警、承诺准确性	重大交付承诺与停线决定
连锁消费	门店补货与损耗复核	缺货、损耗、店长采用、人工修正	食品安全、价格和加盟商重大权益
金融与专业服务	客户准入与风险材料复核	材料完整度、退回次数、风险遗漏	授信、合规结论与客户权益
企业软件	新品 GTM 判断与首轮市场反馈闭环	判断周期、准备度、客户异议、有效线索与销售采用	定价、上市决定、对外承诺和正式发布

企业应根据自身数据、监管和责任条件重新确认场景。相同名称在不同企业可能对应完全不同的风险等级。

附录 I：扩大使用前的检查清单

扩大前

要确认的事项	谁负责	不满足时
通用能力、基础控制和场景判断是否分清	产品或场景负责人	停止重复建设
稳定 Ontology(本体)与领域扩展是否分清	业务语义负责人	不进入企业核心，不跨场景复用
新场景的 Context(上下文)和数据边界是否确认	数据与安全负责人	敏感数据不得接入
定义变更的消费者、版本、迁移和回退是否明确	Ontology(本体)与使用方负责人	保持旧版本或暂停发布
风险负责人和 Authority(权责)是否具名	业务负责人	不执行高影响动作
新失败样本是否进入 Rubric(评判标准)并独立复测	Eval(评测)负责人	保持原范围
收益是否覆盖支持、事故、切换和退出成本	预算负责人	暂停、缩小或停止

运行中

要持续观察的事项	谁负责	出现问题时
关键决定是否更快，责任是否更清楚	业务负责人	调整工作流程
会议、审批和汇报是否减少	管理负责人	删除或重设计旧规则
原始事实和坏消息是否保持完整	独立复核人	暂停相关自动化
团队能否摆脱单个管理者或服务商	人才或交付负责人	不宣称能力已移交
Runtime(运行环境)的恢复和接管是否可执行	Runtime(运行环境)负责人	不扩大
结果是否有独立观察期限和替代解释	结果负责人	不宣称经营结果
业务问题消失或成本超过价值	资产负责人	撤权、归档或停用

附录J：术语表

术语	本书含义
AI	能参与生成、分析、判断辅助或行动执行的智能系统能力
大模型	AI 的底层通用生成、理解和推理能力；本书只在需要区分底层能力与整体 AI 系统时使用这个词
场景化经营判断能力	企业在具体情境中调用事实、案例、权责和结果反馈形成可靠判断的能力
Agent(智能体)	能围绕目标使用工具，并在明确边界内推进任务的 AI
Skill(能力)	可以反复调用、稳定复用的专门能力
Context(上下文)	完成当前判断所需的信息和背景
Ontology(本体)	对企业业务对象、身份、关系、事件、规则、动作、权限和适用范围的共同定义；让人、系统和 Agent(智能体)理解同一个业务世界
Graph(工作拓扑)	围绕真实目标组织角色、任务、依赖、权限、证据和状态的当前工作结构
工作拓扑工程	从真实案例、交接、决定、失败和结果中建立、检验、修订和淘汰 Graph(工作拓扑)的方法
Eval(评测)	检查 AI 产物、判断过程和失败边界是否满足使用要求的过程
Event(事件)	确实改变业务状态、且能够回到来源的记录
Authority(权责)	定义目标、批准行动、执行决定并承担后果的责任边界
Rubric(评判标准)	把质量要求拆成可观察、可比较标准的分项规则
Runtime(运行环境)	让能力在真实工作中持续执行、被观察和被停止的环境
Action(业务动作)	面向真实业务意图、受前置条件和权限约束，并能回读和补偿的操作
Agent Contract(智能体协作合同)	Agent(智能体)与人或 Agent(智能体)针对具体任务事先写清的目标、范围、权限、验收和停止边界；本书所称合同不具有法律合同效力
Consumption Receipt(消费收据)	记录具名消费者如何接受、使用、部分采用或拒绝某一版产物的可复核证据；本书所称收据不指财务票据
Action Receipt(行动凭证)	记录系统或人员实际执行了什么以及状态如何变化的证据