

Agent-to-Agent (A2A) 研究报告

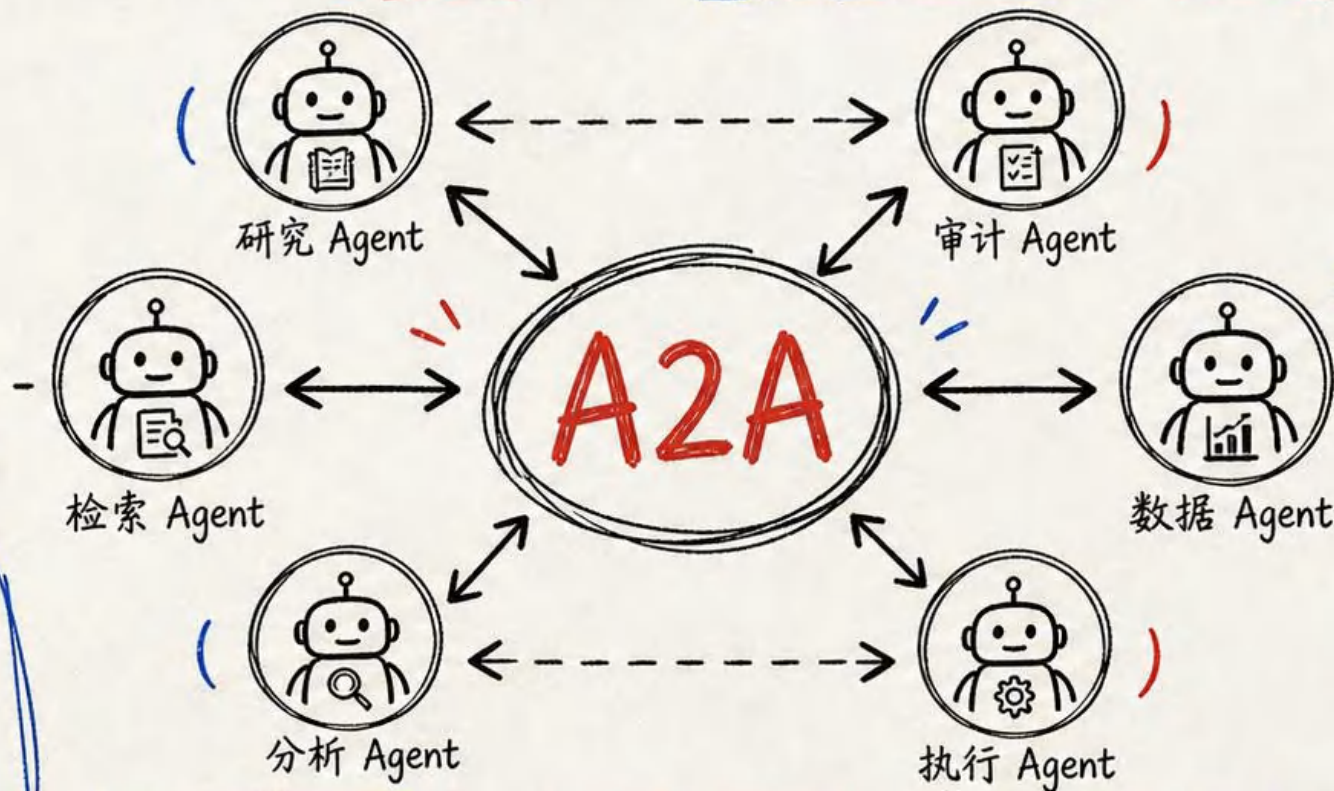
从智能体互操作，到数字委托协议与组织基础设施

研究主线

- ✓ 互操作
- ✓ 可委托
- ✓ 可验证
- ✓ 可追责
- ✓ 可编排

价值所在

- ✓ 连接智能体
- ✓ 协作完成任务
- ✓ 支撑组织运转



不是智能体通信，
而是可编排的
组织基础设施

从通信协议
升级为数字委托协议，
成为组织的“操作系统”

跨厂商

跨框架

跨系统



清新研究



2026年6月4日

更多免费行业报告

并购家

www.ipoipo.cn





@ 清新研究团队简介

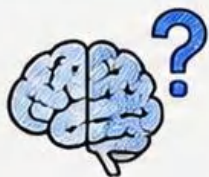
- 领导学术研究团队近30人，指导大数据、AI、人形机器人等多个产业团队。
- 团队坚持：整体主义 实证主义 社会建构 进步主义

沈阳：清华大学新闻学院/
人工智能学院双聘教授、
博导



六大研究方向

1. AI大模型
理论与哲学



2. AI文艺



3. AI应用



4. 新媒体与
网络舆论



5. 大数据



6. XR应用



视频号：
@清新研究



公众号：
@清新研究



邮箱：124739259@qq.com



微博：@清新研究



公众号：@清新研究



清新研究团队 | 2026年

更多免费行业报告

并购家

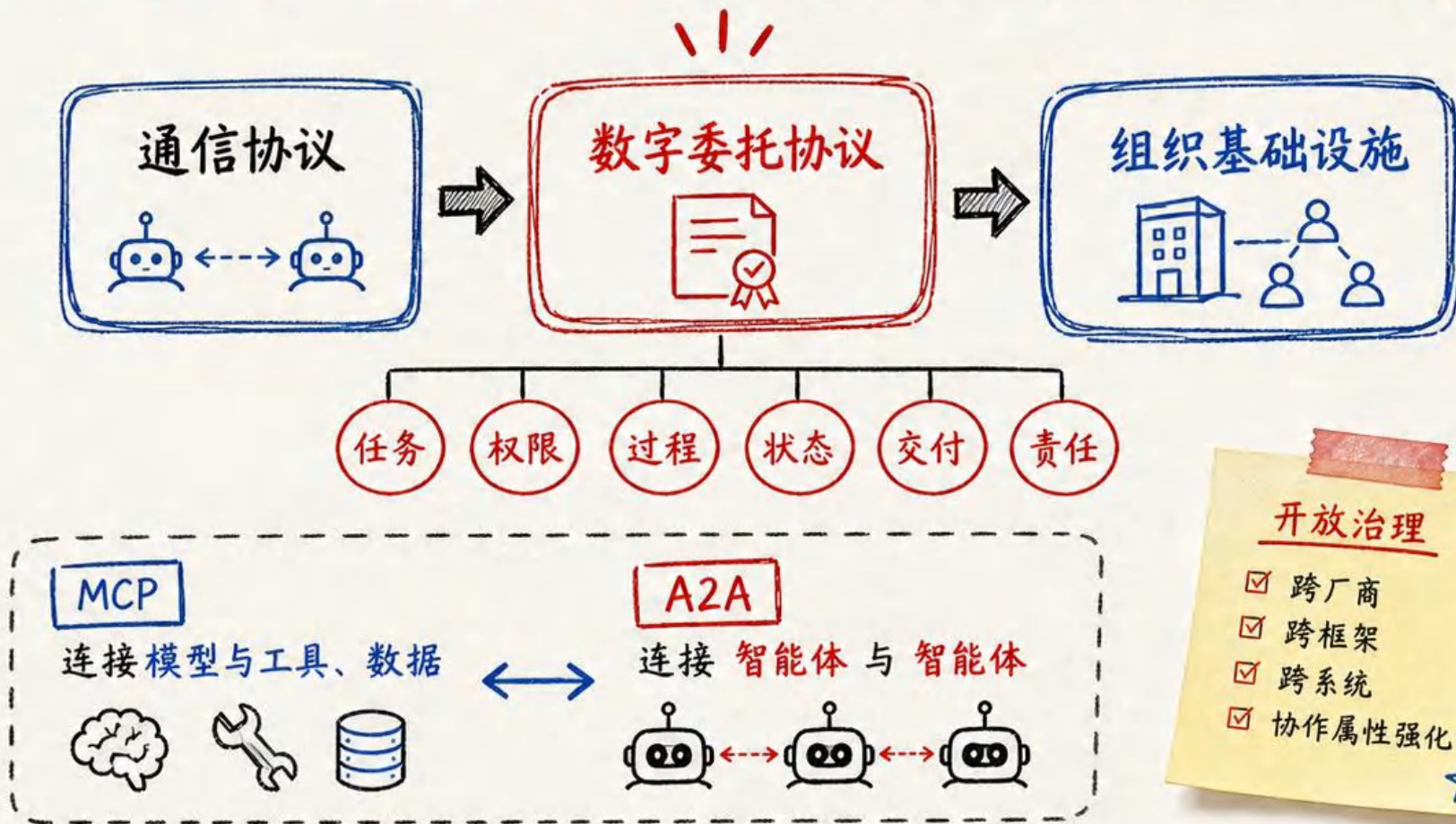
www.ipoipo.cn



A2A 不只是智能体之间的通信协议， 而是AI智能体时代的数字委托协议。



- A2A 不只是智能体之间的通信协议，而是AI智能体时代的**数字委托协议**。
- 它的价值不仅在于连接智能体，更在于让**任务、能力、权限、证据、审计和责任**成为可编排的**组织基础设施**。
- Google 将 A2A 定位为**开放协议**，用于让不同厂商、不同框架构建的 AI 智能体彼此通信、交换信息并协调行动。
- MCP 连接**模型与工具、数据**，A2A 连接**智能体与智能体**。
- A2A 进入 **Linux Foundation** 体系，开放治理强化跨厂商、跨系统协作属性。



开放治理

- ☒ 跨厂商
- ☒ 跨框架
- ☒ 跨系统
- ☒ 协作属性强化



资料来源: <https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/>;

<https://modelcontextprotocol.io/introduction>; <https://www.linuxfoundation.org/press/linux-foundation-launches-the-agent-to-agent/>

更多免费行业报告

并购家

www.ipoipo.cn

清新研究 | 2026年6月4日 |

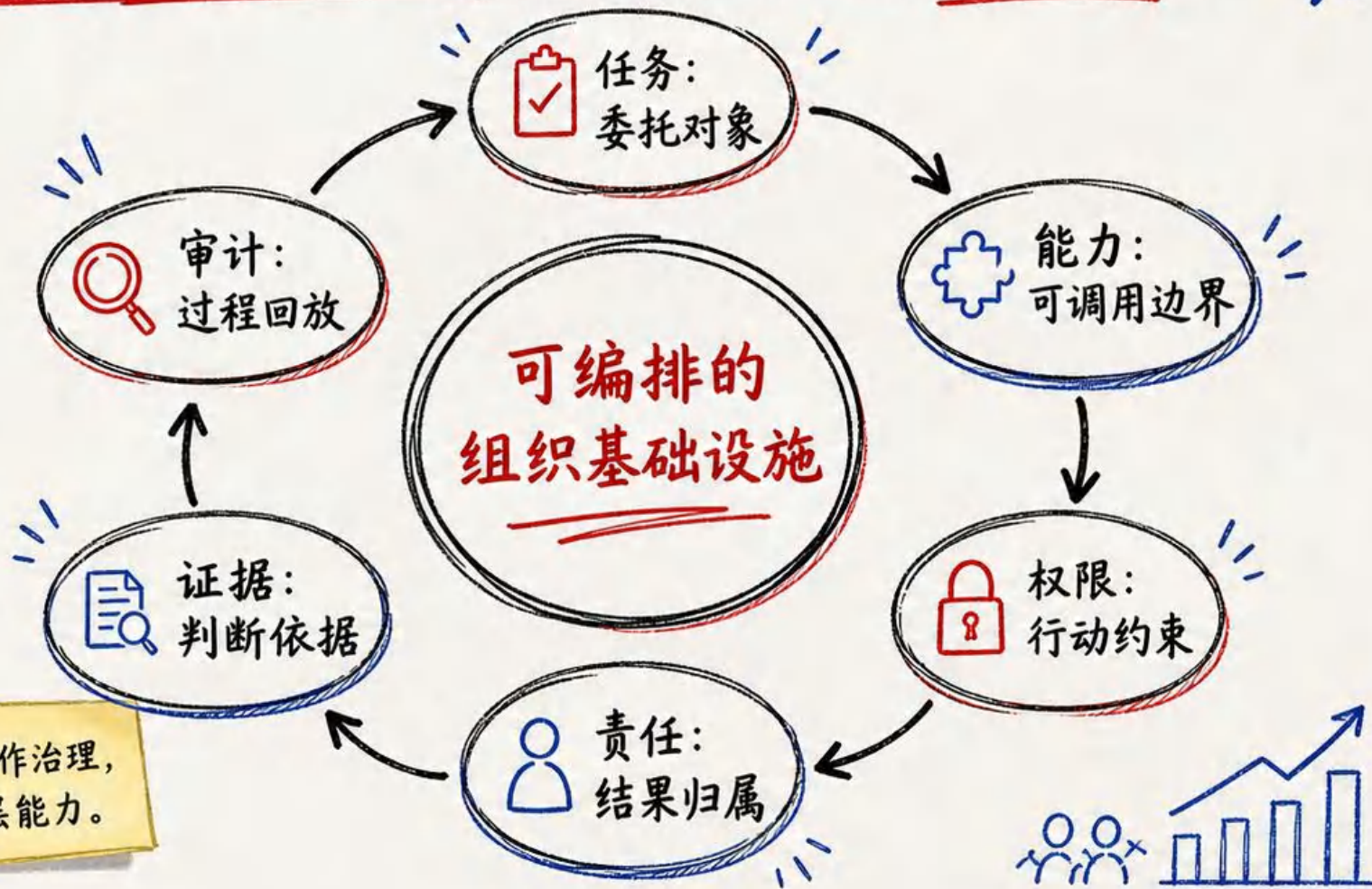
Agent-to-Agent (A2A) 研究报告

☆ A2A 的价值不仅在于连接智能体，更在于让任务、能力、权限、证据、审计和责任成为可编排的组织基础设施

✓ A2A 不只是智能体之间的通信协议，而是AI智能体时代的数字委托协议。

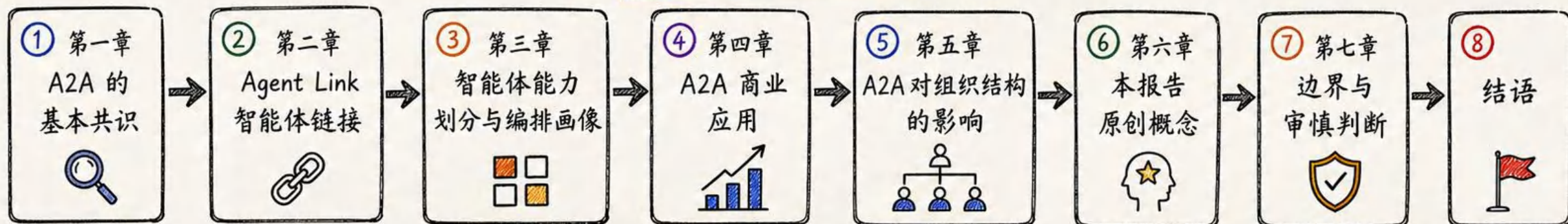
✓ 它的价值不仅在于连接智能体，更在于让任务、能力、权限、证据、审计和责任成为可编排的组织基础设施。

💡 从互操作，走向可信任的协作治理，构建组织智能体时代的底层能力。



本报告不把 A2A 仅仅作为一个技术协议来介绍，
而是将其放进 **更大的智能体基础设施框架中**

报告章节关系



删除独立的“核心创新解释”章节后，
A2A 的 **数字委托判断** 回到各章节中展开。

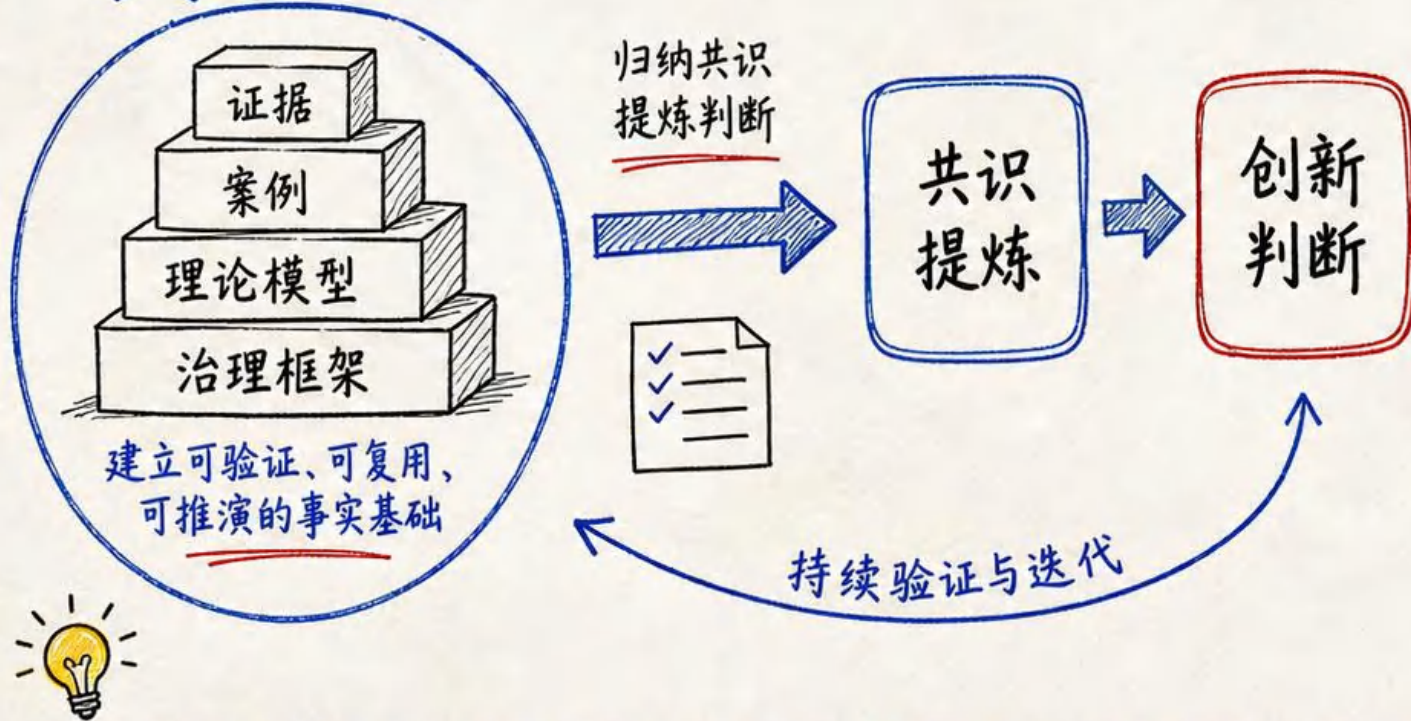
📖 阅读导航小贴士

- ☑ 按章节顺序阅读，构建完整认知路径
- ☑ 可按兴趣跳读，章节间相互引用支撑
- ☑ 建议结合术语表与附录，提升理解效率

先建立 A2A 的事实底座，再提出本文的创新判断

第一章 A2A 的基本共识

事实底座



① 多智能体协作会成为趋势



② A2A 解决的是智能体与智能体的互操作问题



③ MCP 和 A2A 是互补关系，不是替代关系



④ A2A 标准化的是任务协作过程，而不只是消息传递



⑤ A2A 正从厂商提案走向开放治理



⑥ A2A 会改变企业软件入口



⑦ A2A 的难点不只是技术，而是可信协作



⑧ A2A 是否是“AI 时代 HTTP 协议”尚无定论



多智能体协作会成为趋势

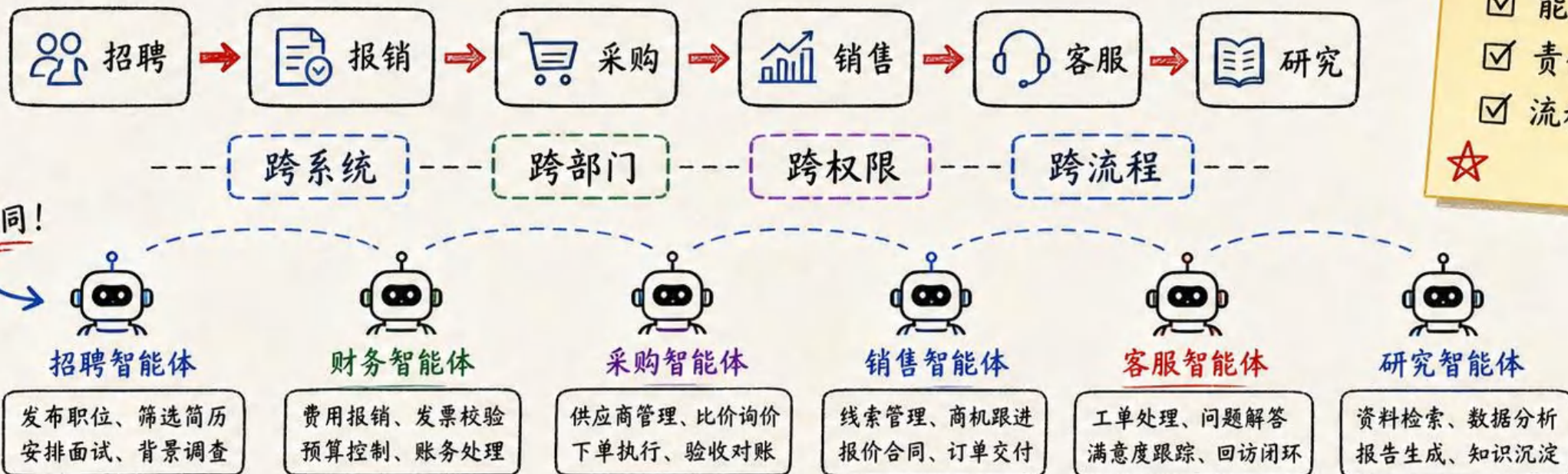
- ① 企业任务通常不是单点动作，而是跨系统、跨部门、跨权限、跨流程的复杂链条。
- ② 招聘、报销、采购、销售、客服、研究都需要多个专业能力共同完成。
- ③ 一个万能智能体在企业里会面临权限过大、专业过杂、系统分散和责任模糊的问题。

协作的价值

- ☑ 专业分工
- ☑ 能力复用
- ☑ 责任清晰
- ☑ 流程串联



多个专业
智能体协同!



一个**万能**智能体在企业里会面临权限过大、专业过杂、系统分散和责任模糊的问题

! 权限过大:

一个智能体不应同时拥有HR、财务、法务、IT的全部权限



! 专业过杂:

招聘、财务、法务、销售、客服等领域判断标准不同



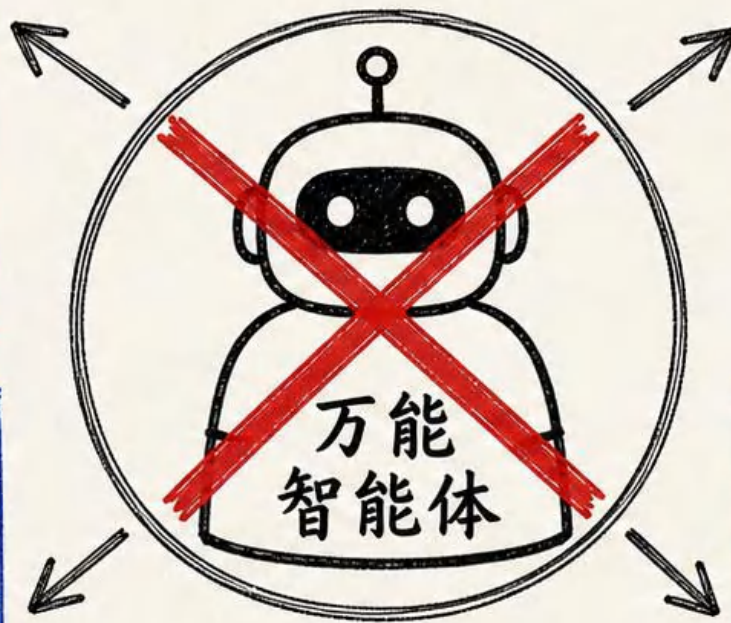
! 系统分散:

企业软件来自不同厂商，数据、流程、接口不统一



! 责任模糊:

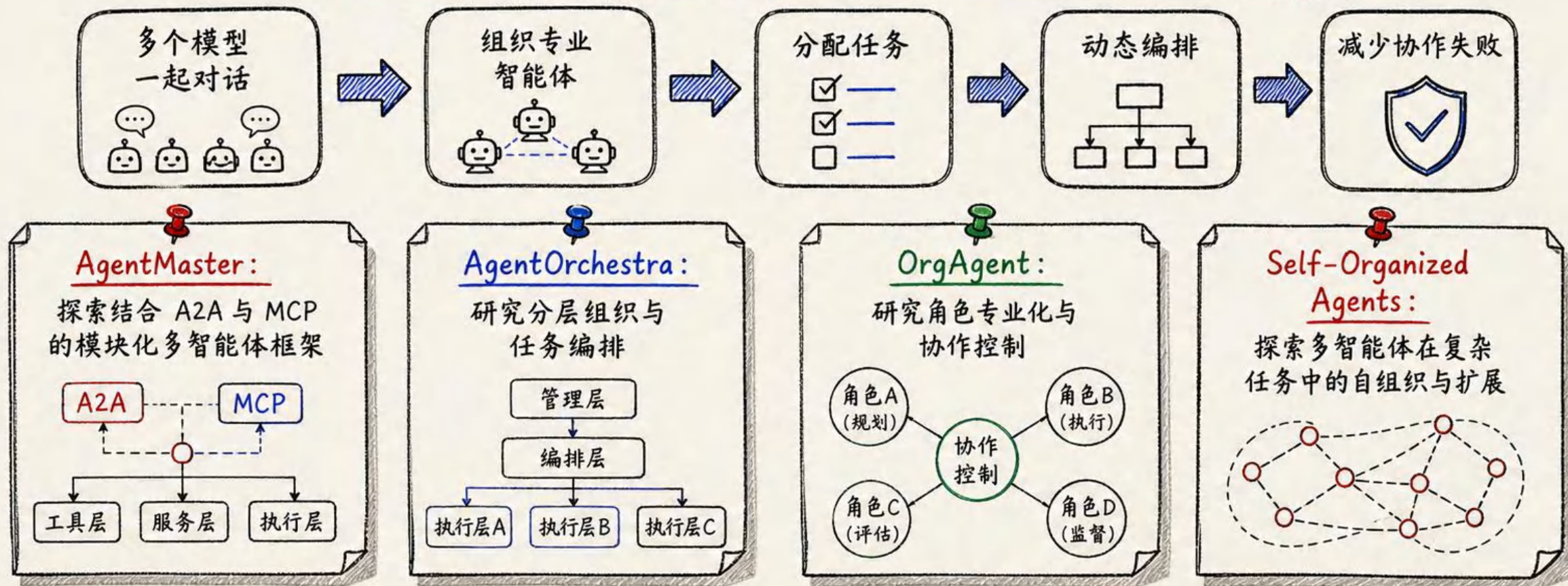
如果所有动作都由一个智能体完成，出错后难以定位责任链



1.1 多智能体协作会成为趋势

近年的多智能体研究正在从“多个模型一起对话”

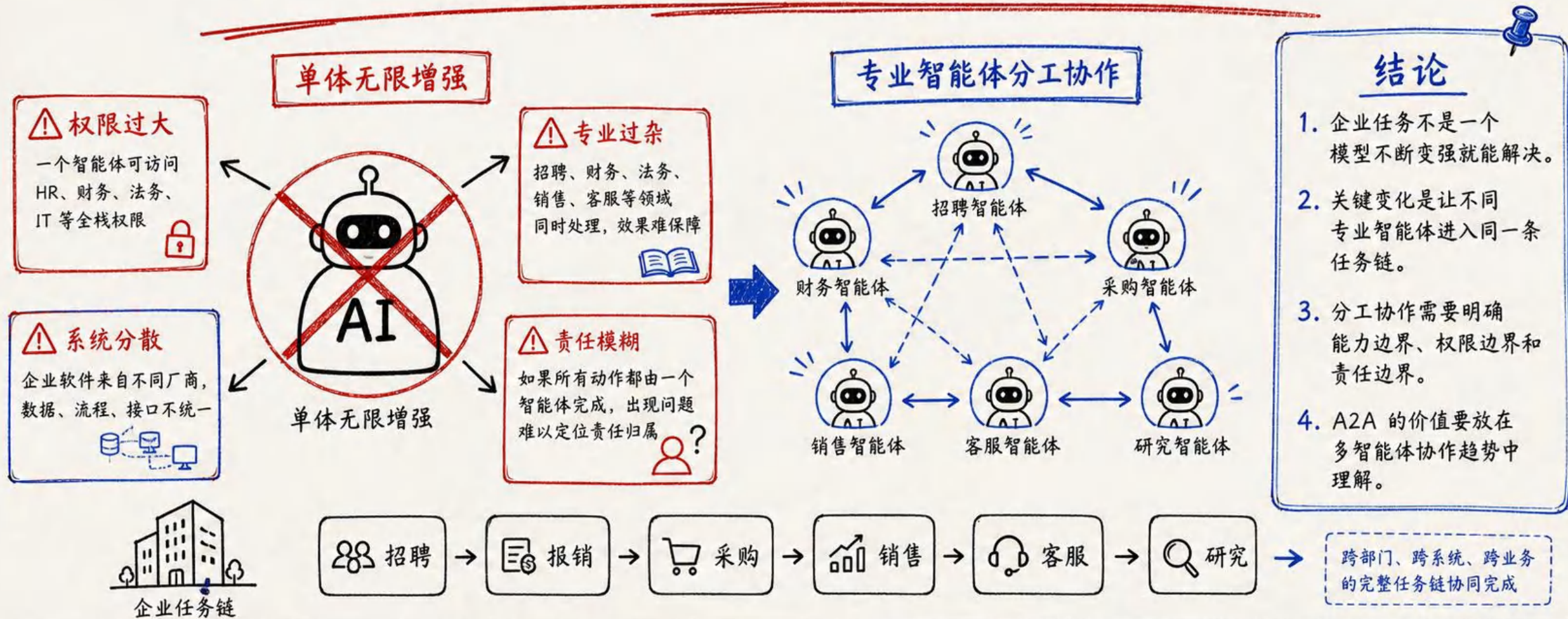
转向“如何组织专业智能体、如何分配任务、如何动态编排、如何减少协作失败”



来源: <https://arxiv.org/abs/2507.21105>; <https://arxiv.org/abs/2506.12508>; <https://arxiv.org/abs/2604.01020>; <https://arxiv.org/abs/2404.02183>



AI 智能体的下一步不是单体无限增强， 而是一组专业智能体分工协作



≡ A2A 解决的是跨厂商、跨框架、跨系统的智能体互操作问题 ≡

它要回答：

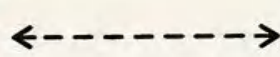


- 一个智能体如何发现另一个智能体？
- 如何知道对方能做什么？
- 如何发送任务？
- 如何跟踪状态？
- 如何交付结果？

★ A2A 官方规范围绕 Agent Card、Task、Message、Artifact，任务状态、流式更新、推送通知和安全控制展开。

跨厂商：

厂商 A



厂商 B

跨框架：

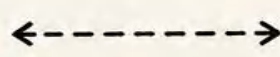
框架 X



框架 Y

跨系统：

系统 1



系统 2

≡ 协议对象 ≡



Agent Card



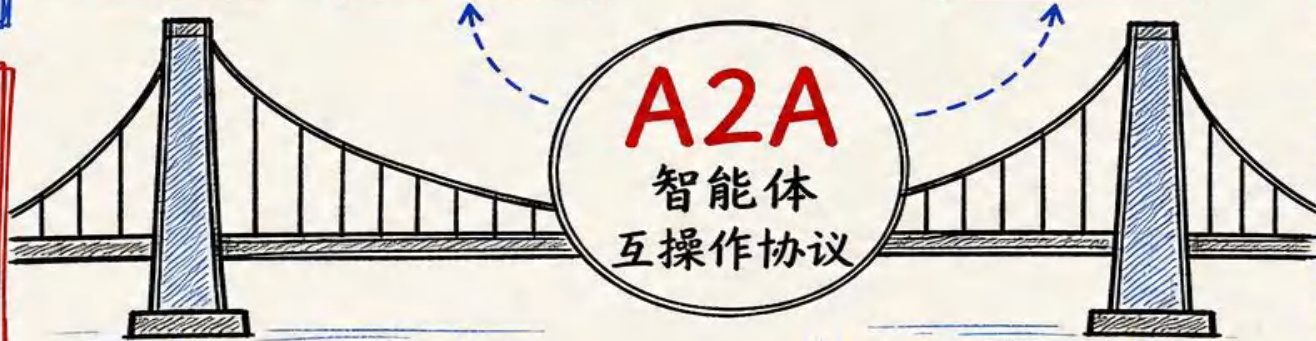
Task



Message



Artifact



任务状态

跟踪与管理任务进展



流式更新

实时传递状态变化



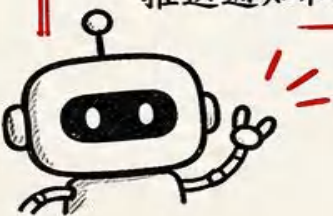
推送通知

事件驱动及时推送



安全控制

认证、授权与审计



来源：<https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/>；<https://a2a-protocol.org/latest/specification/>



清新

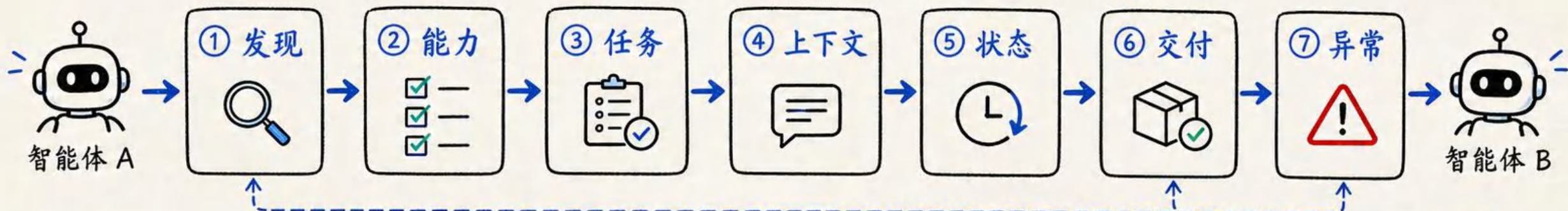
更多免费行业报告

并购家

www.ipoiipo.cn

研究报告

A2A 要回答的是发现、能力、任务、状态和交付问题 ☆



💡 A2A 提供**标准化**机制，让不同智能体像**团队成员**一样协作，而不是各自为战。



① **发现**：一个智能体如何发现另一个智能体？

② **能力**：如何知道对方能做什么？

③ **任务**：如何发起任务？

④ **上下文**：如何传递必要信息？

⑤ **状态**：如何跟踪任务进展？

⑥ **交付**：如何接收结果？

⑦ **异常**：如何处理失败、拒绝和权限问题？

★ 解决这些问题，才能真正实现智能体之间的**互操作**！

关键目标

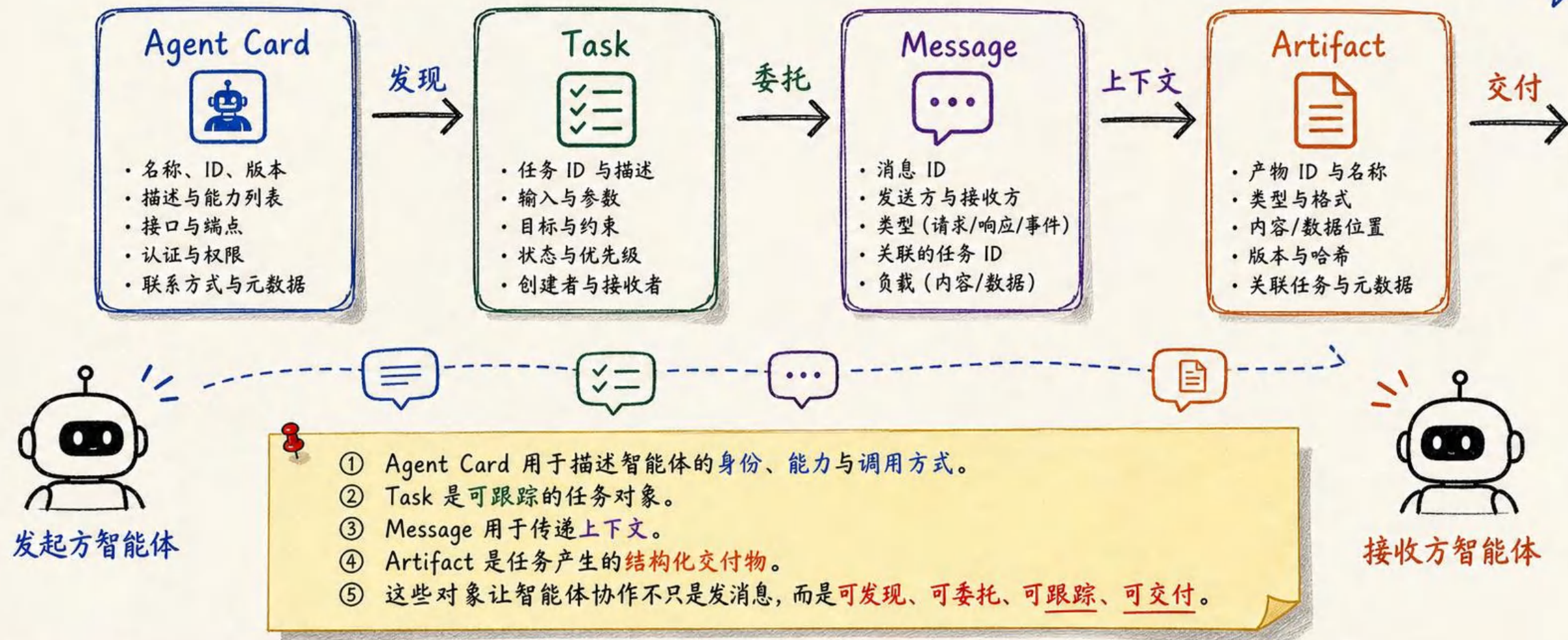
- ☒ 可发现
- ☒ 可理解
- ☒ 可协作
- ☒ 可追踪
- ☒ 可交付
- ☒ 可治理



来源：<https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/>；<https://a2a-protocol.org/latest/specification/>



Agent Card、Task、Message、Artifact 构成 A2A 的核心协议对象



MCP 和 A2A 是互补关系，不是替代关系

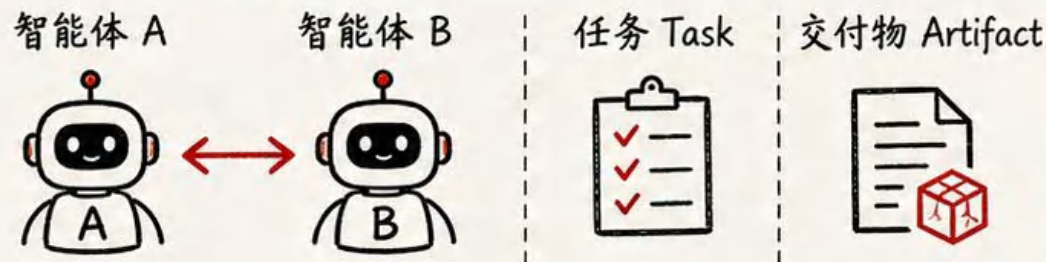


MCP: 模型 / 工具 / 数据



- ✓ MCP 连接模型与工具、数据。
- ✓ MCP 解决“智能体能用什么”。

A2A: 智能体 / 智能体



- ✓ A2A 连接智能体与智能体。
- ✓ A2A 解决“智能体能委托谁”。

💡 MCP 让智能体连接外部世界的能力资源

工具接入 + 协作委托

💡 A2A 让智能体连接其他智能体的智能能力



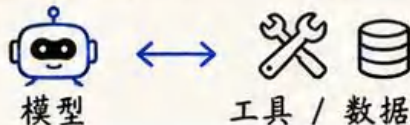
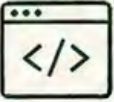
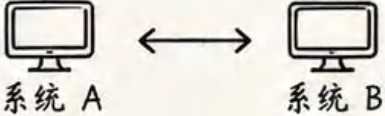
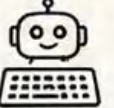
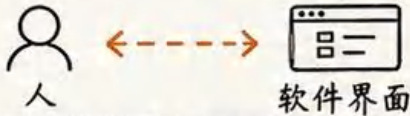
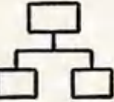
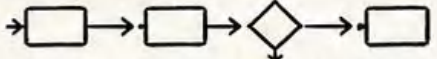
两者共同支撑智能体从调用工具走向协作委托。



清



MCP、A2A、API 接口、机器人流程自动化和工作流系统各有分工

机制	连接对象	适合问题	局限
 MCP: 模型连接工具和数据	 模型 工具 / 数据	让模型安全、可靠地访问工具、数据库、API 和文件等资源。	不负责多个智能体之间的协作、委托和结果交付。
 A2A: 智能体之间发现、委托、跟踪和交付	 智能体 智能体	需要多个专业智能体协同完成复杂任务的场景。	不直接连接底层工具和数据，需要依赖 MCP 或 API。
 API 接口: 系统之间调用能力或数据	 系统 A 系统 B	系统间直接调用能力或数据交换的场景。	缺乏语义、状态、上下文管理，协作能力有限。
 机器人流程自动化: 模拟人在界面上的操作	 人 软件界面	没有开放接口的老旧系统或必须走界面的操作场景。	脆弱、易变、难以维护，不具备智能协作能力。
 工作流系统: 预设流程、审批和状态流转	 流程 / 任务 / 人	结构化、标准化、可预测的流程与审批管理场景。	灵活性弱，难以处理开放式任务和跨系统智能协作。

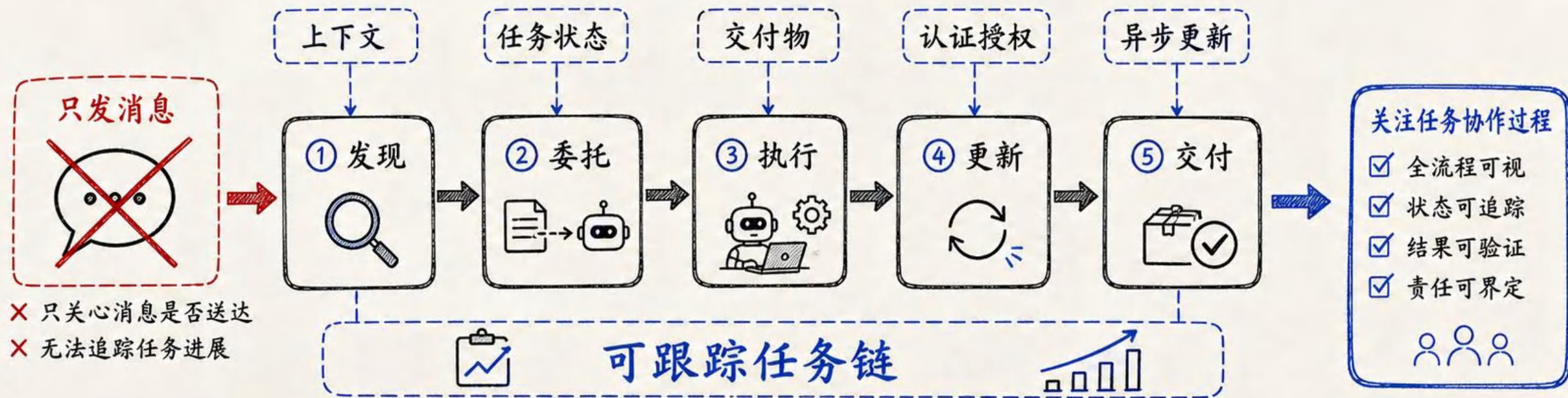
 不是谁替代谁，而是 不同层次的协作拼图。



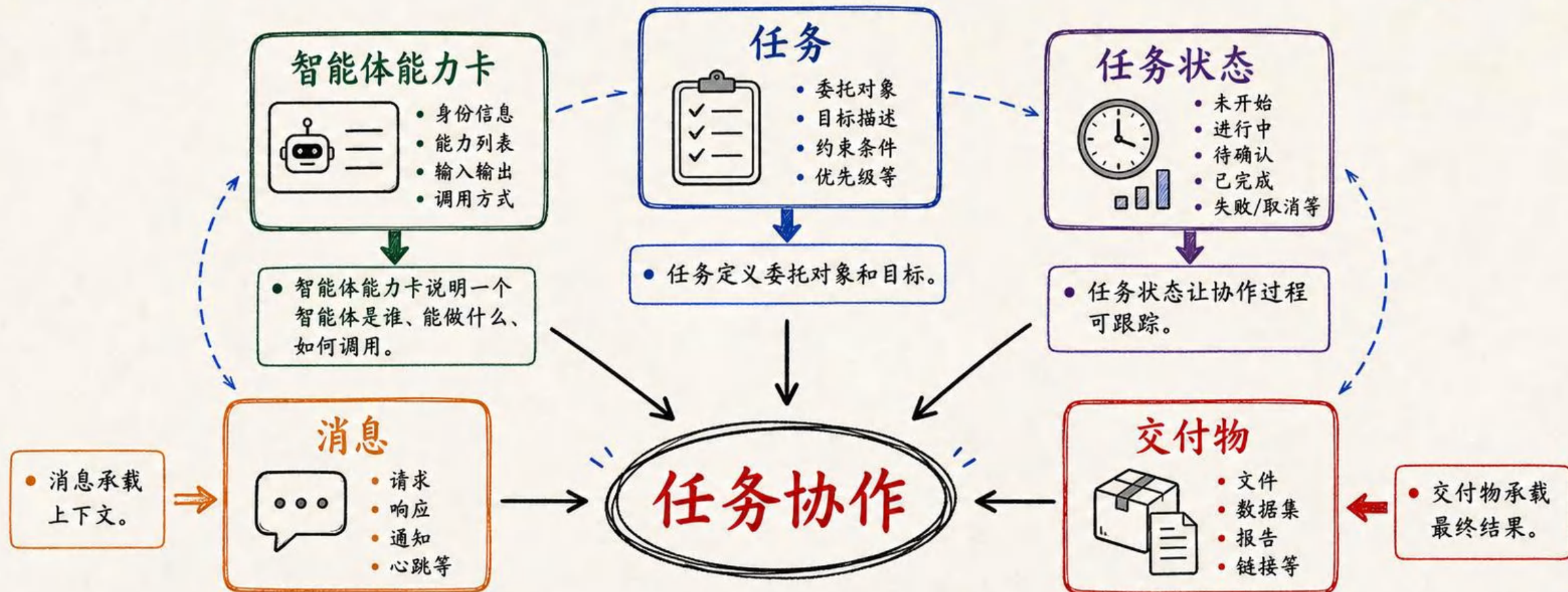
≧ A2A 标准化的是 任务协作过程，而不只是消息传递 ≦



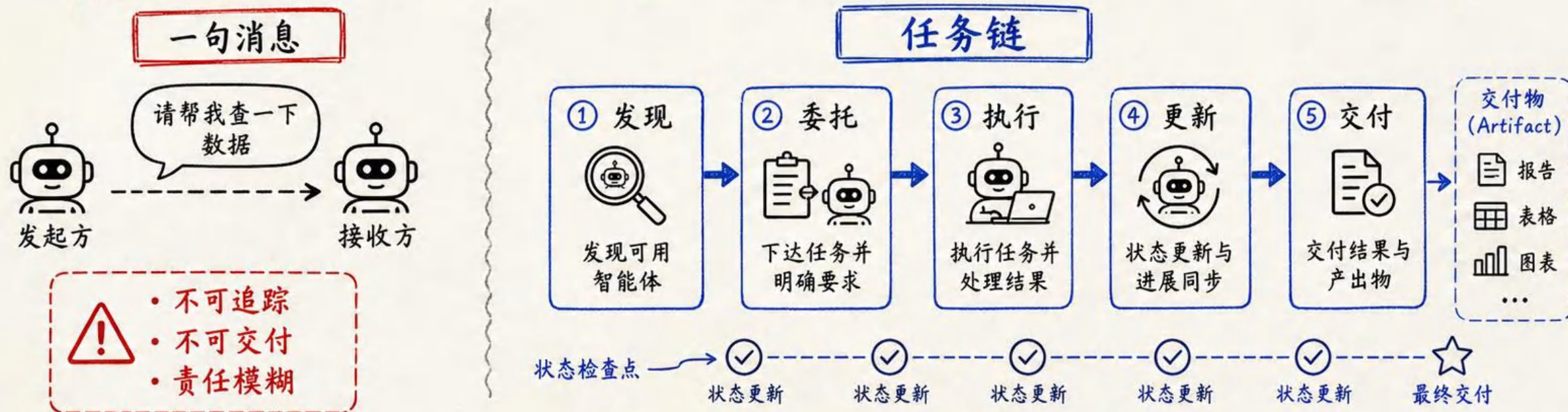
- ✓ A2A 关注任务状态、上下文、交付物、异步更新、认证授权。
- ✓ 协作不是一句消息是否送达，而是任务是否被发现、委托、执行、更新和交付。
- ✓ 长任务、流式更新和推送通知让任务协作过程可跟踪。



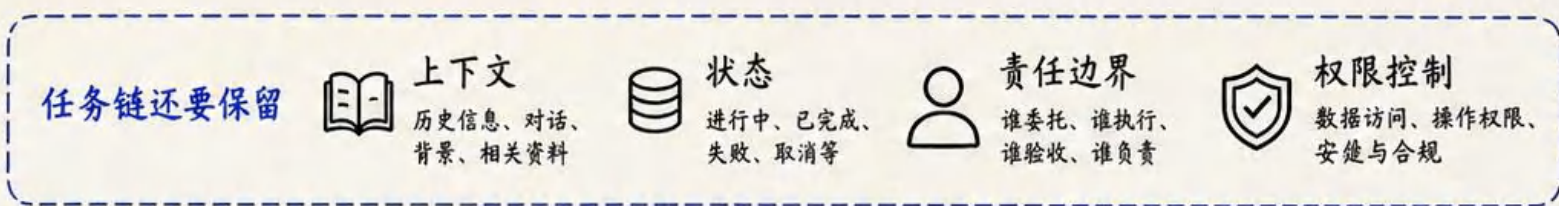
智能体能力卡、任务、任务状态、消息、交付物共同定义任务协作



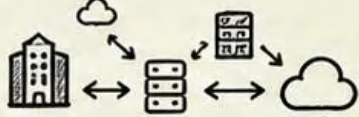
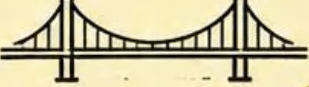
A2A 标准化的不是一句消息， 而是一条从发现、委托、执行、更新到交付的任务链

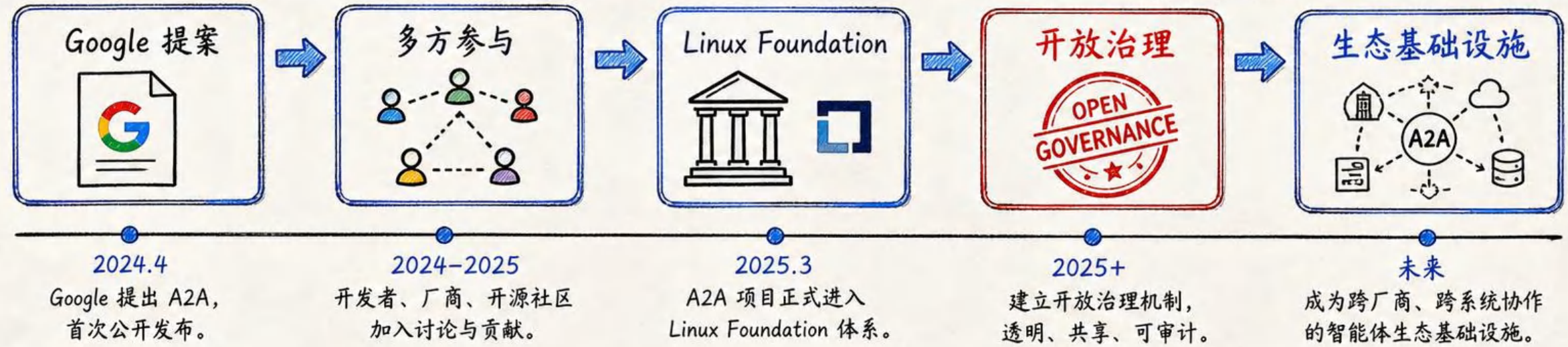


- ① 一句消息只能表达请求。
- ② 任务链要表达发现、委托、执行、状态更新、结果交付。
- ③ 任务链还要保留上下文、状态和责任边界。
- ④ 因此 A2A 的关键不只是通信，而是任务协作过程。

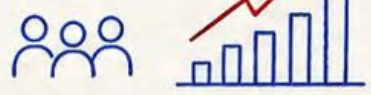


A2A 正从厂商提案走向开放治理

- A2A 最初由 Google 推动提出。
- A2A 项目进入 Linux Foundation 体系。
- 开放治理强化了跨厂商、跨系统协作的基础设施属性。
- 协议要成为生态基础设施，不能只停留在单一厂商提案。

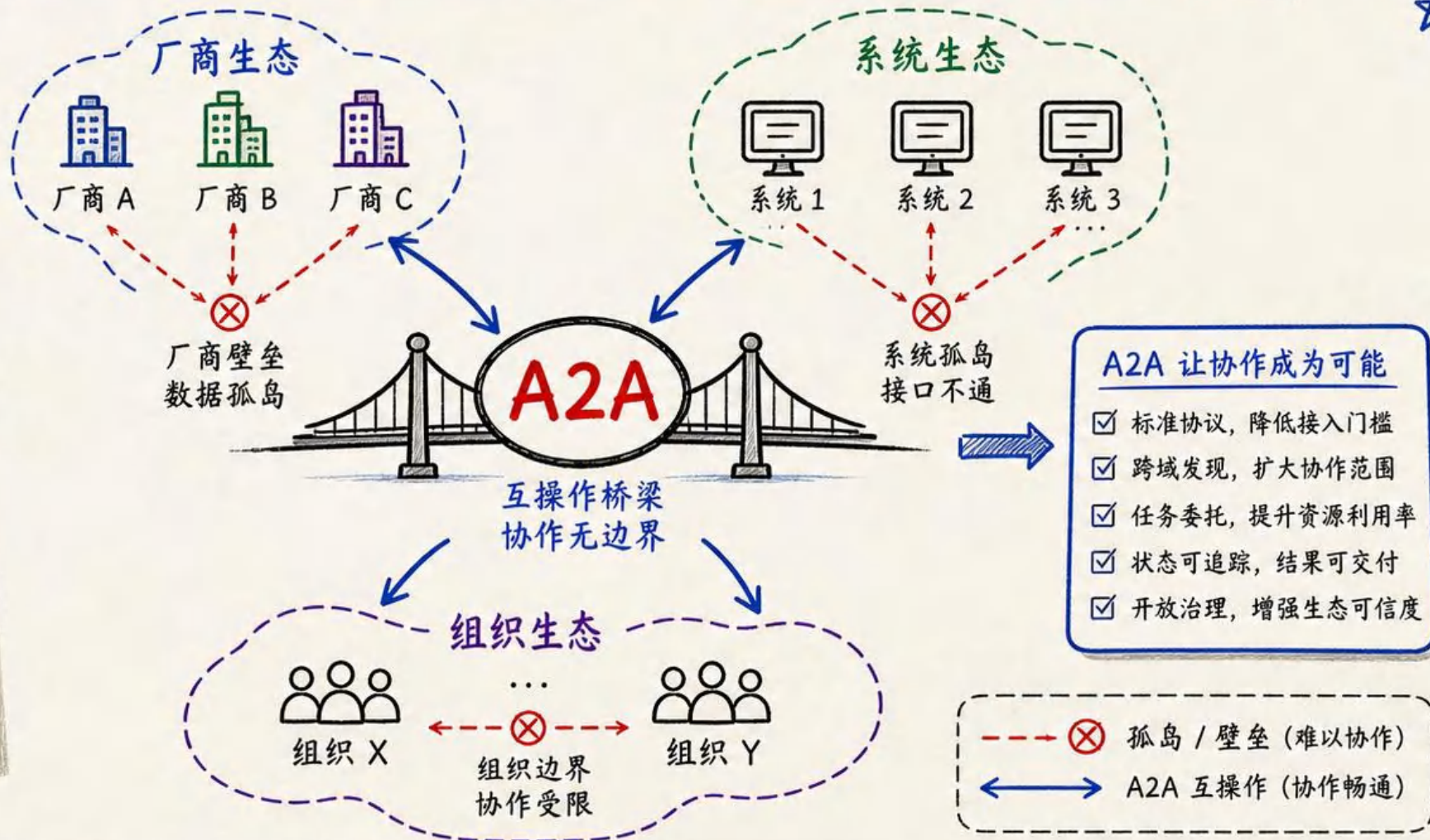


来源: <https://www.linuxfoundation.org/press/linux-foundation-launches-the-agent2agent-protocol-project>

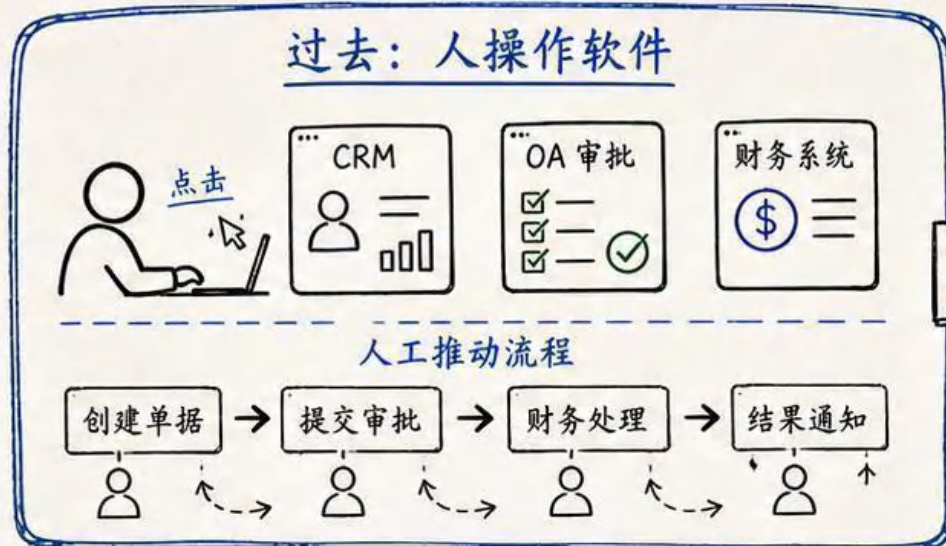


A2A 的基础设施价值来自跨厂商、跨系统、跨组织协作的需求

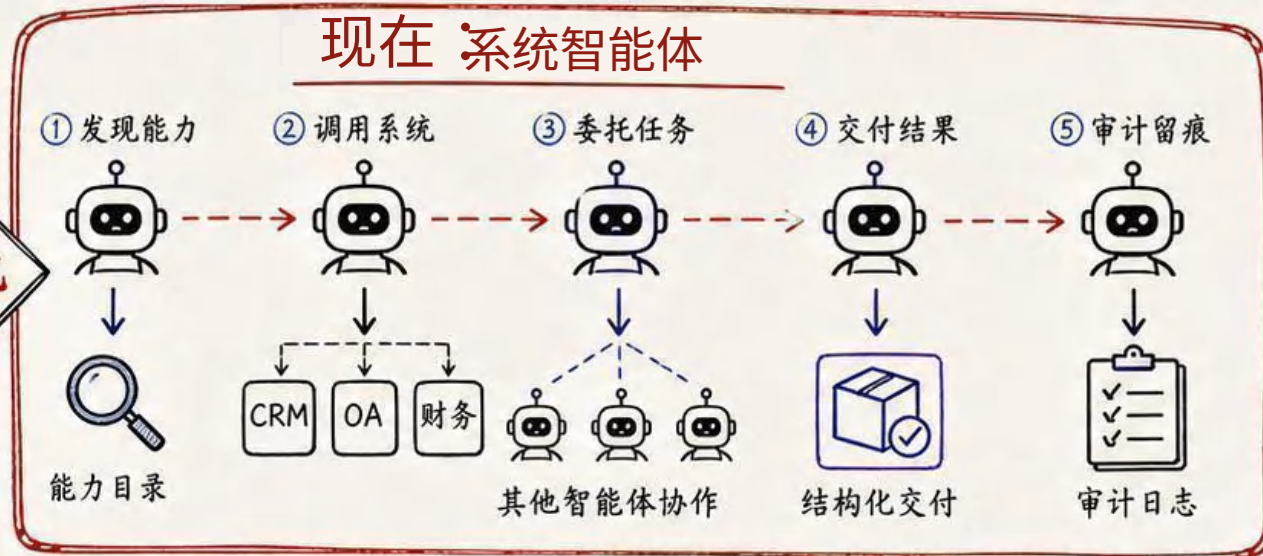
- ① 企业环境通常由多个厂商的软件、多个系统、多个组织边界组成。
- ② 智能体如果只能在单一厂家或单一系统内协作，价值会被限制。
- ③ A2A 的基础设施价值来自跨厂商、跨系统、跨组织协作。
- ④ 开放治理有助于降低生态协作的不确定性。



A2A 会改变企业软件入口



依赖人工点击、切换、等待，流程效率低且难跟踪。



智能体发现、调用、委托、跟踪、交付并留痕，流程自动化、可观察、可审计。

- ✓ 过去：人操作软件，人推动流程。
- ✓ 现在：系统智能体
- ✓ 企业软件入口会从“给人使用的界面”扩展为“可被智能体发现、理解、调用和审计的能力”。
- ✓ 软件价值还取决于能否进入智能体任务链。

核心观点

企业软件的未来入口，
在智能体，而不只是人。



界面入口

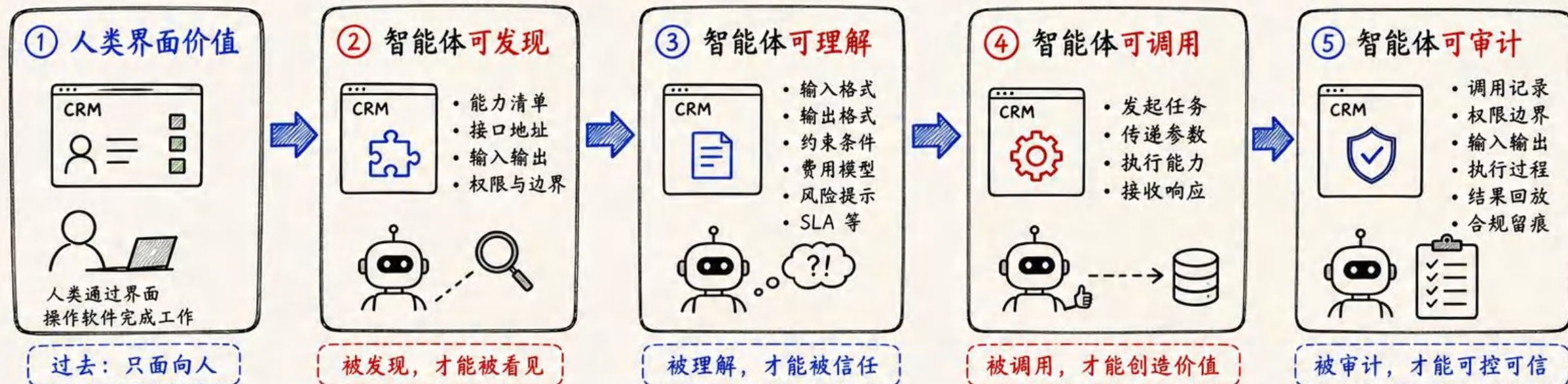
能力入口

人工点击

任务委托



软件价值还取决于能否被智能体**发现**、**理解**、**调用**和**审计**



💡 软件价值正在从“界面可用”走向“能力可用”

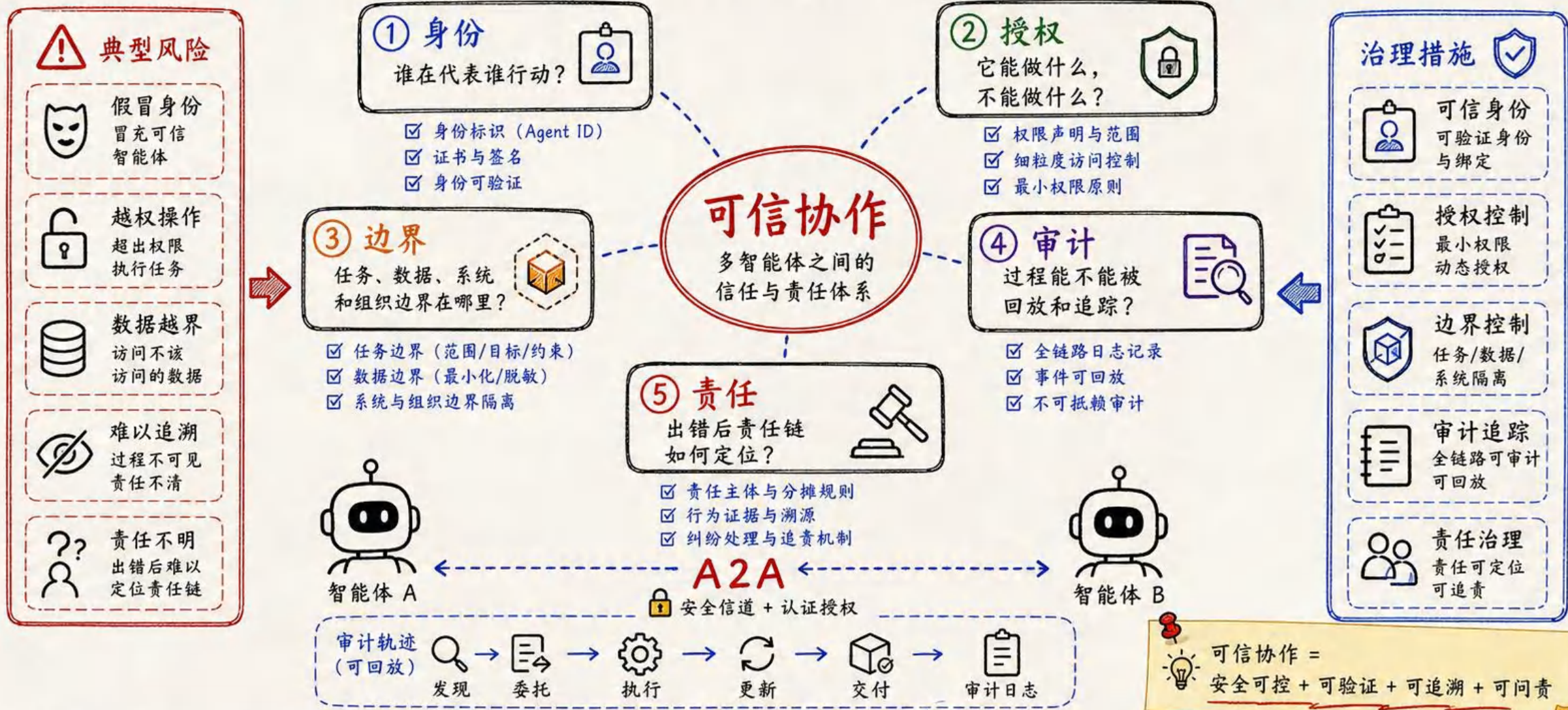
- ① 企业软件不再只面向人类界面，也要面向**智能体任务链**。
- ② 软件需要被智能体**发现**：能力、接口、边界可被识别。
- ③ 软件需要被智能体**理解**：输入、输出、约束、费用和 risk 可解释。
- ④ 软件需要被智能体**调用**：任务可以进入系统能力。
- ⑤ 软件需要被智能体**审计**：调用过程、权限边界和结果可回放。

☆ 核心判断

软件能否进入智能体任务链，将决定其未来的**生态位**和**商业价值**。



A2A 的难点不只是技术，而是可信协作



A2A 是否是“AI 时代 HTTP 协议”尚无定论

① HTTP 是 Web 时代的基础通信协议类比。

② A2A 有成为多智能体协作基础协议的可能。

③ 但 A2A 仍处在生态形成阶段，标准稳定性、采用范围和治理成熟度仍需观察。

④ 更谨慎的判断是：A2A 是重要方向，但尚不能简单等同于 HTTP。

类比对象	相似处	差异	审慎判断
 HTTP (Web 基础协议)	- 通信标准 - 通用、可扩展、跨平台 - 构建生态的关键基石	- HTTP 非常简单，成功在于简洁 - A2A 语义更复杂（任务、状态、权限、履约等） - 所解决问题空间更大	部分相似但不完全等同 ★★★★★☆☆
 API (应用接口)	- 定义调用方式 - 能力暴露与发现 - 支撑系统互联互通	- API 面向系统/功能调用 - A2A 面向智能体协作（发现、委托、跟踪、交付） - 生命周期与状态模型更完整	有相似性但定位更高层 ★★★★★☆☆
 工作流 (流程编排)	- 任务分解与编排 - 状态流转与跟踪 - 支持多方协作	- 工作流由单一系统预设流程 - A2A 是跨系统、跨智能体的动态协作网络 - 强调自治与协商	互补而非替代 ★★★★★☆☆
 SWIFT (金融报文网络)	- 跨组织安全报文传递 - 标准化与治理重要 - 网络效应驱动价值	- SWIFT 专注金融结算场景 - A2A 面向通用智能体协作，场景更广 - 语义、对象与安全模型更复杂	治理方向类似，场决与语义不同 ★★★★★☆☆
 企业服务总线 (ESB)	- 集成异构系统 - 路由、转换、适配 - 降低集成复杂度	- ESB 多为中心化/网关式 - A2A 面向点对点/多对多协作 - 强调智能体自治与协商式交互	有借鉴价值但架构范式不同 ★★☆☆☆☆



A2A 是否成为“AI 时代 HTTP”，取决于生态规模、标准稳定性与治理成熟度！

关键观察点

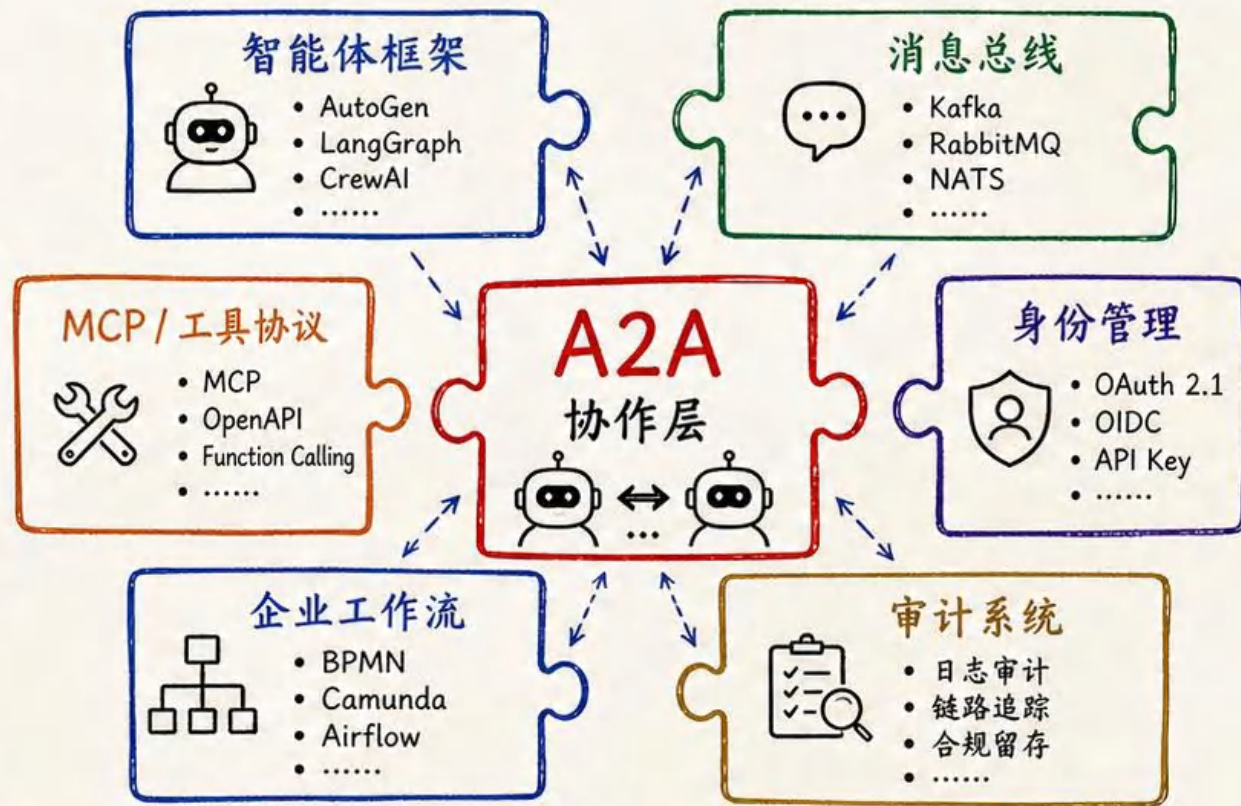
- ☒ 标准稳定性与版本演进是否形成长期共识 
- ☒ 厂商与框架的广泛采用与互操作落地深度 
- ☒ 治理机制是否开放、透明、可持续 
- ☒ 能否支撑大规模、复杂任务协作的真实场景 
- ☒ 安全、信任、隐私与合规体系是否健全 

结论：A2A 有潜力成为 AI 时代**重要基础协议**，但尚未达到“**HTTP 级别**”的确定性。 



≡ A2A 更像一个方向，会和不同框架、消息总线、工具协议一起组合落地 ≡

- ① A2A 不会孤立落地。
- ② 它会和不同智能体框架、消息总线、工具协议、身份管理和企业工作流组合。
- ③ 适合的场景包括多轮对话、任务编排、异步回调、状态跟踪、长链路协作。
- ④ 因此更合理的判断是组合式落地，而不是单一协议吞并全部系统。



💡 关键判断

A2A 更像一个方向和基础设施能力，需要与多种现有技术生态组合，才能在真实场景中创造价值。

适合的场景 →



多轮对话



任务编排



异步回调



状态跟踪



长链路协作





A2A 解决智能体如何协作，智能体链接解决智能体如何认知、信任 and 选择



A2A：如何协作

关注任务协作过程



核心对象

Agent Card

Task

Message

Artifact

...

目标：让智能体之间高效协同完成任务

从连接协议



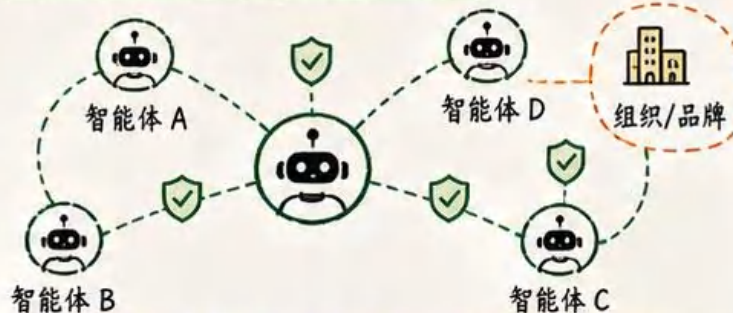
数字委托网络



Agent Link：如何认知、信任和选择

关注认知信任与选择决策

可信实体
品牌声誉
能力画像
历史记录
...



信任链路

信息源

证据

验证

信任结论

目标：让智能体选择可信对象，做出可靠决策



① A2A 关注智能体之间如何发现、委托、跟踪和交付任务。

② Agent Link 关注智能体如何判断信息源、实体、品牌、能力和其他智能体是否可信。

③ 没有智能体链接，协作只停留在连接层。

④ 有了智能体链接，A2A 才可能进入真正的数字委托网络。

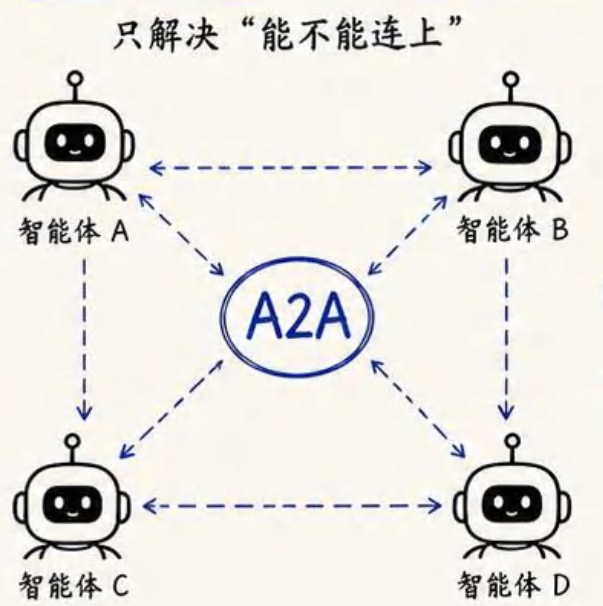
协作靠连接，
信任靠链接！



没有智能体链接，A2A 只是连接协议； 有了智能体链接，A2A 才可能成为真正的数字委托网络

只有 A2A：连接协议

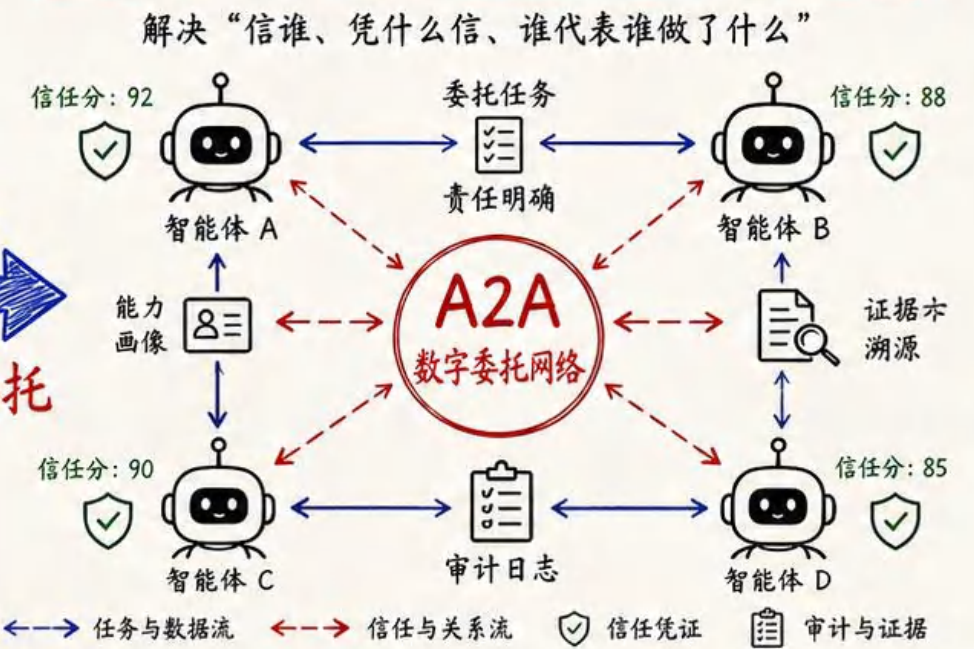
- 连接协议特征
- ✓ 发现智能体
 - ✓ 发送任务
 - ✓ 跟踪状态
 - ✓ 交付结果



- 局限
- ⚠ 不知道对方是否可信
 - ⚠ 不知道能力是否可靠
 - ⚠ 不知道谁代表谁
 - ⚠ 出错后无法追责

连接
可信委托

A2A + Agent Link：数字委托网络



- 数字委托网络特征
- ✓ 关系信誉
 - ✓ 能力可验证
 - ✓ 证据可溯源
 - ✓ 委托可审计
 - ✓ 责任可追踪

核心价值
从“能连上”到
“敢委托、可追责、
可持续协作”

信任是协作的
操作系统

- ① 连接协议回答“能不能连上”。
- ② 智能体链接进一步回答“信谁、凭什么信、信任之后谁代表谁做了什么”。
- ③ 数字委托网络需要关系信誉、证据溯源和委托审计。
- ④ A2A 与 Agent Link 组合后，协作才从连接走向可信委托。

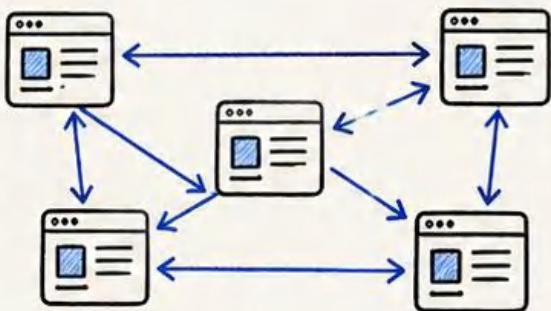
让我放心
地去协作！



从网页排名算法 (PageRank) 到智能体链接 ☆

PageRank: 网页 ↔ 网页

Web 时代, PageRank 通过网页之间的
链接关系衡量重要性。



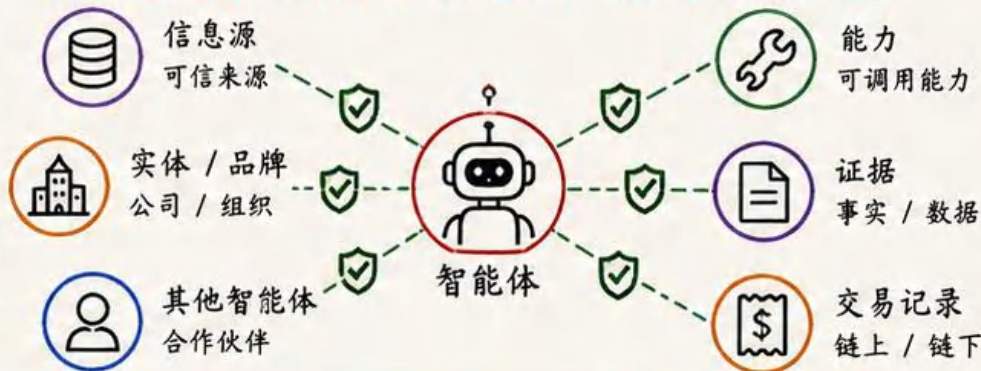
排序网页, 衡量重要性

排序网页

选择可信
能力节点

Agent Link: 智能体 / 能力 / 证据 / 交易

智能体时代, 被判断的不只是网页, 而是
信息源、实体、品牌、能力、智能体和交易记录。



选择可信能力节点, 支撑智能体决策与协作委托

- ✓ 智能体链接不是简单超链接, 而是**关系信誉、证据溯源**和**委托审计**的综合结构。
- ✓ 它决定智能体是否信任、是否选择、是否委托。

- ✓ 信誉来自关系网络与行为记录。
- ✓ 证据来自可验证的事实、数据和交易。
- ✓ 审计来自可回放的委托执行与责任链。



核心思想

从“链接越多越重要”,
走向“可信越强越可委托”。

公式化表达:

Agent Link

=

关系信誉

+

证据溯源

+

委托审计

智能体时代被排序的不再只是网页， 而是信息源、实体、品牌、能力、智能体和交易记录

Web 时代：网页 / 内容 / 点击



主要排序网页和内容，
追求被人看见和点击。

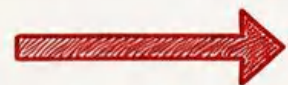
- 1 网页 A
- 2 网页 B
- 3 网页 C
- 4 网页 D
- 5 网页 E
- ...

排序信号 (PageRank 思想)

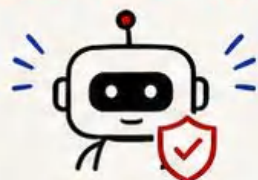
- ✓ 链接数量
- ✓ 链接质量
- ✓ 点击量
- ✓ 访问量
- ✓ 停留时间
- ✓ ...



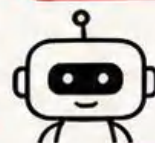
网页排序



能力与信任排序



智能体时代：信息源 / 实体 / 品牌 / 能力 / 智能体 / 交易记录

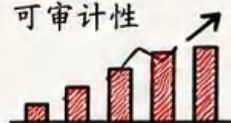


要排序可信信息源，
让智能体信任和调用。

- 1 信息源
- 2 实体
- 3 品牌
- 4 能力
- 5 智能体
- 6 交易记录
- ...

排序信号 (信任与效果导向)

- ✓ 可信度 / 声誉
- ✓ 权威性 / 引用证据
- ✓ 能力表现 / 成功率
- ✓ 使用反馈 / 满意度
- ✓ 安全性 / 合规性
- ✓ 交易完成度 / 纠纷率
- ✓ 透明度 / 可审计性
- ✓ ...



- ① Web 时代主要排序网页和内容。
- ② 智能体时代要排序可信信息源。
- ③ 还要排序实体、品牌、能力、智能体和交易记录。
- ④ 排序对象变化意味着商业竞争从“被人看见”扩展为“被智能体信任和调用”。



核心洞察

谁能获得智能体的信任和调用，谁就能获得新时代的流量、订单和增长。



商业结果将从获得**点击**扩展为获得**推荐、调用、委托、下单和支付**

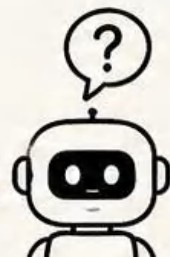


- ① 过去商业结果主要围绕**曝光、点击和转化**。
- ② 智能体时代，商业结果会扩展为获得**推荐、调用、委托、下单和支付**。
- ③ 品牌和软件需要被智能体看见、信任、比较、调用。
- ④ 竞争焦点从人类注意力扩展到**智能体选择权**。

过去：Web 商业



人类用户



智能体决策

- ✓ 看见与发现
- ✓ 信任与评估
- ✓ 比较与选择

决策依据



可信身份



能力描述



口碑评价



价格与成本

现在：智能体商业



--- 在候选中被纳入

--- 能力被调用执行

--- 任务被委托给你

--- 订单被创建

--- 完成支付与结算



企业 / 服务提供方



核心提示

商业价值不再只取决于是否被人看到，而是能否进入并赢得智能体的**任务链路**。



人类注意力



智能体选择权

- ✓ 从被动曝光 → 主动选择
- ✓ 从一次点击 → 多步协作
- ✓ 从单点转化 → 全链路交易

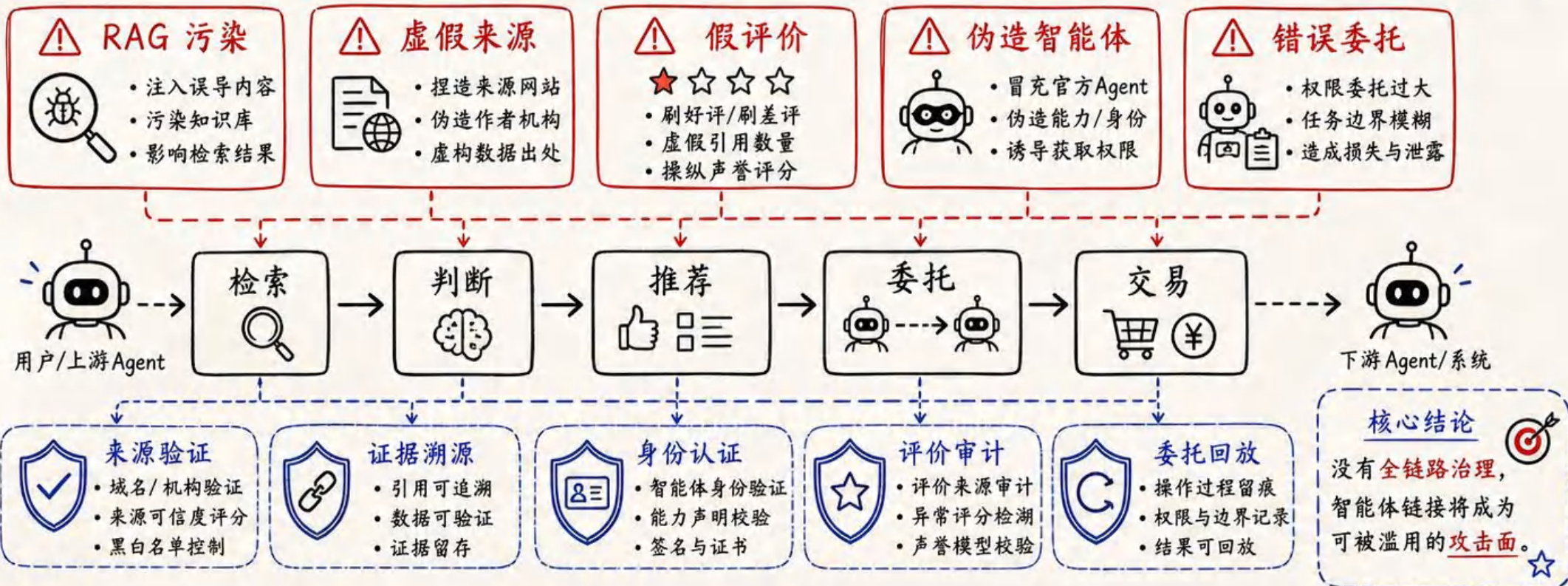


智能体链接的风险不只是垃圾链接接，还包括 RAG 污染、虚假来源、假评价、伪造智能体和错误委托

- Web 时代的链接风险包括垃圾链接和操纵排序。
- 智能体时代的链接风险会进入推理、检索、委托和交易链路。
- 关键风险包括 RAG 污染、虚假来源、假评价、伪造智能体和错误委托。
- 风险治理必须覆盖来源、证据、身份、评价和委托过程。

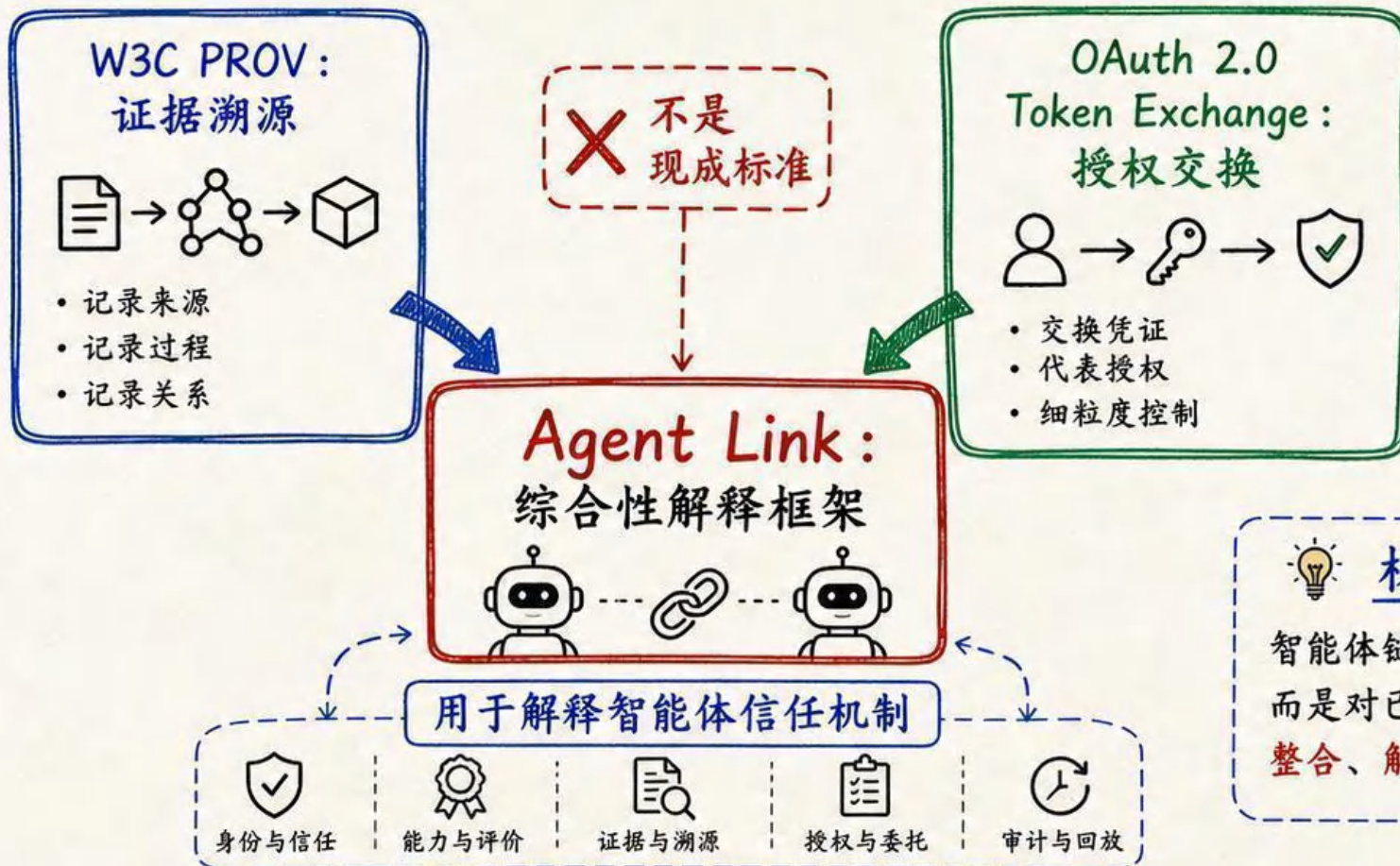
风险来源

- ☑ 开放网络
- ☑ 激励偏差
- ☑ 低成本造假
- ☑ 信息不对称
- ☑ 信任可被操纵



智能体链接 是对已有信任机制的综合性解释

- ① 智能体链接是本报告提出的**研究框架**。
- ② 它借鉴已有信任机制
- ③ W3C PROV 支撑**证据溯源**思路。
- ④ OAuth 2.0 Token Exchange 支撑**授权交换**思路。



Agent Link 有三个信任锚点：关系信誉、证据溯源、委托审计



① 关系信誉解决“信谁”。

② 证据溯源解决“凭什么信”。

③ 委托审计解决“信任之后谁代表谁做了什么”。

④ 三个锚点共同构成智能体链接的信任底座。



解决“信谁”

解决“凭什么信”

解决“信任之后谁代表谁做了什么”

⚠️ 风险提醒

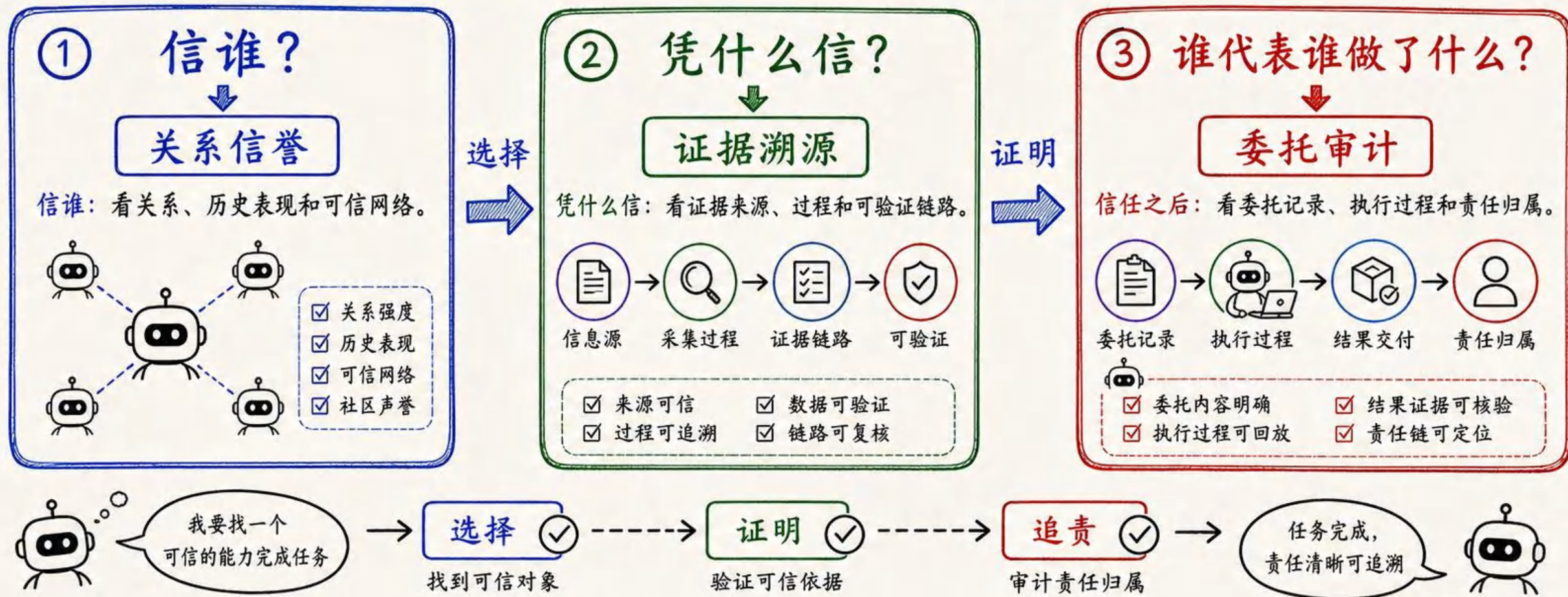
- ❌ 虚假关系与刷信誉
- ❌ 伪造证据与断链
- ❌ 越权委托与责任缺失
- ❌ 审计缺失与不可追溯

🛡️ 信任底座价值

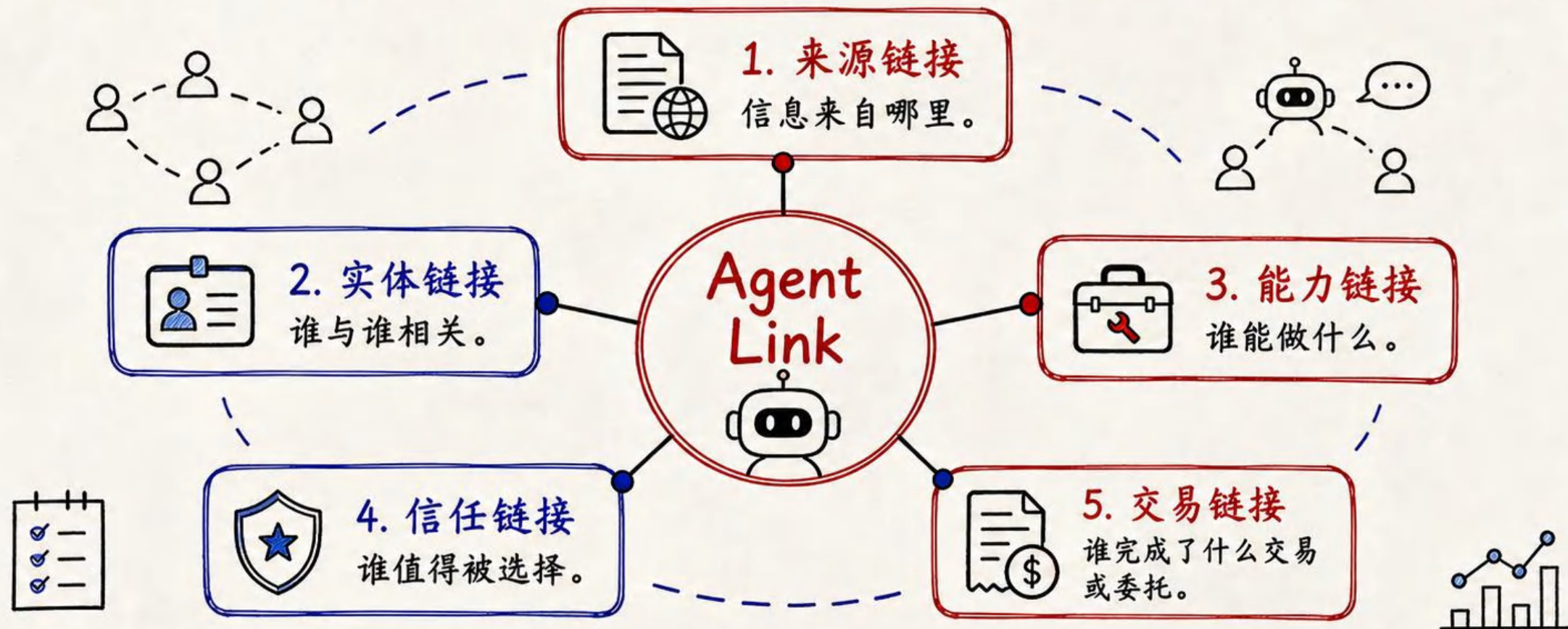
- ✅ 让智能体敢选择
- ✅ 让任务可委托
- ✅ 让结果可追责
- ✅ 让生态可持续



关系信誉解决“信谁”，证据溯源解决“凭什么信”，委托审计解决“信任之后谁代表谁做了什么”



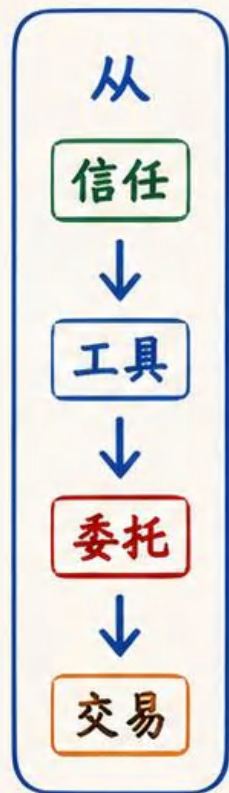
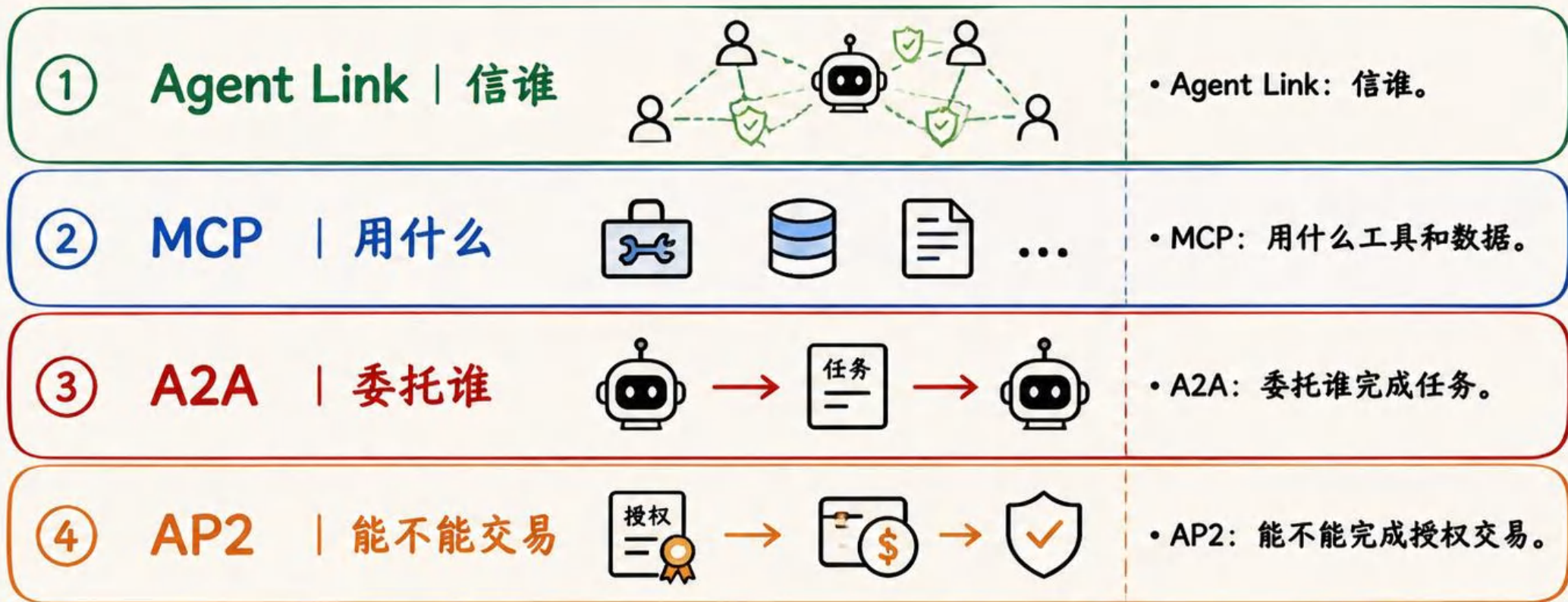
Agent Link 包括来源链接、实体链接、能力链接、信任链接和交易链接



链接不只是连接，而是可被智能体读取的信任关系。



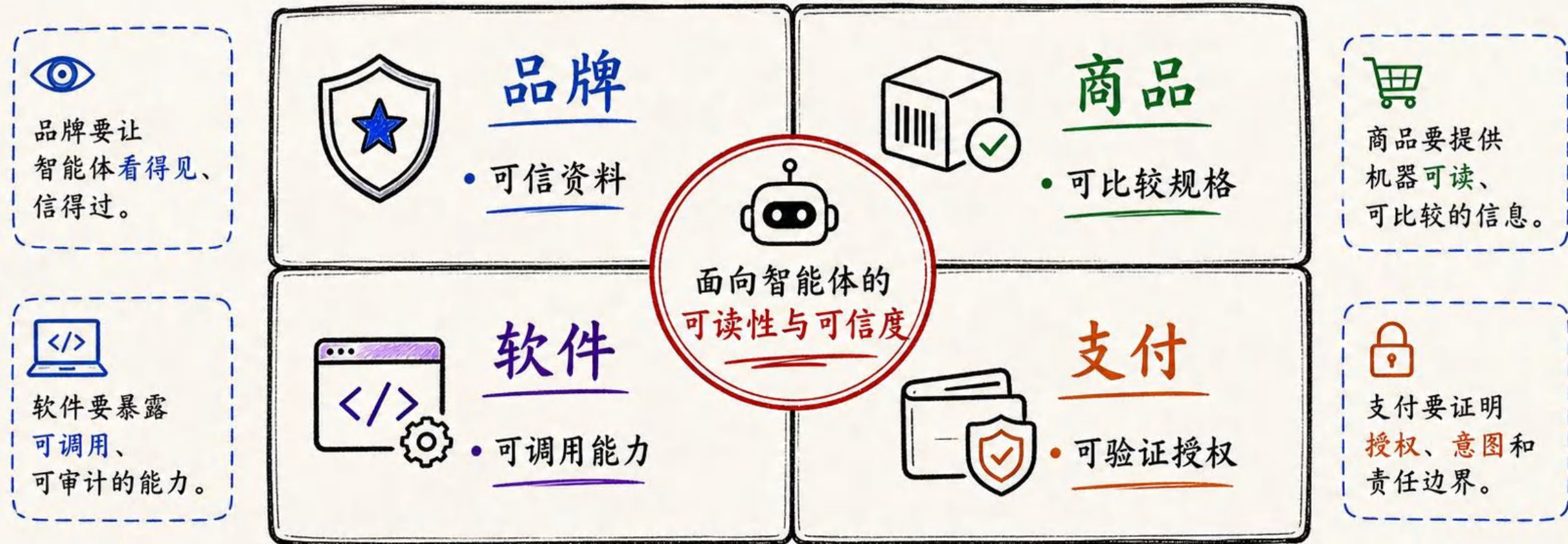
智能体链接决定智能体信谁；MCP 决定智能体能用什么；
A2A 决定智能体能委托谁；AP2 决定智能体能不能交易



💡 四层协同，构成智能体从信任判断到能力使用、任务委托，再到价值交换的完整闭环。



Agent Link 的商业后果是品牌、商品、软件 and 支付 都要增加面向智能体的可读性与可信度



Agent Link 的组织后果是形成智能体能力图谱， 让任务流向可信能力节点

① 组织不只管理岗位，
也要管理**可调用能力**。



② 智能体能力图谱
记录谁能做什么。



③ 任务会流向**可信、
可审计、边界清晰**的
能力节点。

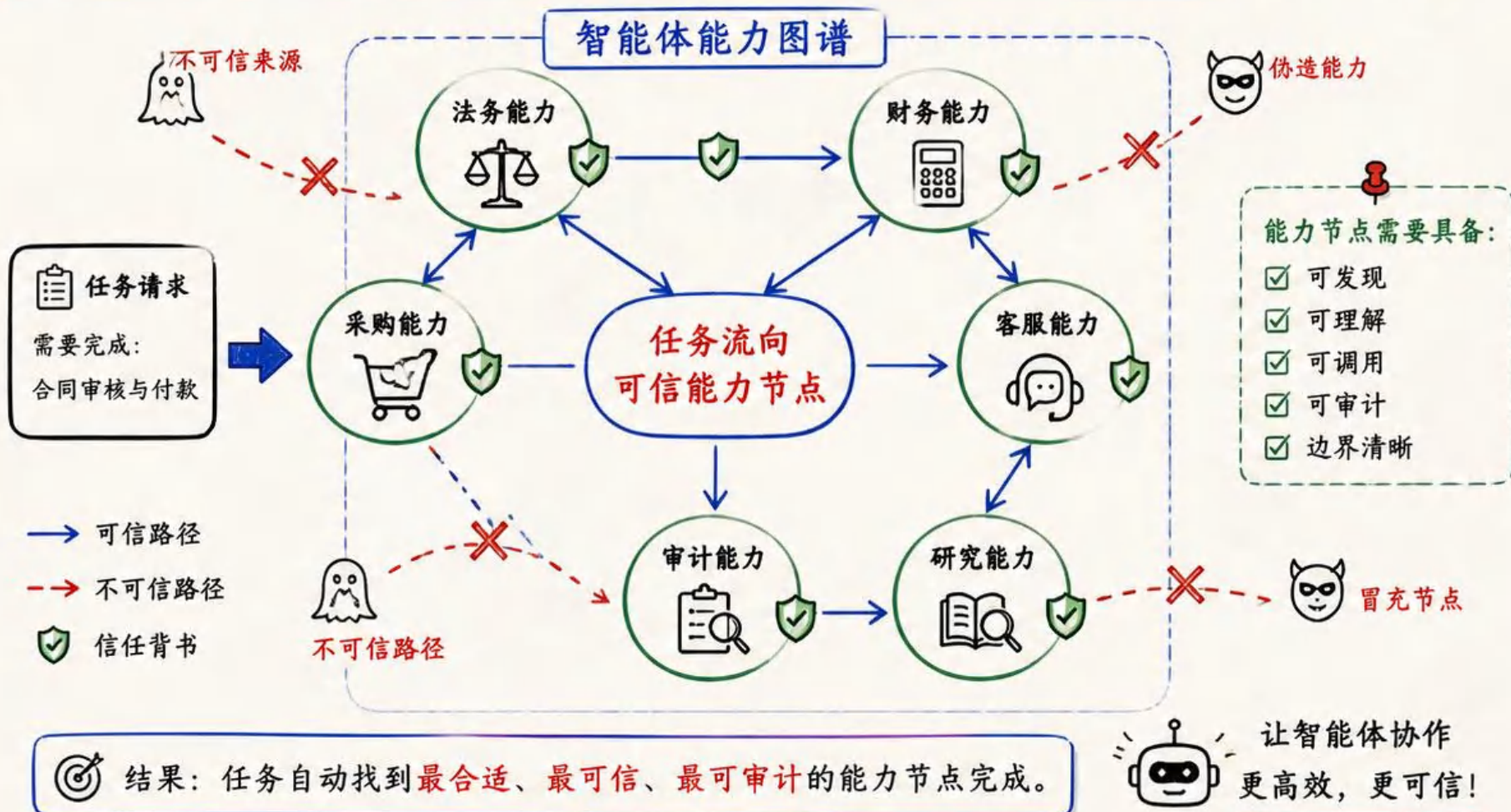


④ Agent Link 让能力
节点之间形成可选择的
信任网络。



核心观点

组织的核心资产，将从“人力”
结构”扩展为“**能力结构**”。
可被**发现**、可被**信任**、可被**调用**
的能力，才会产生商业价值。



企业真正要管理的不是智能体的智能水平， 而是它能访问什么、修改什么、代表谁行动

能访问什么？



系统数据



文件资料



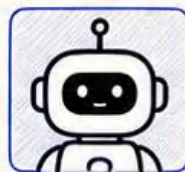
API 服务



应用资源



关键：访问边界必须清晰可控，
最小权限原则。



智能体能力与编排画像

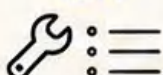
名称：财务报销智能体

类型：流程型智能体

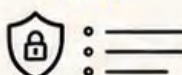
版本：v1.0.0

描述：处理员工报销申请

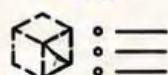
能力



权限



边界



能修改什么？



写入数据



更新状态



触发流程



关键：修改边界必须限定范围，
防止越权操作。

代表谁行动？



员工



部门



组织



角色/职责



关键：身份与责任必须明确，行动可追溯。

进入任务链之前，先定义边界



发现



评估



授权



执行



交付

可发现、可评估、可授权、可执行、可审计

从智能水平



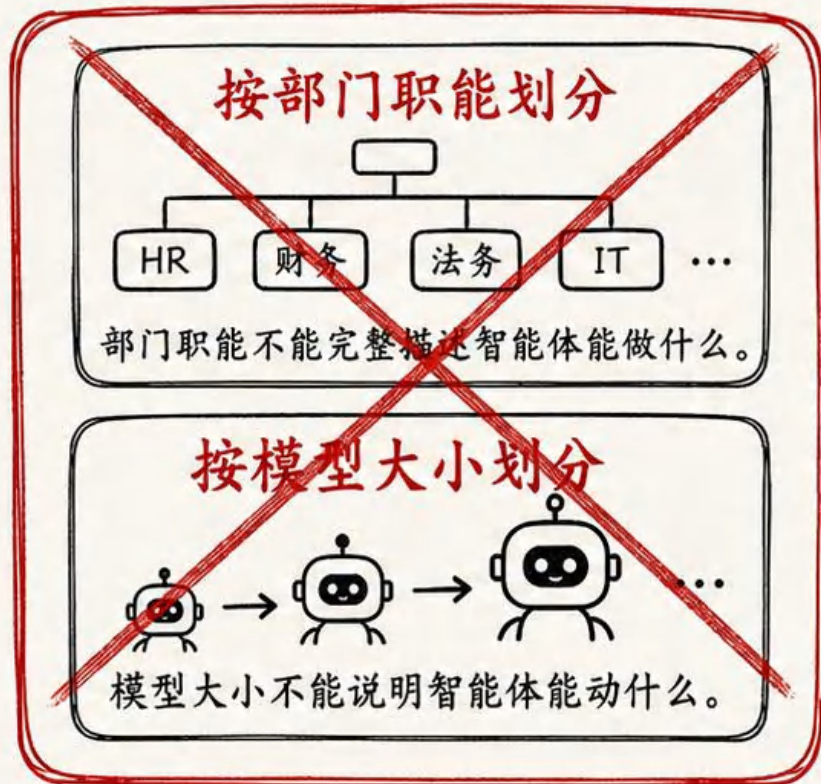
边界管理



- ✓ 企业管理的核心不是模型大小。
- ✓ 关键是访问边界、修改边界、代表谁行动。
- ✓ 智能体进入企业任务链前，必须先定义能力和权限画像。



智能体能力不应只按部门职能划分，更不能按模型大小划分



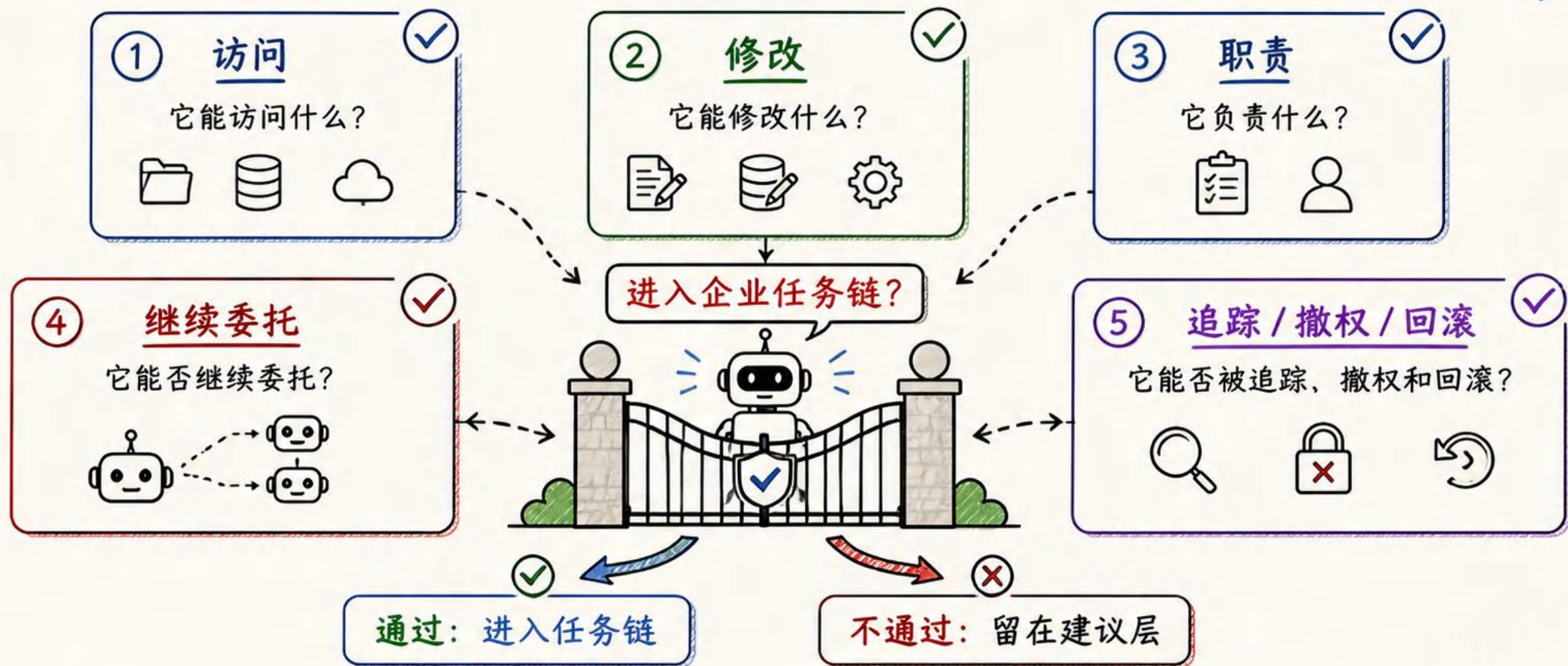
从粗分类



可治理画像



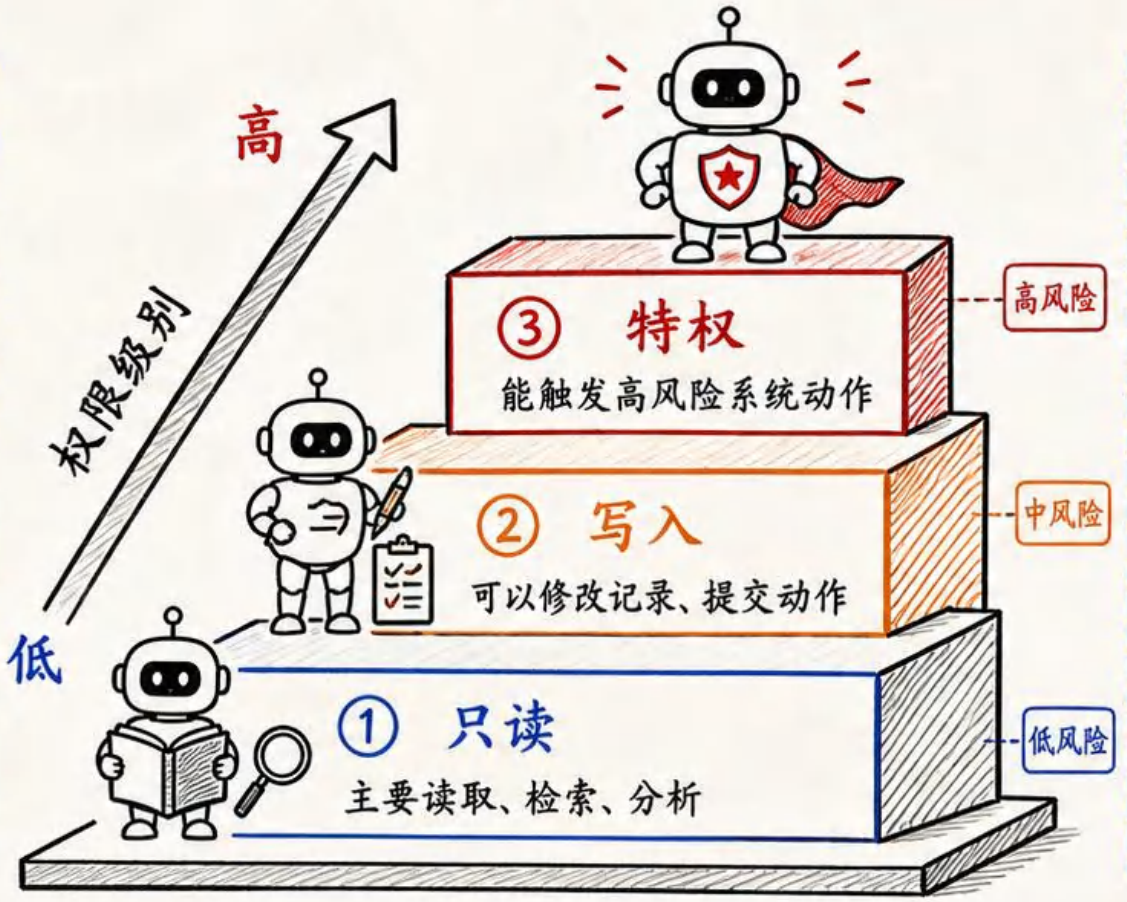
一个智能体能否进入任务链，至少要回答五个问题



3.1 权限面：它能动什么，风险有多大

权限面决定智能体能动什么，
以及风险有多大

- ① 只读智能体：
主要读取、检索、分析。
- ② 写入智能体：
可以修改记录、提交动作。
- ③ 特权智能体：
能触发高风险系统动作。
- ④ 权限越高，
审批、审计和回滚要求越强。



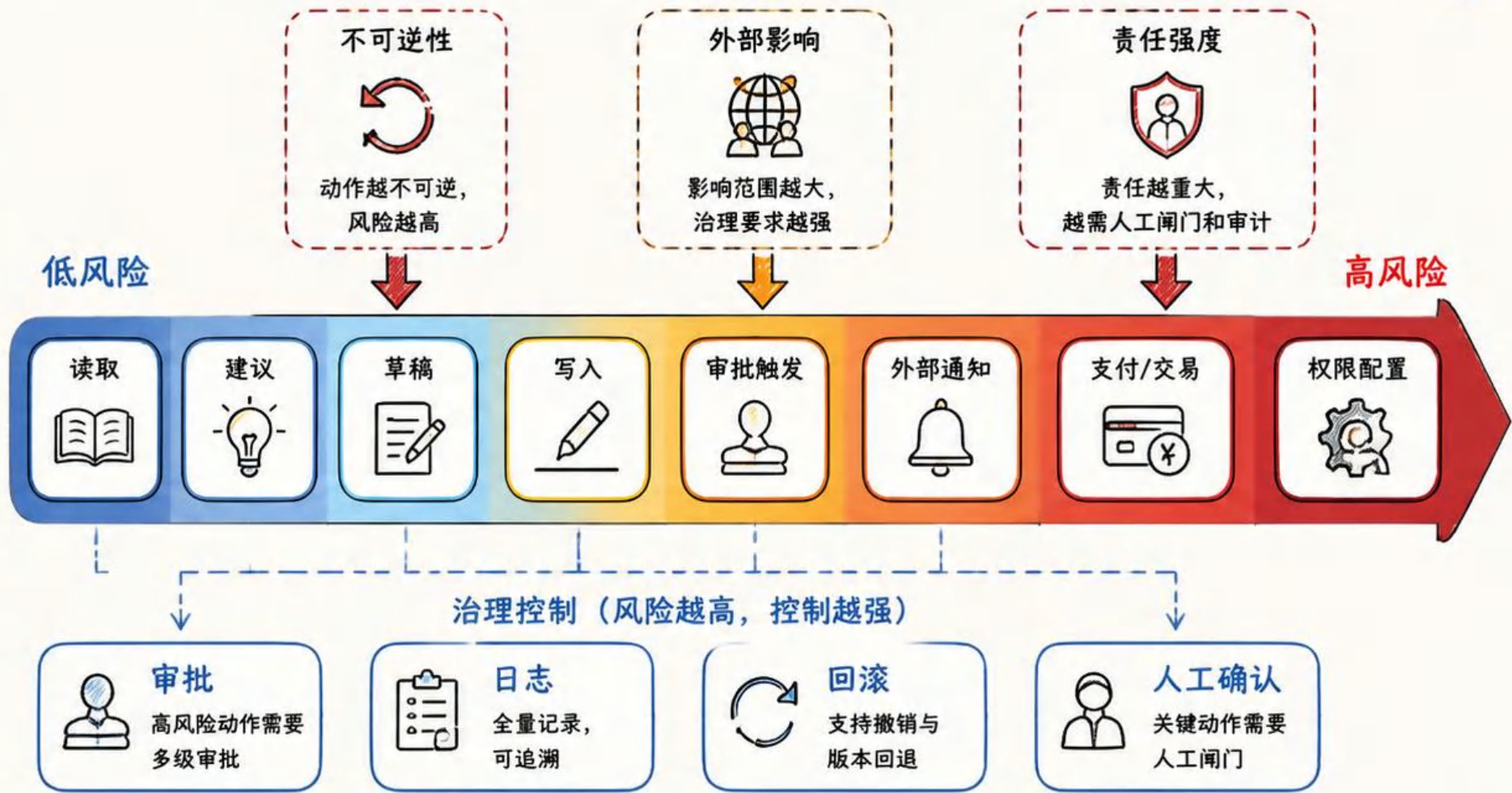
- 治理控制（随权限增强）
- 人工闸门
关键操作需人工确认
 - 回滚
更强的回滚能力
 - 审计
更完整的操作记录
 - 审批
需要更高权限审批

要求强度递增



风险等级会随动作不可逆性、外部影响和责任强度上升

-  动作越不可逆，风险越高。
-  外部影响越大，治理要求越强。
-  责任强度越高，越需要人工闸门和审计。
-  权限类型应按风险梯度管理，而不是一刀切授权。



智能体最好按系统或业务域拆分， 而不是设计成连接所有系统的万能智能体

① 企业系统天然分散在HR、财务、采购、销售、客服等业务域。



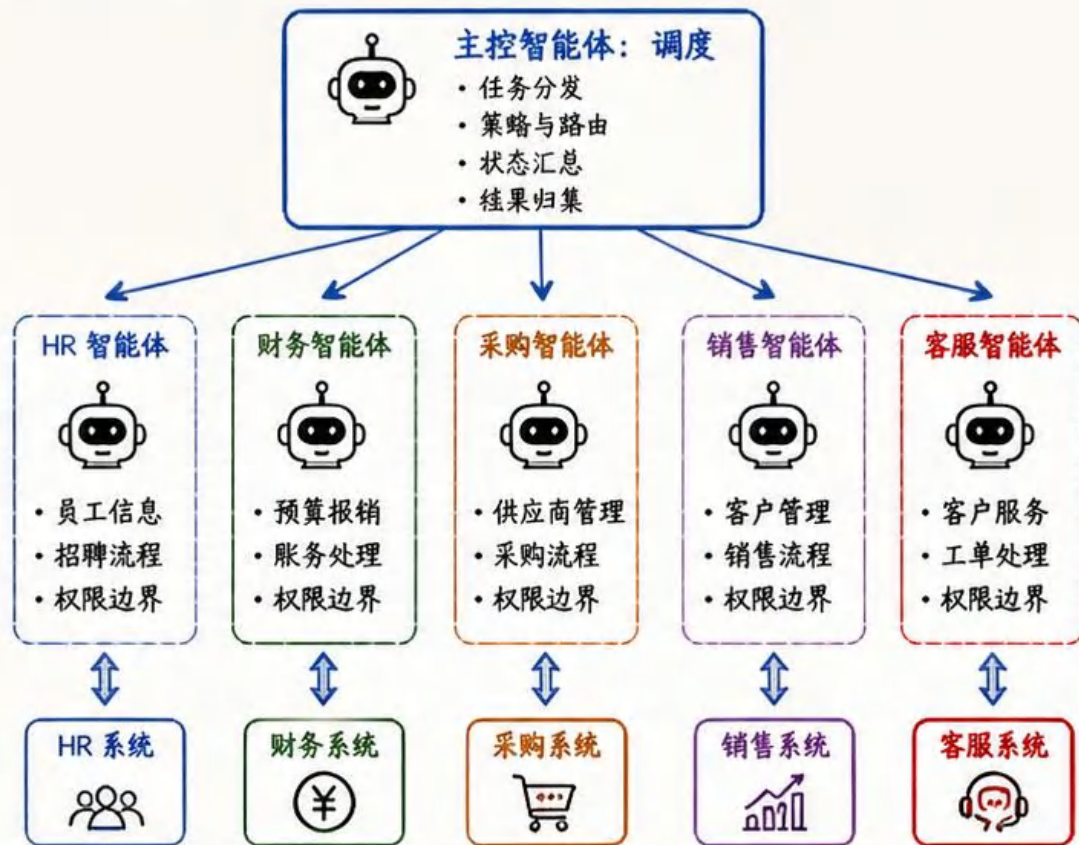
② 万能智能体会扩大权限面和责任模糊。



③ 按系统或业务域拆分，边界更清楚。



④ 主控智能体负责调度，不应持有所有系统权限。



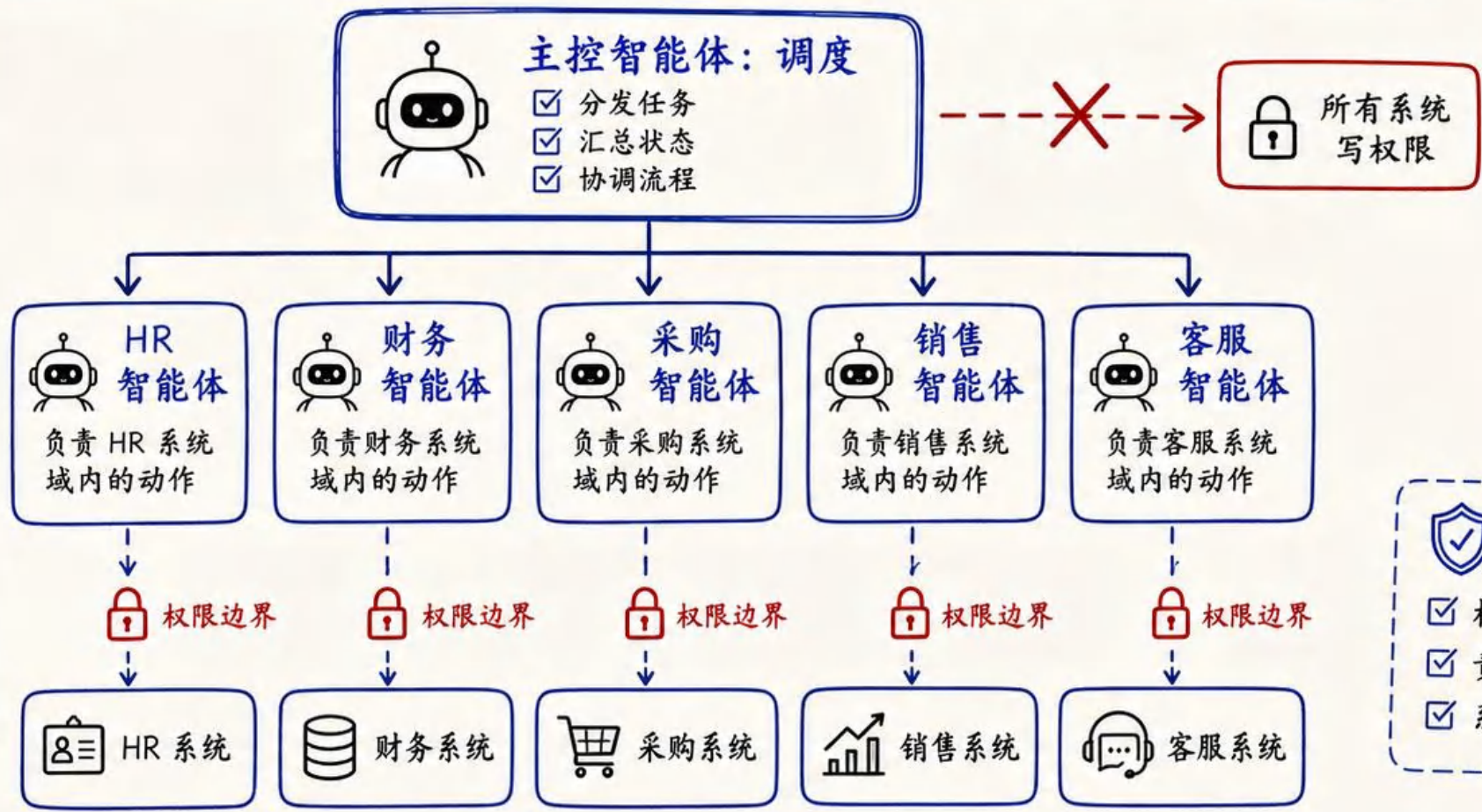
按系统或业务域拆分，边界清晰，可控、可审计、可追责



系统智能体越专，权限边界越清晰；
主控智能体负责调度，而不是持有所有系统权限

核心观点

- ① 主控智能体负责分发任务、汇总状态和协调流程。
- ② 系统智能体负责各自系统域内的动作。
- ③ 专业化降低权限集中和责任模糊。
- ④ 主控不应持有所有系统写权限。



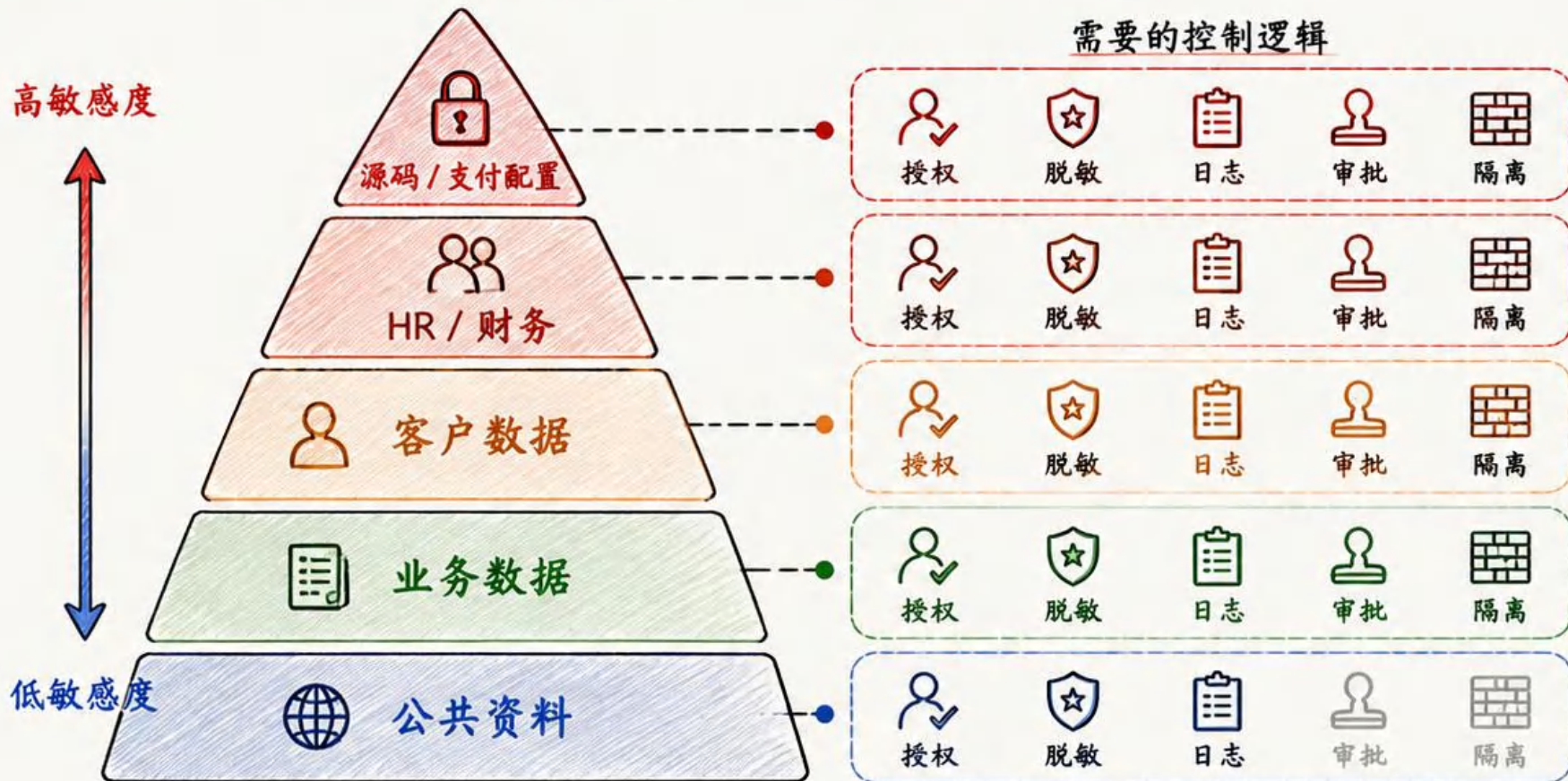
价值

- 权限边界清晰
- 责任链可追溯
- 系统安全可控

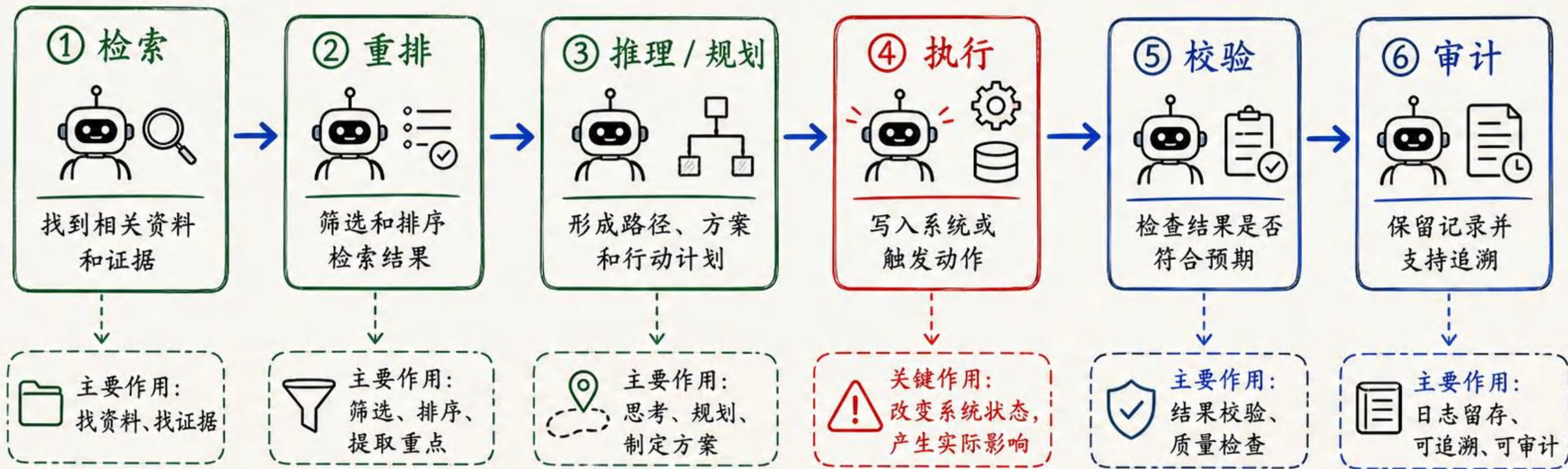


数据域决定智能体能接触哪类数据， 以及需要什么控制逻辑

- ☆ 公共资料：低敏感度，可用于检索和总结。
- ☆ 业务数据、客户数据：需要授权、脱敏和日志。
- ☆ HR、财务、源码、支付配置：需要更严格审批和隔离。
- ☆ 数据域越敏感，控制逻辑越强。



职责面决定智能体在任务链中负责什么



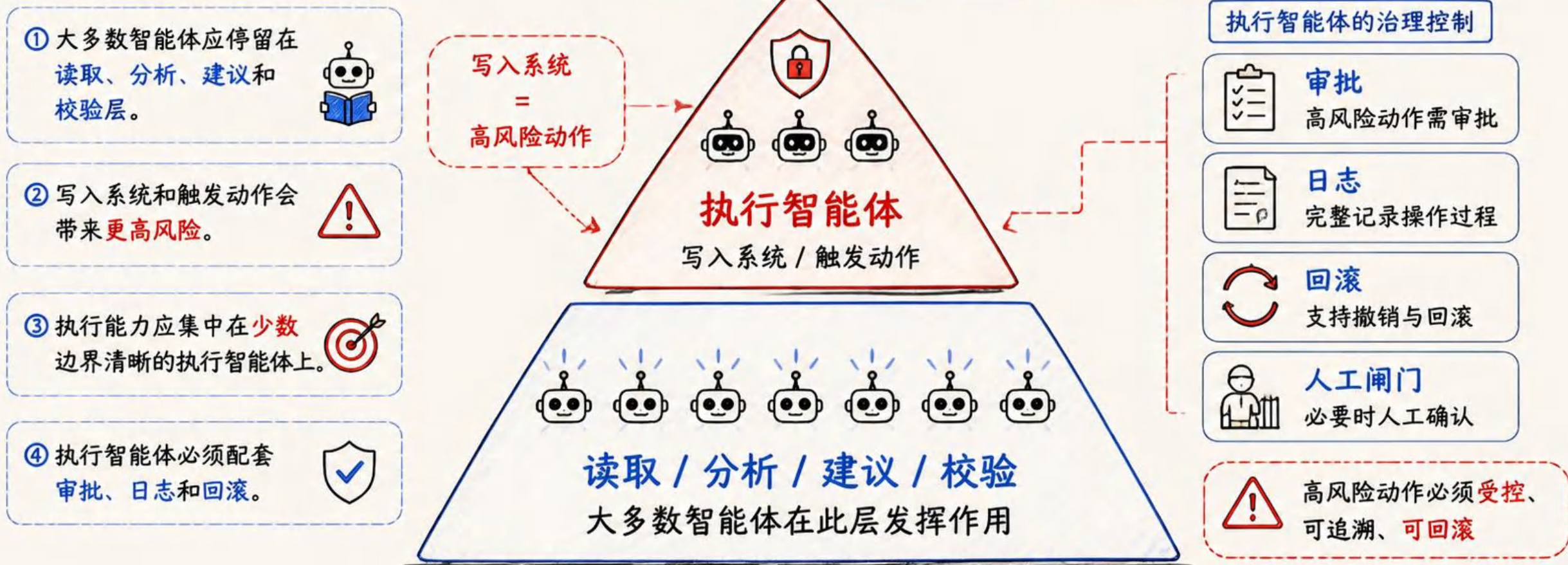
职责清晰，边界清楚，才能让智能体协作可控、可追溯、可信任。

执行改变系统，
校验保证质量，
审计留下证据。



3.4 职责面：它在任务链中负责什么

真正会写入系统、触发动作的能力 应集中在少数执行智能体上

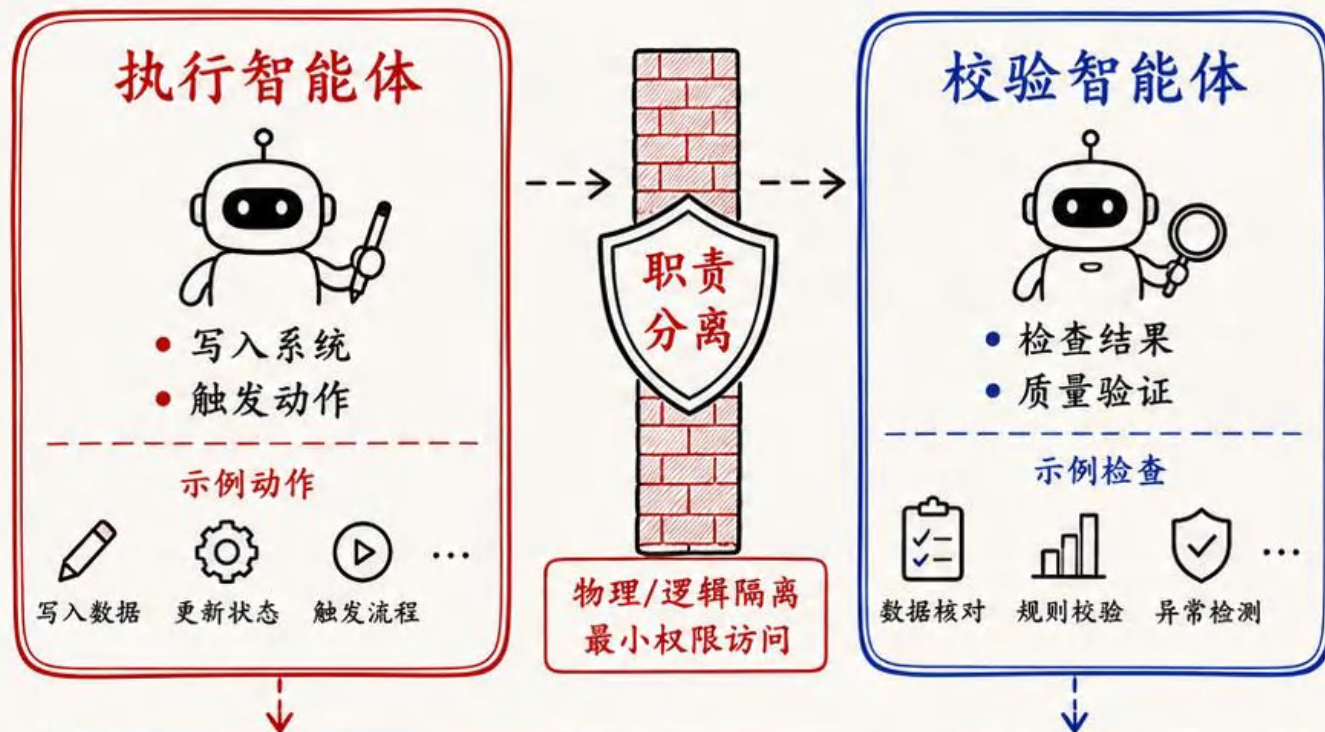


执行与校验必须分离， 执行智能体不应自己验证自己

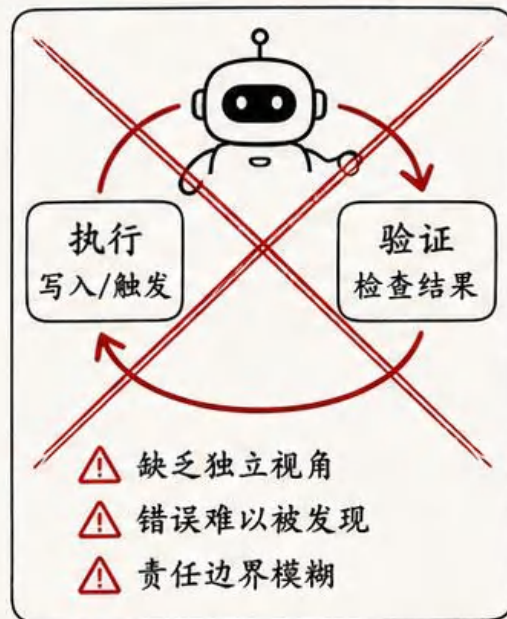
关键点

- ✓ 执行智能体负责写入系统或触发动作。
- ✓ 校验智能体负责检查结果是否符合预期。
- ✓ 自己执行、自己验证，会放大错误和责任盲区。
- ✓ 执行与校验分离，是企业级协作的基本控制点。

控制点
而非可选项



错误做法：自我执行与验证



审计日志：记录执行过程、校验结果、异常与处理，支持追溯与责任定位



编排面决定智能体在协作网络中扮演什么角色



① 主控：拆解目标、分发任务。



② 路由：把任务送到合适能力节点。



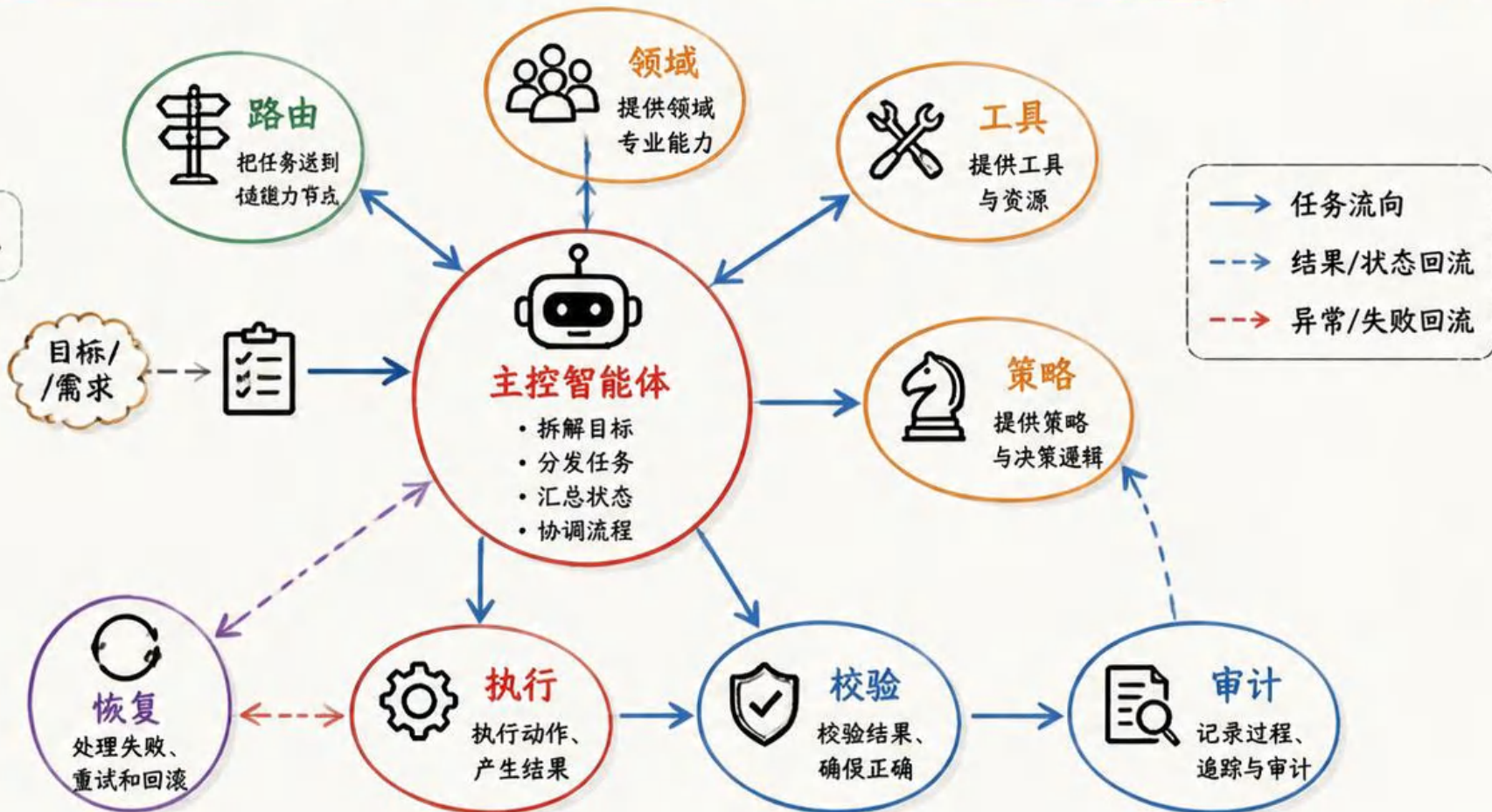
③ 领域 / 工具 / 策略：
提供专业能力。



④ 执行 / 校验 / 审计：
完成动作并控制风险。



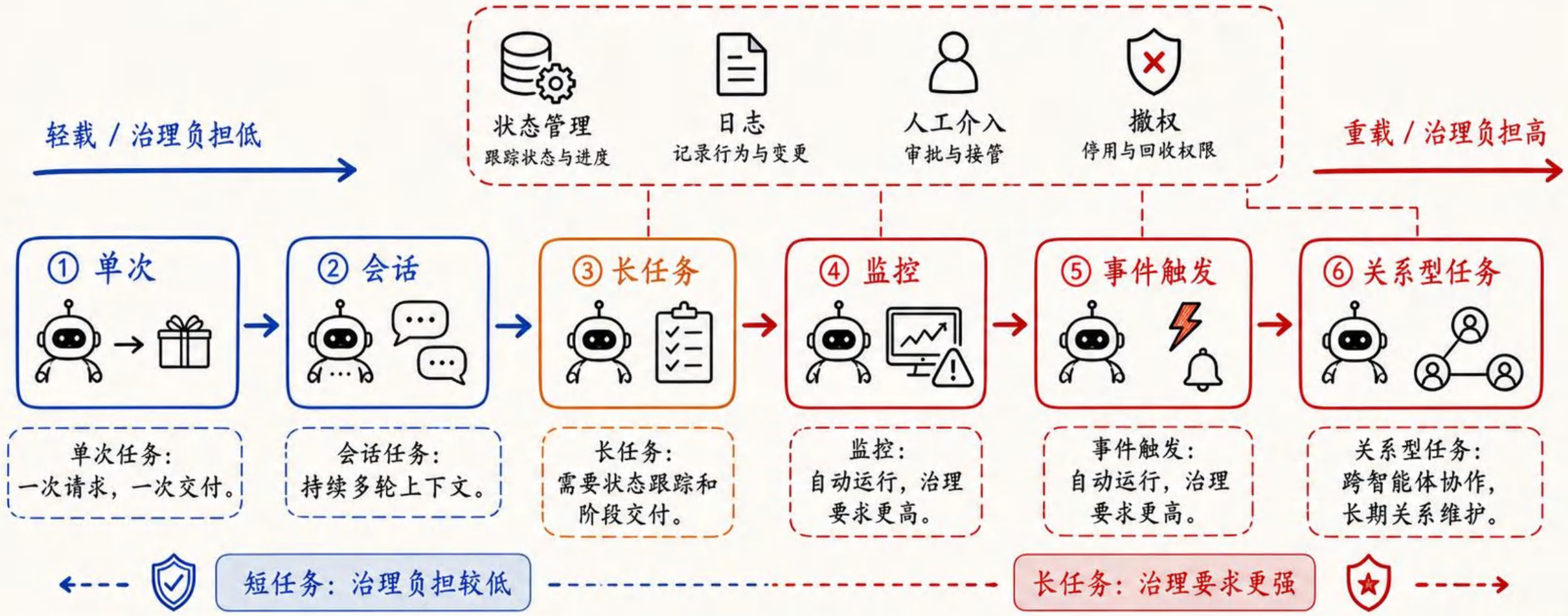
⑤ 恢复：处理失败、
重试和回滚。



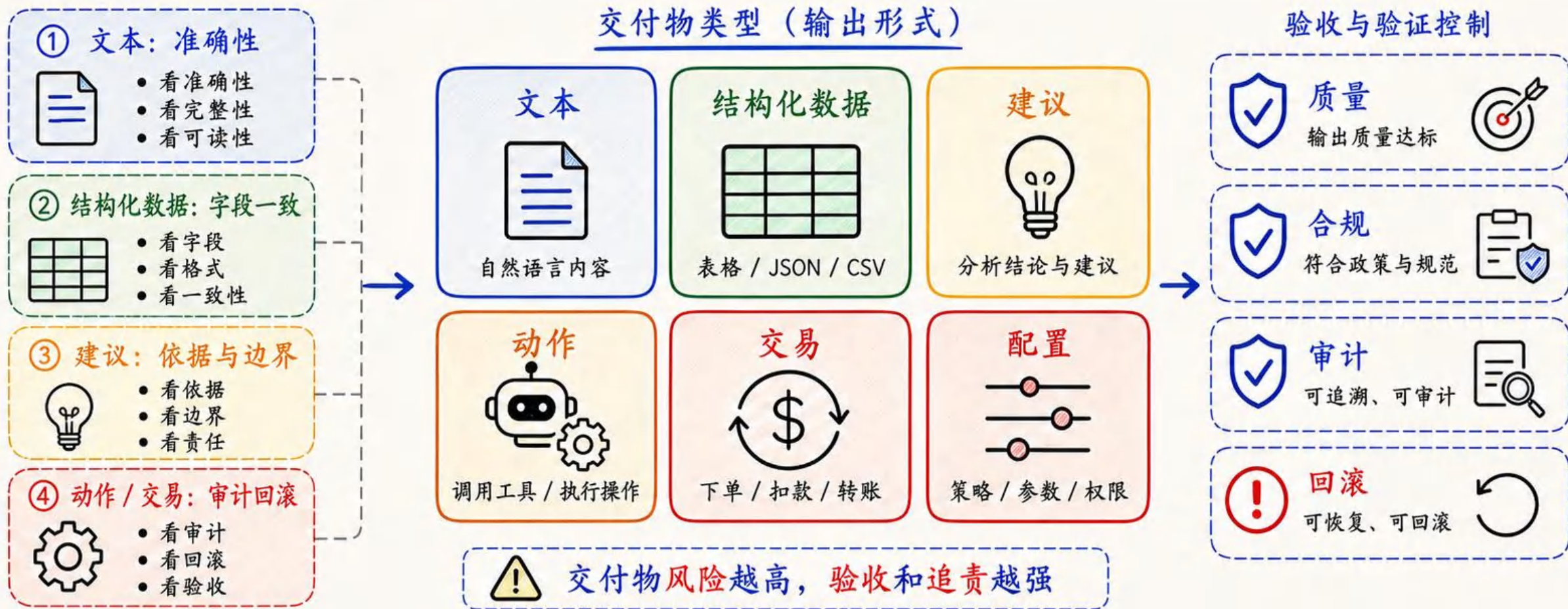
来源: <https://arxiv.org/abs/2507.21105>; <https://arxiv.org/abs/2506.12508>; <https://arxiv.org/abs/2604.01020>; <https://arxiv.org/abs/2404.02183>



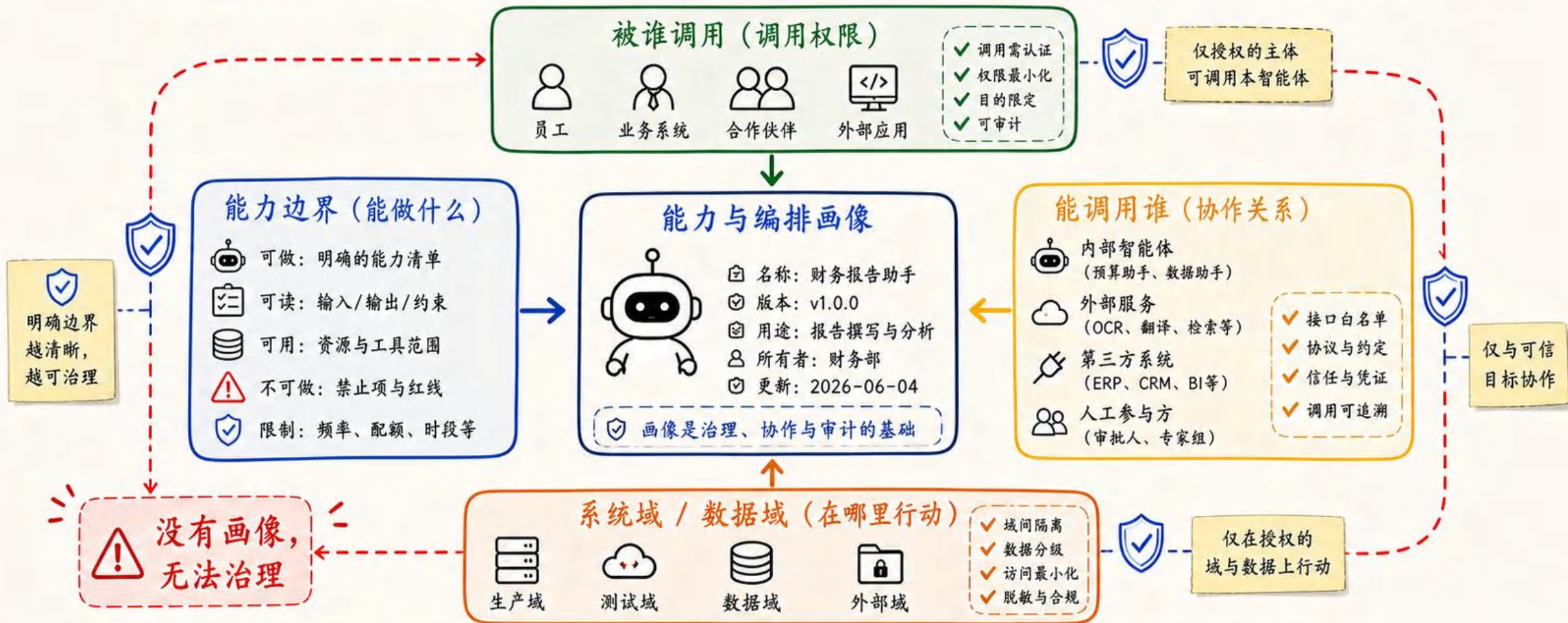
运行面决定智能体是一次性任务，还是长期运行任务



交付面决定智能体最终交付什么，以及如何验收



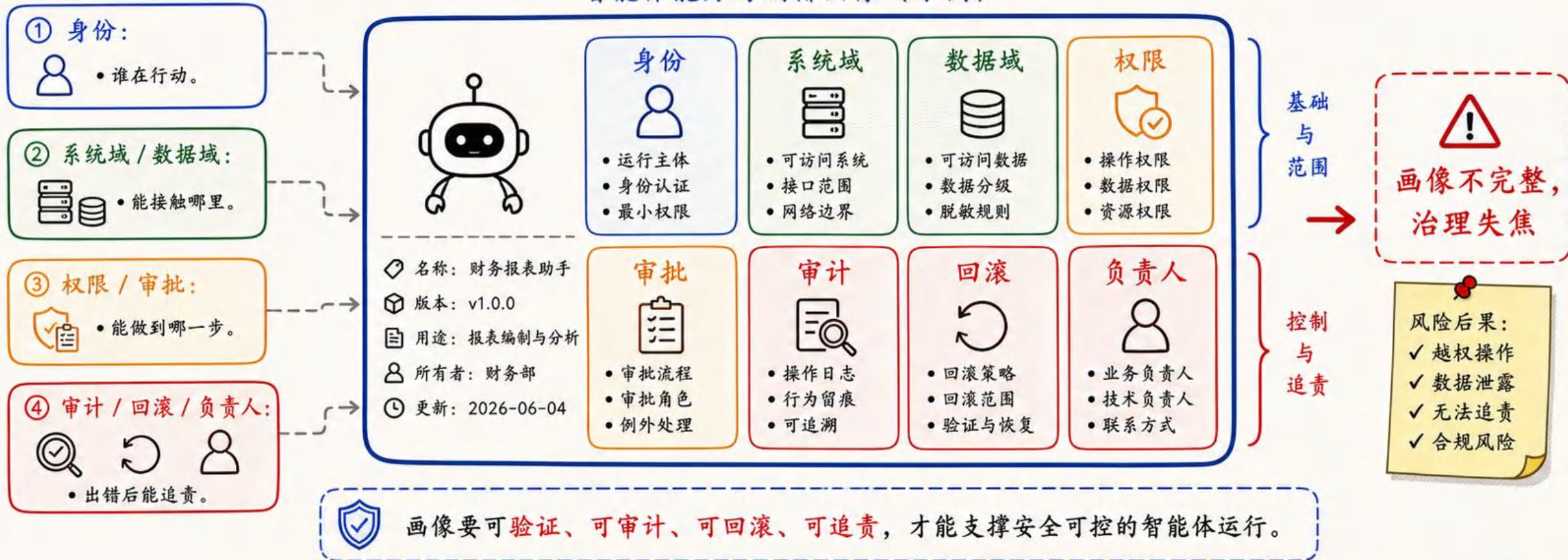
智能体画像不仅描述能做什么，还描述被谁调用、能调用谁、在哪些边界内行动



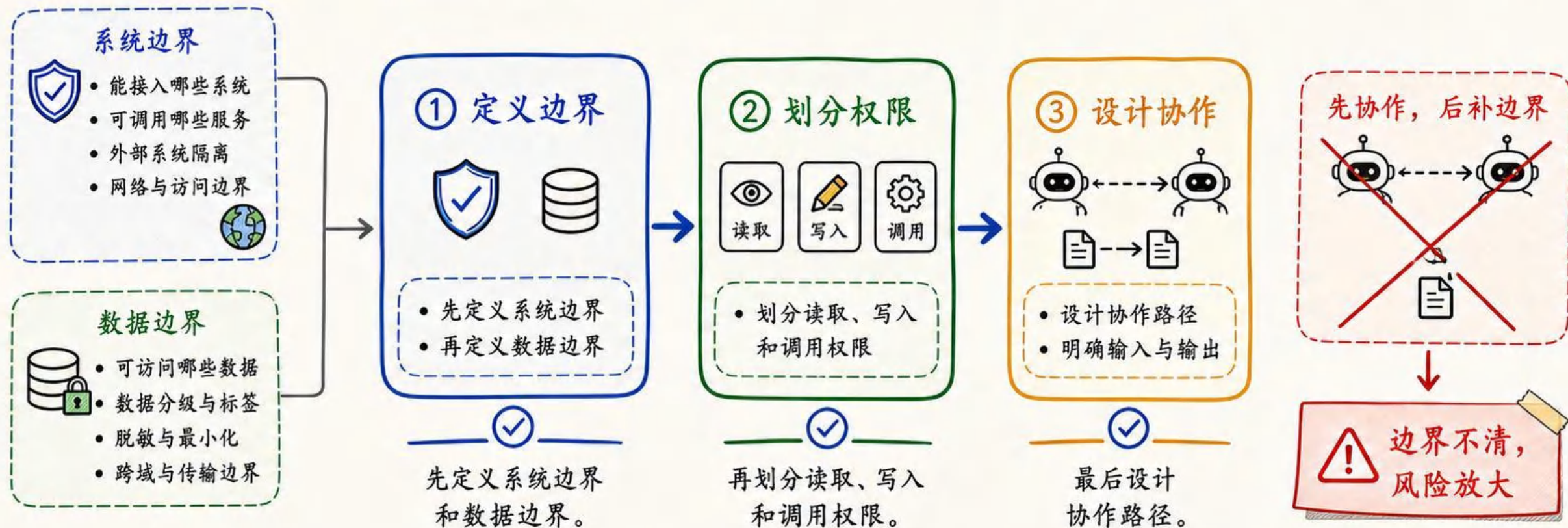
智能体能力与编排画像必须覆盖身份、系统域、数据域、权限、审批、审计、回滚和负责人

画像治理要覆盖的关键点

智能体能力与编排画像（示例）



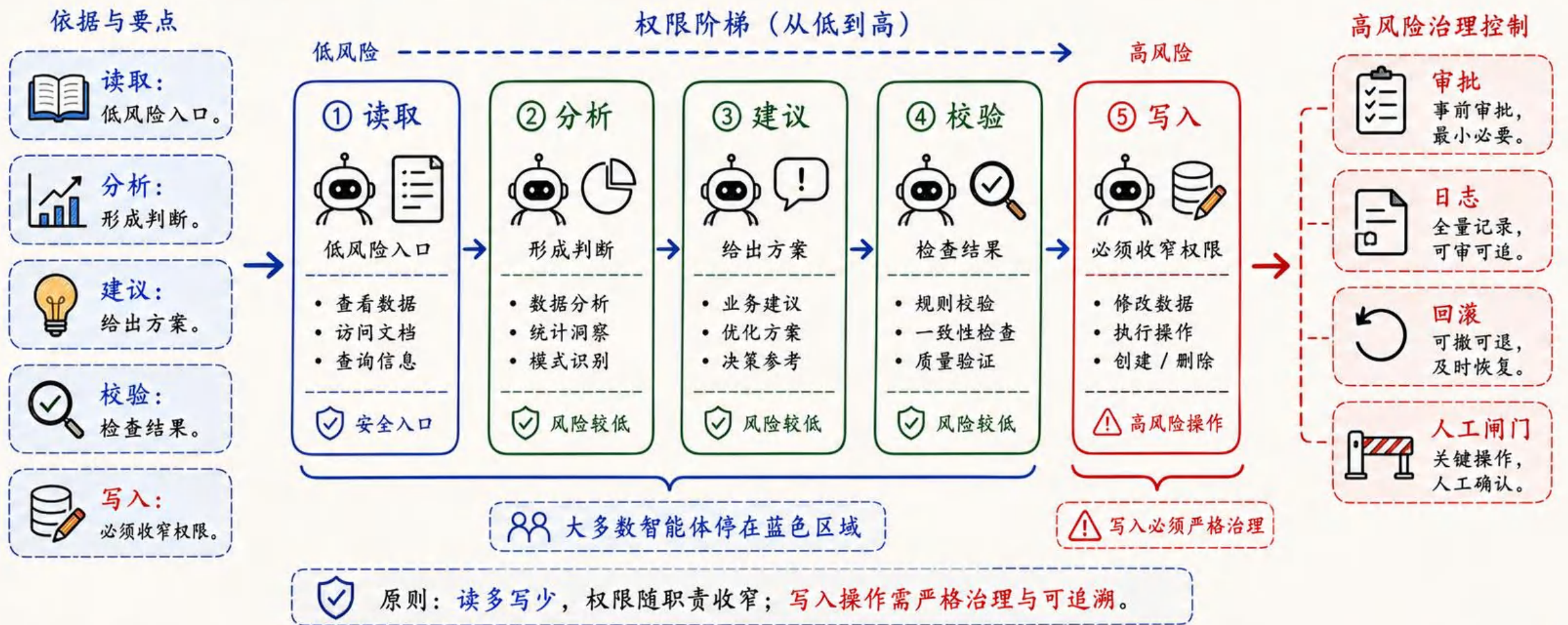
1. 先定义边界，再设计协作



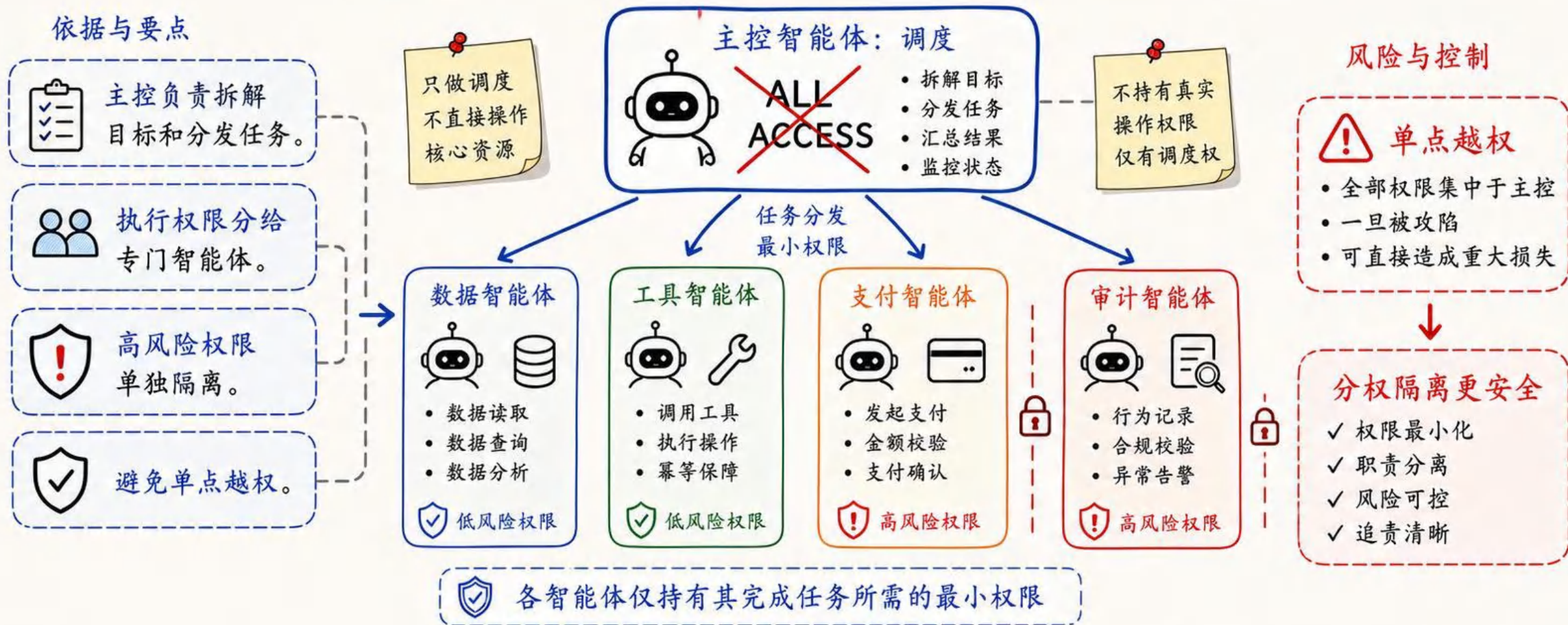
原则：先定义边界，再划分权限，最后设计协作；边界不清，协作越多风险越大。



读多写少，大多数智能体应停留在读取、分析、建议和校验层



4. 主控智能体不应持有全部权限



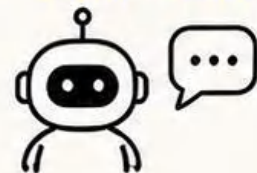
5. 高风险动作必须有**人工闸门**

依据与要点

- ① 支付、删除、外发：
 - 需要人工确认。
- ② 高影响动作：
 - 必须可暂停。
- ③ 审批记录：
 - 进入审计链。
- ④ 人工闸门：
 - 是风险保底。

建议与依据
上下文信息

智能体建议



- 任务建议
- 风险等级评估
- 方案与影响说明

审批

- 身份校验
- 权限校验
- 多级审批

暂停

- 可随时暂停
- 阻断执行

人工闸门



- 人工判断
- 权限校验
- 签字确认

记录

- 审批记录
- 操作日志
- 证据留存

回滚

- 可撤回
- 可回滚
- 可恢复

高风险动作

支付

删除

外发

影响大、不可逆、
涉及数据与资金安全

审计链

- 全量记录
- 不可篡改
- 可追溯



风险放大!

- 审批缺失
- 日志缺失
- 无法回滚

必须通过人工闸门
才能执行。



原则：高风险动作必须有**人工闸门**，确保可控、可追溯、可恢复。



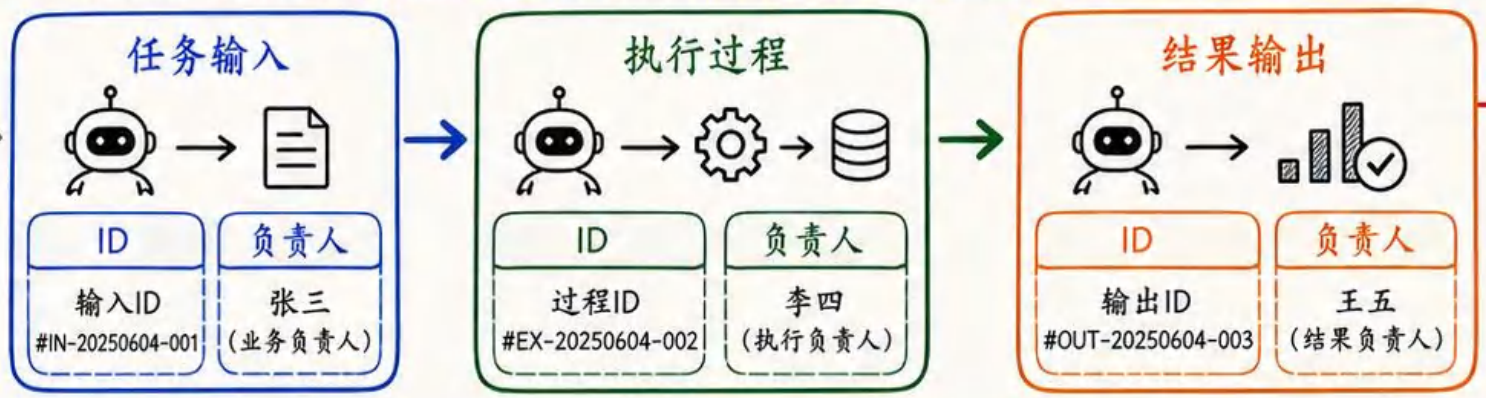
每条任务链都要可回放，每个智能体都要有身份和负责人

依据与要点

- ① 任务输入可追踪。
 - 来源清晰
 - 参数可查
 - 元数据完整
- ② 执行过程可回放。
 - 步骤记录
 - 调用链路
 - 状态快照
- ③ 结果输出可核验。
 - 输出可对比
 - 结果可验证
 - 证据可留存
- ④ 每个节点都有身份和负责人。
 - 身份标识
 - 负责人可追责

审计目标：完整、可追溯、可回放、可核验、可追责

🕒 时间戳 — 指纹 身份标识 — 人 负责人 — 文档 版本与签名



禁止：匿名与无责

匿名智能体

- 无身份
- 无负责人
- 无法追责

审计日志 (不可篡改)

- 全量记录
- 不可更改
- 可追溯
- 可导出

审计证据链 (示例)

🕒 时间戳	🏷️ 环节	ID 节点ID	🤖 智能体ID	👤 负责人	📋 操作/结果	🛡️ 签名/哈希
2026-06-04 09:15:32	输入	#IN-20250604-001	BOT-IN-01	张三	接收请求, 参数校验通过	a1b2...9f7e
2026-06-04 09:15:45	过程	#EX-20250604-002	BOT-EX-07	李四	调用工具X, 执行完成	d4c3...8a2b
2026-06-04 09:16:08	输出	#OUT-20250604-003	BOT-OUT-03	王五	生成报告, 校验通过	7c9e...2d5a

身份与负责人是风险可控的基础条件!

原则：可追踪、可回放、可核验、可追责，确保每一步都可查、可证、可问责。



商业链路将从人搜索、人比较、人下单， 转向智能体搜索、比较、委托、下单和支付

旧链路：人主导

人搜索



查找信息

人比较



对比选择

人下单



点击下单



旧链路：人搜索、人比较、人下单。



问题：信息分散、决策耗时、体验割裂。



局限：需人工反复操作，易受主观影响。

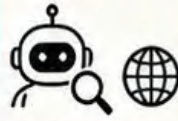
关键变化：



交易从点击
转向委托。

新链路：智能体主导

智能体搜索



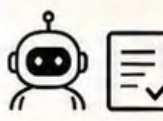
• 全域检索
• 信息聚合

比较



• 多维对比
• 智能筛选

委托



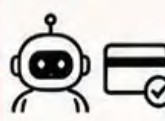
• 明确意图
• 委托执行

下单



• 执行下单
• 确认条款

支付



• 选择方式
• 完成支付



新链路：智能体搜索、比较、委托。



优势：自动化、个性化、连贯性更强。



价值：提升效率，减少人为错误与操作成本。

治理重点



身份

• 身份认证
• 实体溯源



授权

• 权限分级
• 最小授权



支付

• 支付凭证
• 风险控制



追责

• 操作留痕
• 责任可追



治理原则：先定义边界，再授权协作；全程留痕，可审计、可回放、可追责。

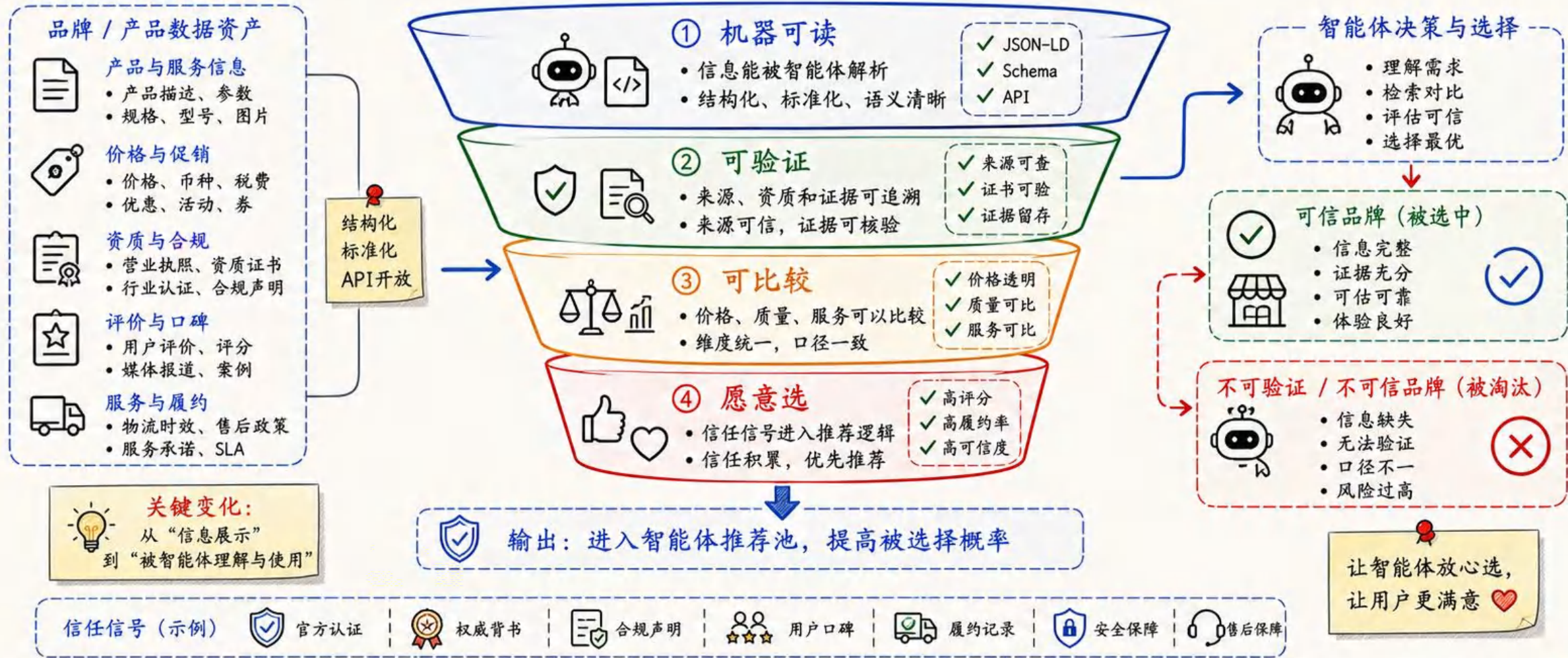


智能体支付让支付从结账动作变成委托控制层

智能体支付委托控制层（四道防线）



智能体可见性要求品牌让智能体看得见、信得过、愿意选



智能体下单让下单从人工点击变成自动化决策链



企业对企业采购可能更早进入 A2A 协作

场景：企业采购从需求到结算的全链路自动化协作

采购方智能体



采购智能体

- 提出需求
- 选择策略
- 监督执行
- 验收结果

目标：合规、高效、降本

① 询价



- 采购智能体发起需求
- 向多家供应商询价

输出：询价单

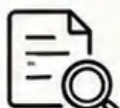
② 预算校验



- 财条规则自动检查
- 预算余额与科目校验

输出：预算通过

③ 合同初审



- 条款合规检查
- 风险与条款预筛

输出：预算通过

④ 下单



- 生成采购订单
- 锁定价格与交付条件

输出：PO 已生成

⑤ 付款



- 按审批支付
- 支付凭证生成
- 资金安全控制

输出：支付完成

⑥ 对账



- 交付与发票核对
- 三单匹配
- 差异处理

输出：对账完成

供应商智能体



供应商智能体

- 应答询价
- 提供报价
- 确认订单
- 开具发票
- 发货交付

目标：及时、准确、合规



合规与权限控制

- 角色/权限隔离
- 最小权限原则
- 敏感操作审批



合同风险控制

- 黑名单校验
- 高风险条款预警
- 必备条款检查



支付风险控制

- 收款主体校验
- 支付限额控制
- 异常交易拦截



差异与异常处理

- 差异自动识别
- 异常工单自动生成
- 闭环处理跟踪



审计与追溯

- 全链路日志留存
- 关键操作不可篡改
- 可追溯、可审计、可追责



询价记录



预算记录



合同记录



订单记录



支付凭证



对账记录



审批与签署

- 多级审批流程
- 电子签章存证
- 合规留痕



价值：提升效率，降低采购成本与风险，确保合规可控，形成数据驱动的采购能力。



企业对企业销售要面向客户采购智能体

目标：让客户采购智能体理解、信任并优先选择我方方案

客户采购智能体



- 理解需求
- 评估供应商
- 自动决策
- 执行采购



需求与目标

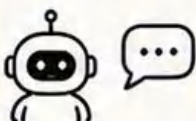
- 业务目标与场景
- 技术与服务要求
- 交付与时间窗口

预算与规则

- 预算范围与周期
- 审批规则与流程
- 采购政策与合规

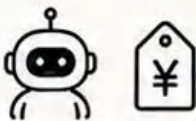
需求发布

① 需求



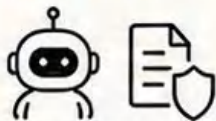
- 需求结构化
- 规格与数量
- 交付与时间

② 报价



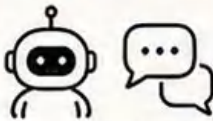
- 产品与价格
- 服务与SLA
- 有效期说明

③ 资格判断



- 资质与认证
- 成功案例
- 有规与审查

④ 协商



- 商务条款
- 交付与验收
- 责任与赔付

达成一致

反馈 / 修改 / 补充信息

风险与控制（合同与支付相关）



合同风险

条款审查
风险预警



支付风险

分级授权
限额控制



履约风险

里程碑管理
交付验收



争议处理

证据留存
责任划分

审计链（全程可追溯）



需求/招标
记录



报价/响应
记录



资格/证据
记录



合同/签署
记录



下单/支付
记录



交付/验收
记录

企业销售智能体



- 发布可被理解的信息
- 响应需求与招标
- 提供可信证据
- 完成合同与交付

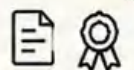


销售包（机器可读）



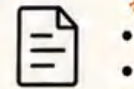
- 产品规格
- 价格与计价方式
- 提务等级与保障

资格与证据



- 公司资质与认证
- 成功案例与评价
- 合规文件与声明

提案与合同



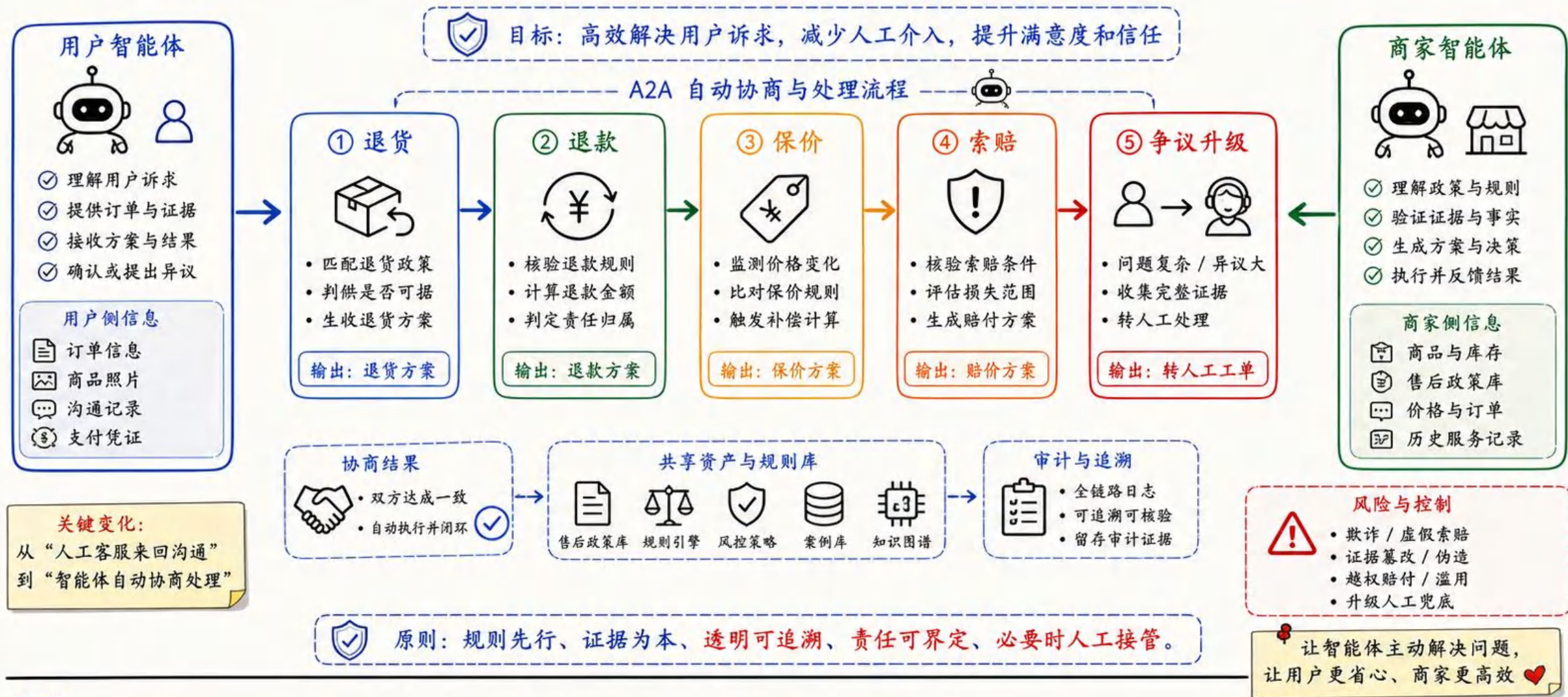
- 需求建议书/响应书
- 合同条款与附件
- 交付计划与SLA

原则：

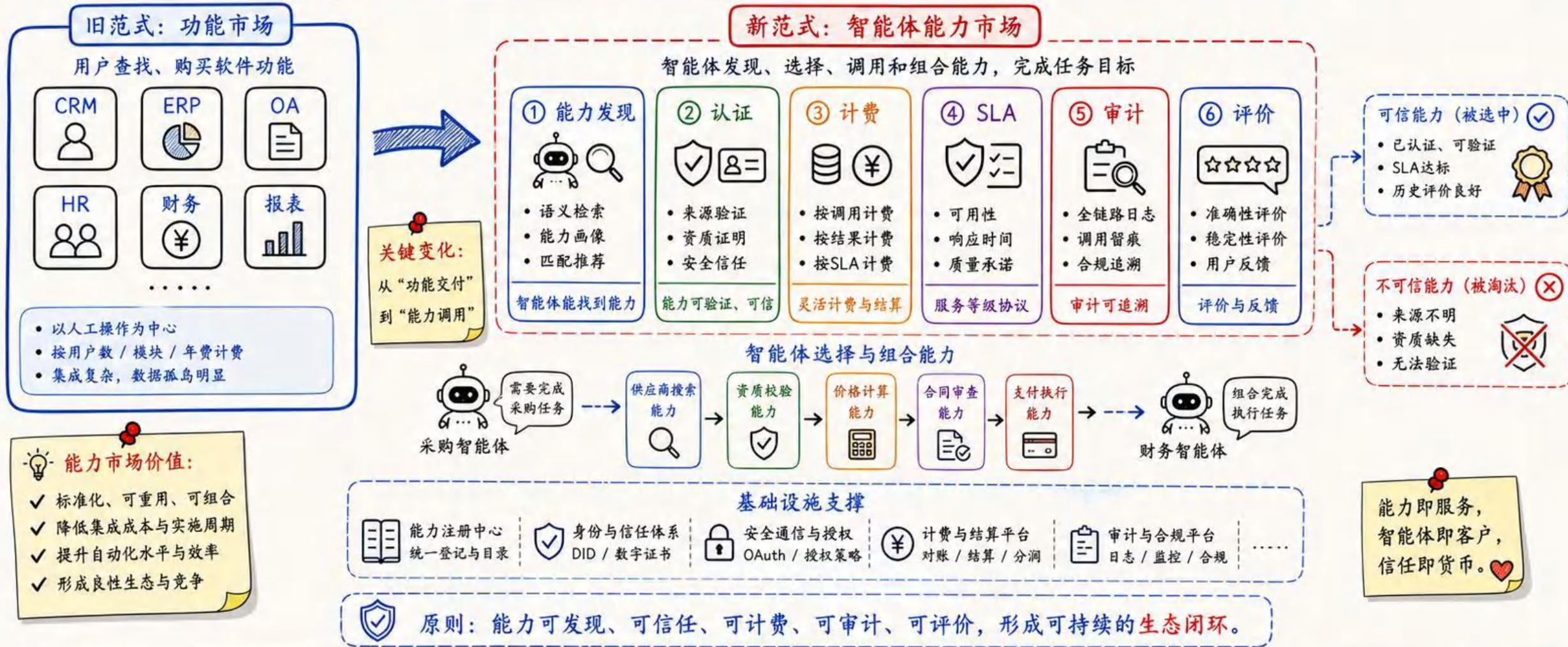
信息可理解、证据可验证、
流程可追溯、责任可追责



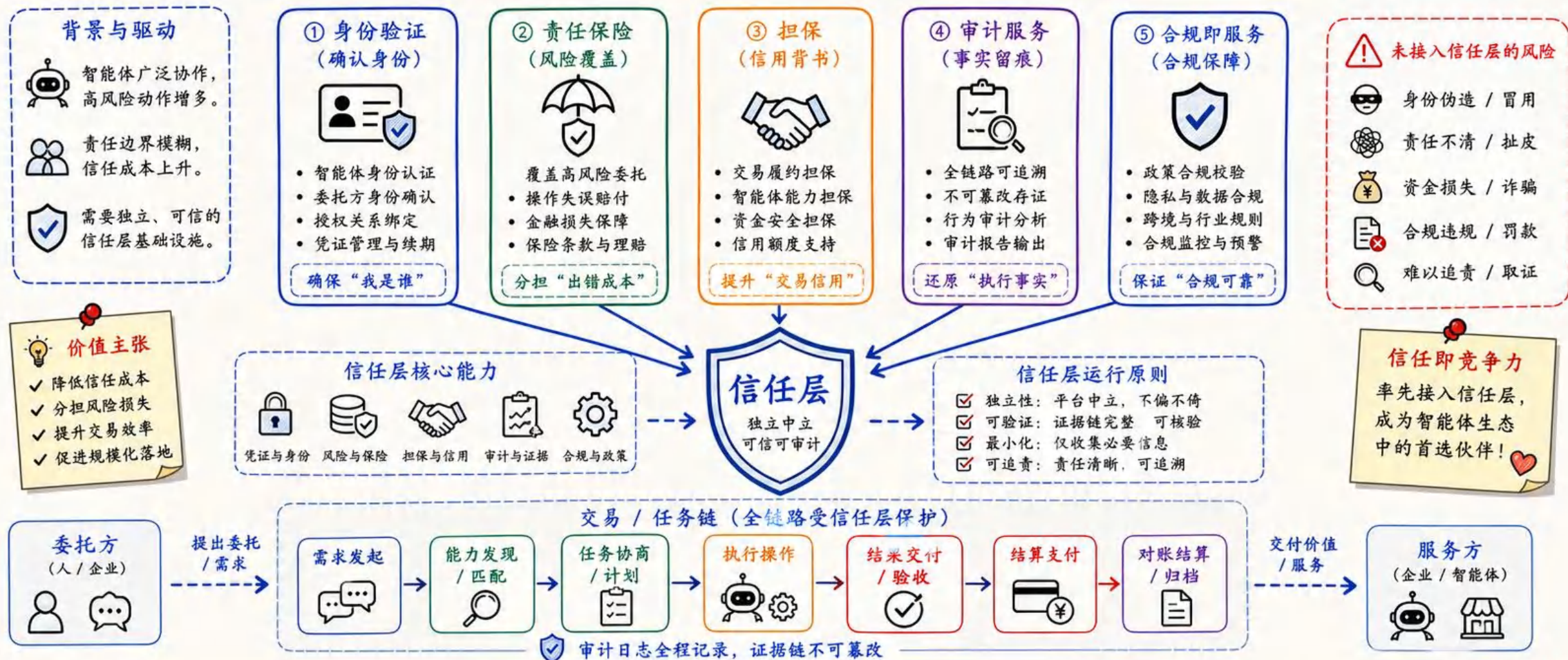
用户智能体与商家智能体可能自动协商退货、退款、保价、索赔和争议升级



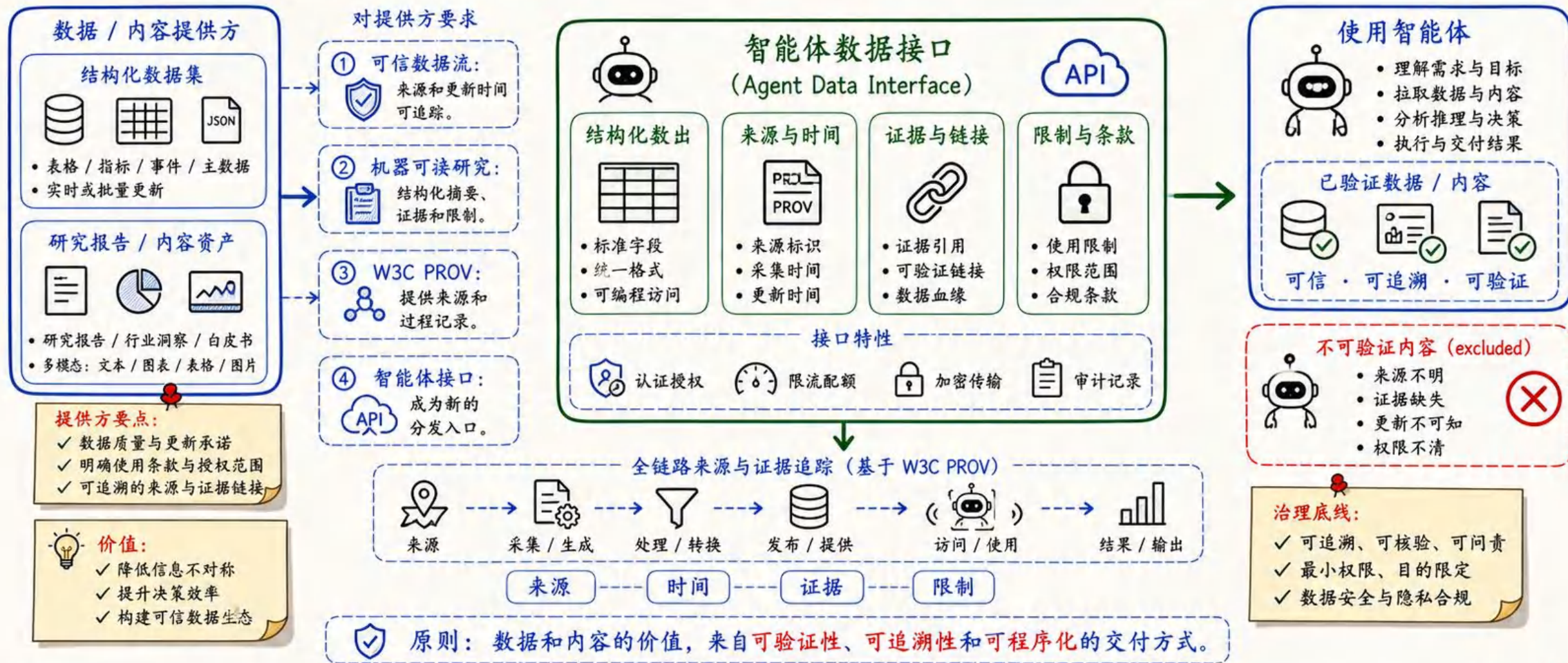
企业软件可能从功能市场走向智能体能力市场



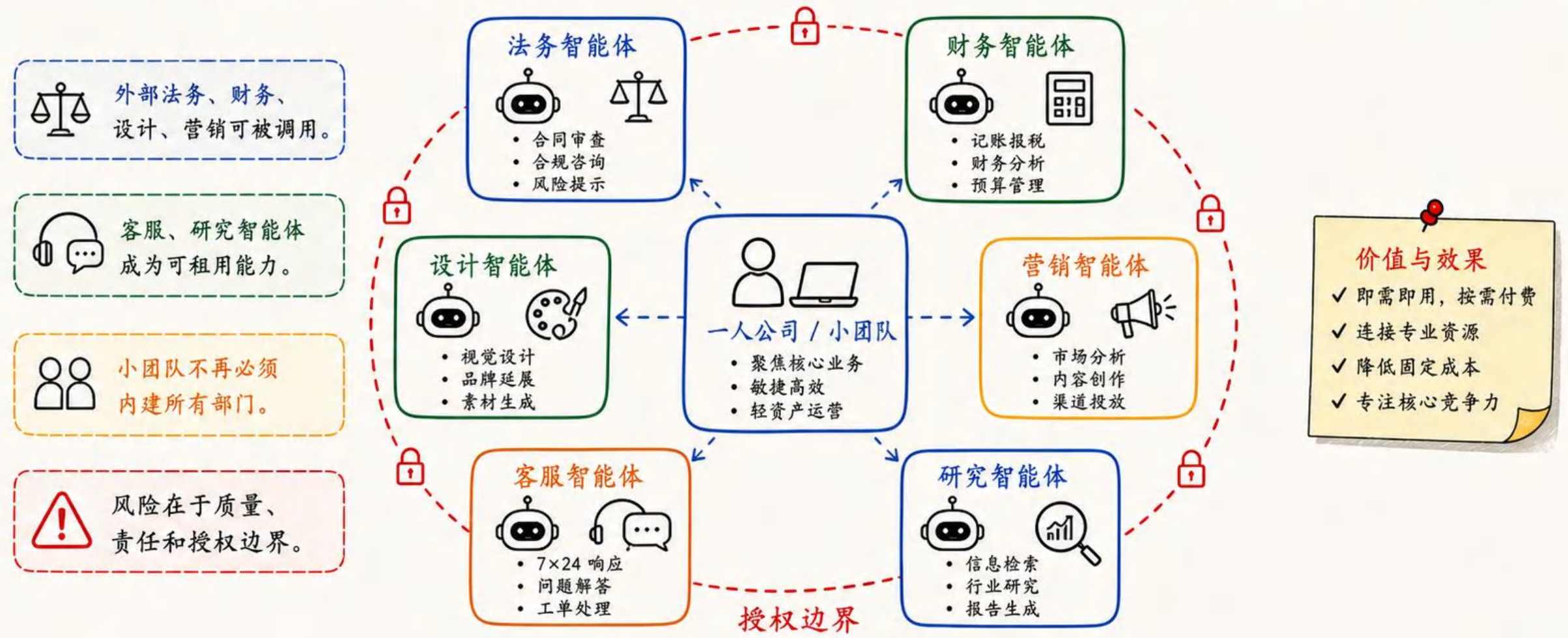
智能体 Trust / 保险可能形成新的信任层商业模式



数据和内容将不仅卖给人，也可能通过智能体数据接口提供给智能体



一人公司和小团队可以通过智能体能力外包获得大组织能力



商业主题可优先关注智能体可见性、授权委托、能力市场、A2A 销售、信任保险和能力外包

优先级原则

商业主题优先级阶梯（从基础到放大）

路径总结

- ① 先做可见性和授权委托。
- ② 再扩展能力市场和 A2A 销售。
- ③ 信任保险支撑高风险交易。
- ④ 能力外包释放小团队杠杆。

价值与影响力
逐步放大



风险与复杂度逐步提升

先基础设施，
后商业放大。

风险提示

- 跳过基础直接放大风险极高。
- 治理滞后将导致信任与合规危机。

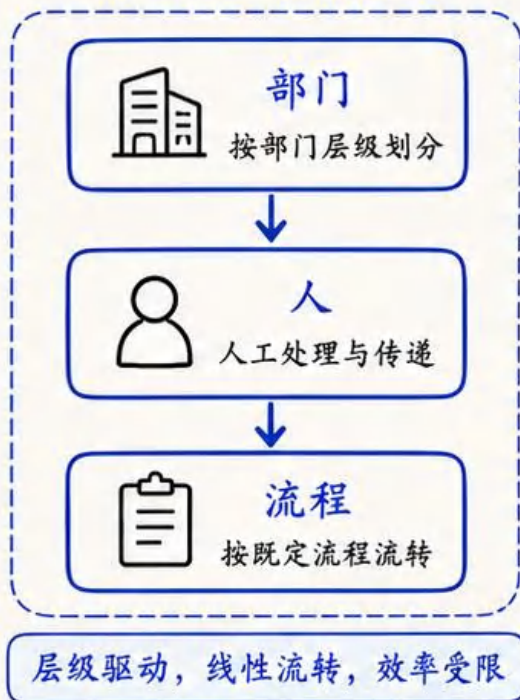


原则：循序渐进，稳步构建，确保可控、可信、可追溯。



A2A 对组织结构的真正影响， 是改变任务在组织内部流转的方式

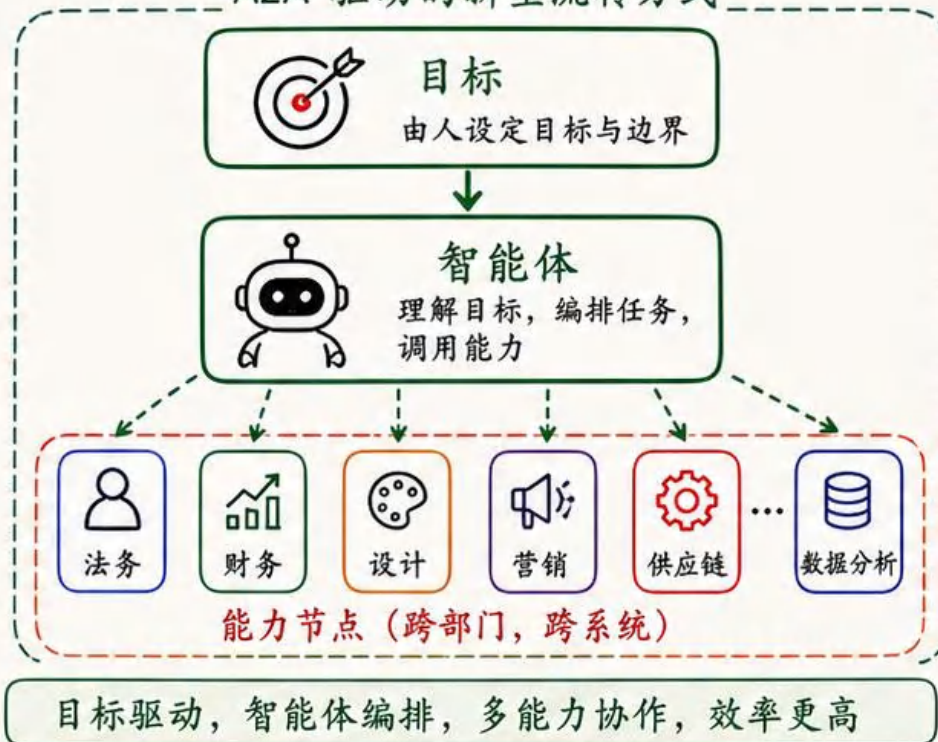
传统组织流转方式



从层级流转

到能力协作

A2A 驱动的新型流转方式



关键影响点

① 任务不再只沿部门层级流转。

智能体会调用跨部门能力。

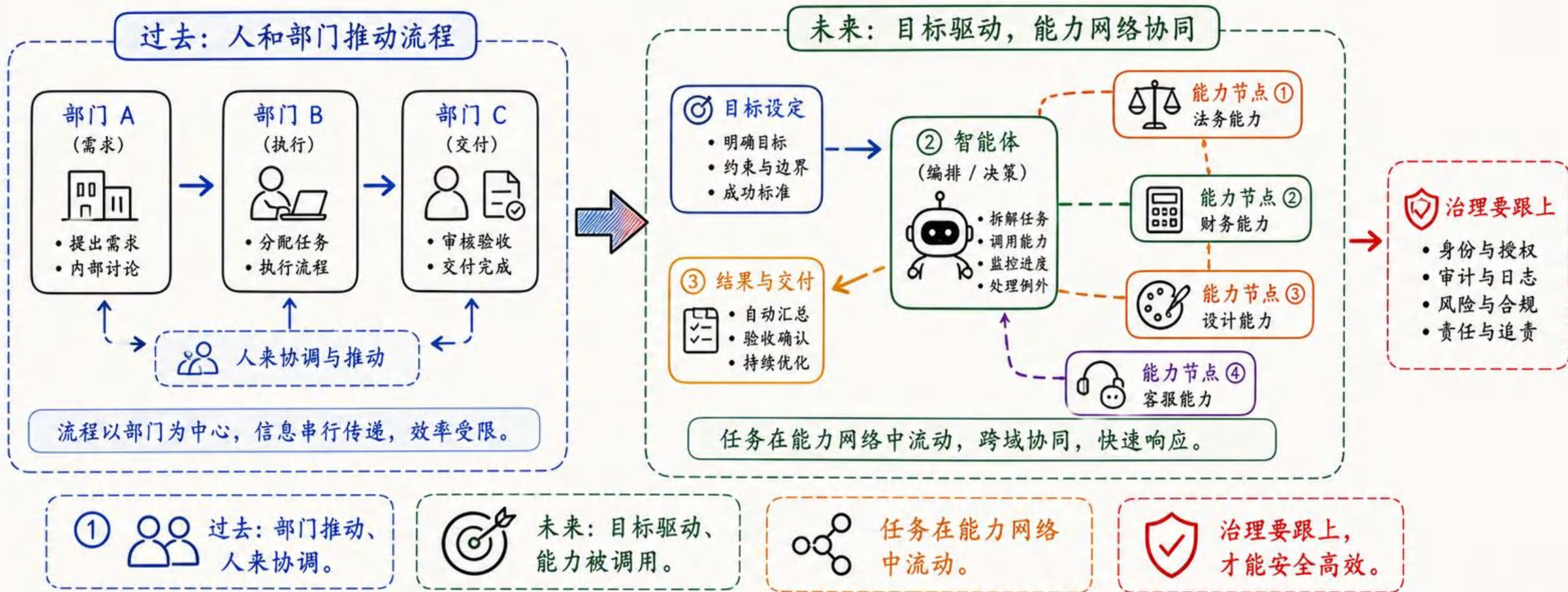
人更多设定目标和处理例外。

治理层需要同步增强。

组织的竞争力，将来自“目标定义力 + 能力编排力 + 治理控制力”。



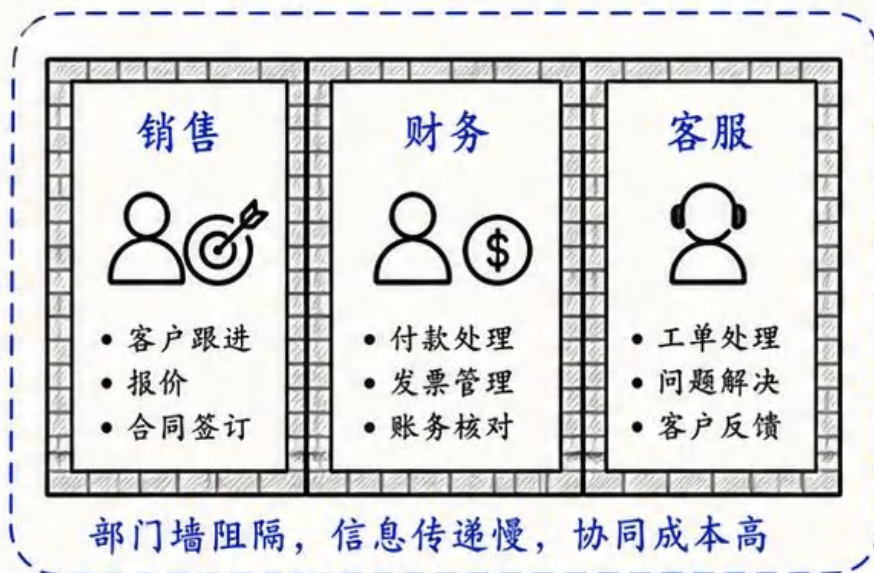
企业会从人和部门推动流程，走向人设定目标、智能体调用能力、任务在能力网络中流转



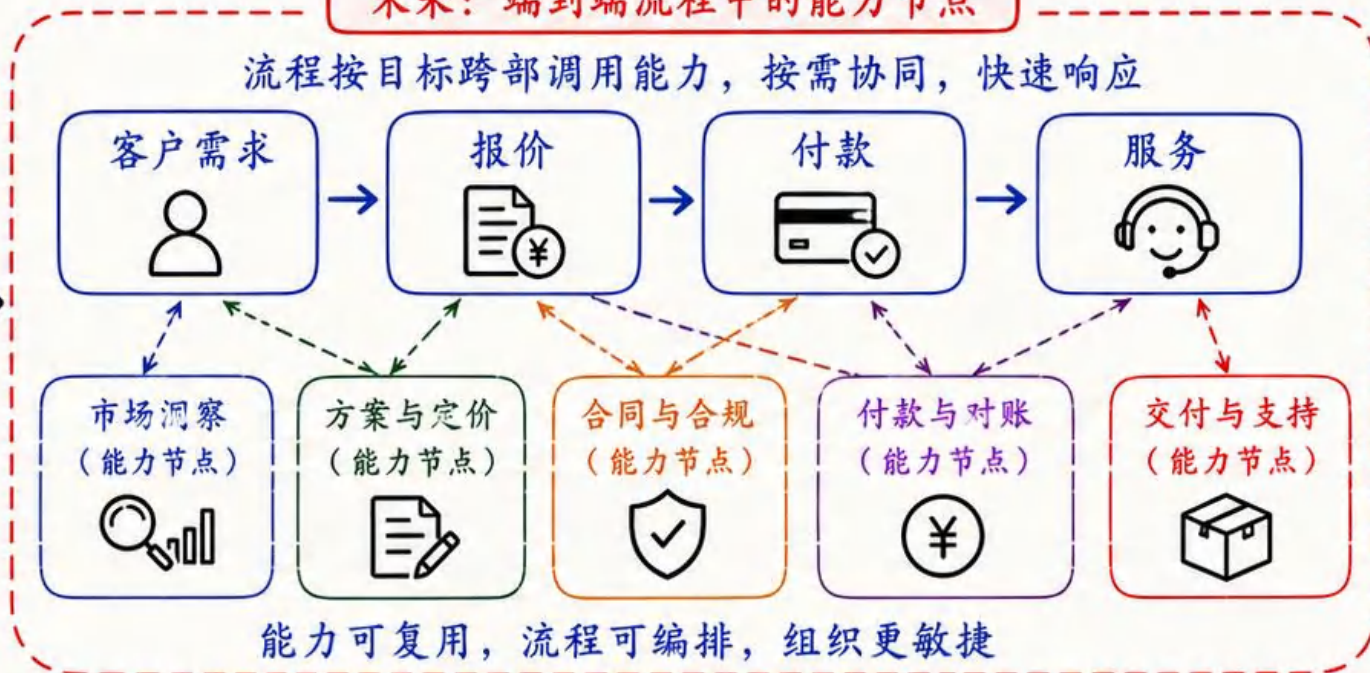
部门不再只是封闭职能单元， 而会变成端到端流程中的能力节点



过去：封闭职能单元



未来：端到端流程中的能力节点



① 部门仍存在，但边界变薄。
从“封闭职能”到“开放协同”。



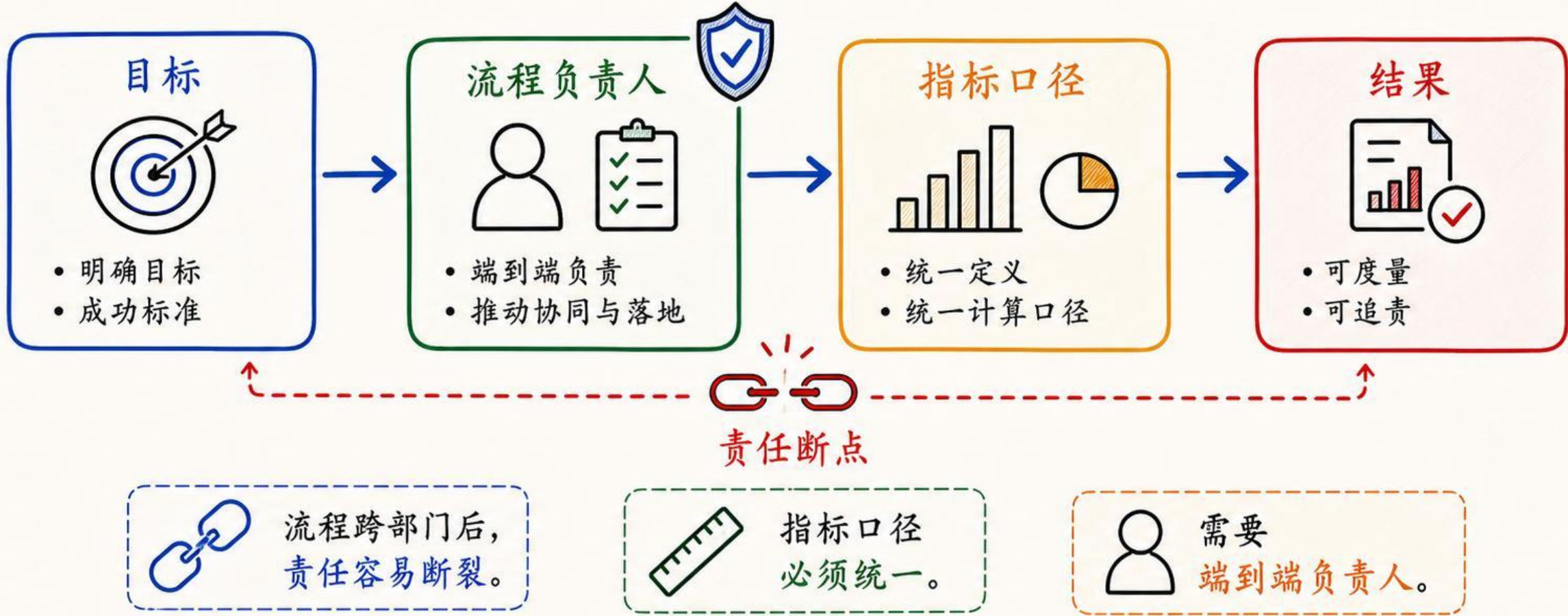
② 流程按目标跨部门调用能力。
以客户价值为中心，端到端流转。



③ 能力节点需要清晰责任人。
定义 Owner、SLA 与考核指标。



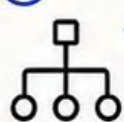
组织需要更清晰的**流程负责人**和**指标口径**，否则没人对结果负责



层级结构会更扁平，但治理层会变强



①



执行链路

减少人工层级。

②



平台 / 赋能团队
更重要。

③



治理与风险团队
必须增强。

治理层（更强）



平台

- 能力编排
- 身份与权限
- 统一标准



治理

- 策略与规则
- 合规与审计
- 验收与度量



风险

- 风险识别
- 监控与预警
- 应急与处置



更扁平，但更可控

执行层（更扁平）



能力节点 A
完成任务单元



能力节点 B
完成任务单元



能力节点 C
完成任务单元



智能体协作 · 跨部门能力 · 结果导向

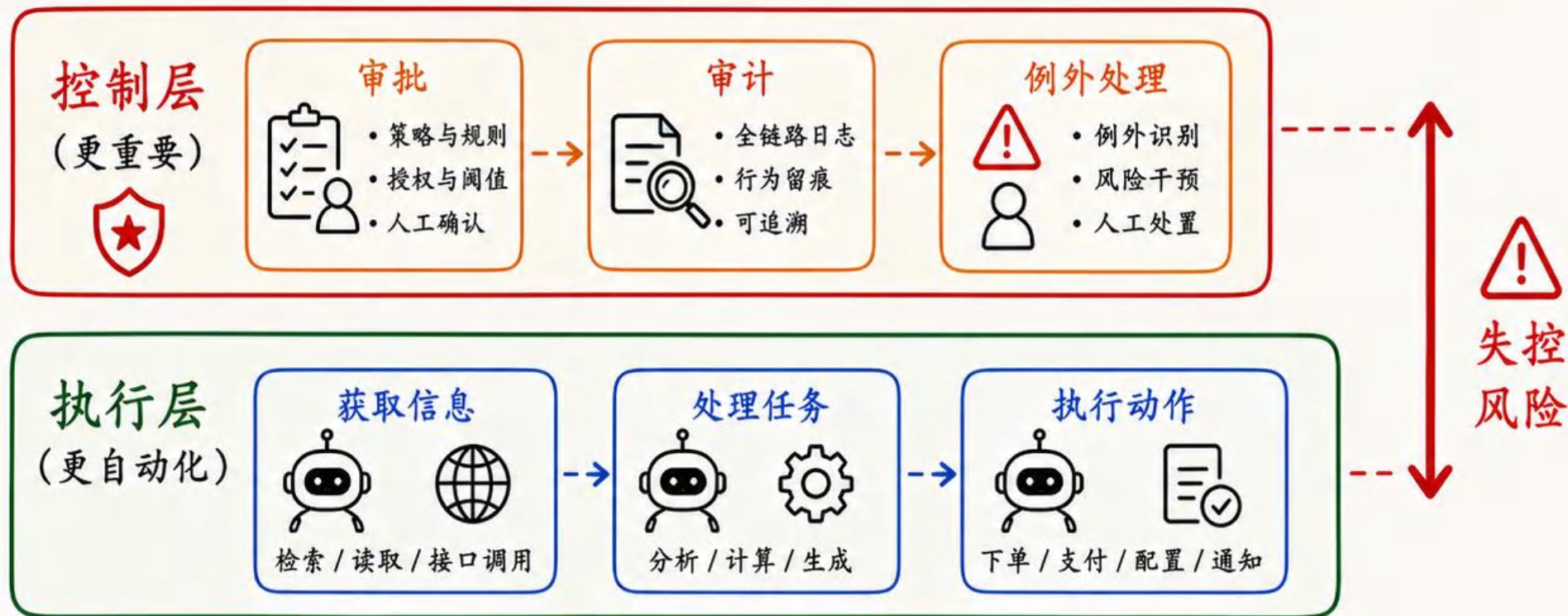


A2A 让执行层更自动化，让控制层更重要

① 执行层：
智能体自动完成任务。

② 控制层：
审批、审计和例外处理。

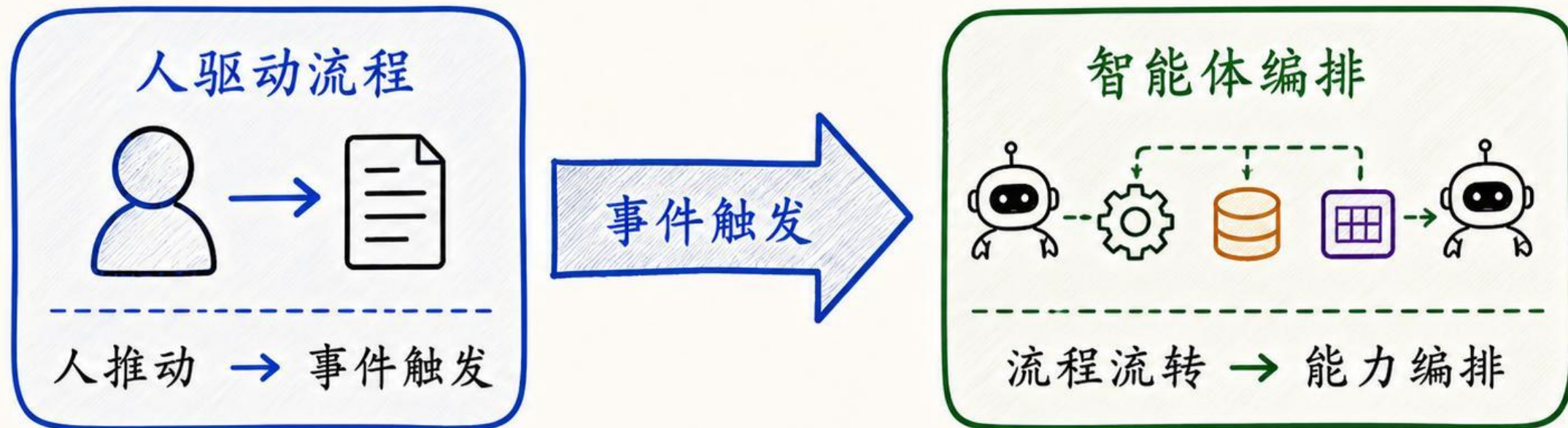
③ 自动化越强，
控制越不能缺位。



原则：放权给执行层，守住控制层，才能实现安全、可靠、可追责的自动化。



workflow 将从人驱动流转变成事件驱动编排



! 新瓶颈：治理与例外





传统岗位将从执行型转向**设计、审核和例外处理**



执行型岗位

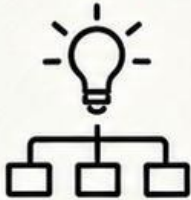


重复执行、操作为主



人转向更高价值岗位

设计



流程设计
方案与规则

审核



质量审核
合规把关

例外处理



复杂例外
决策与处置



① 执行任务被智能体吸收。



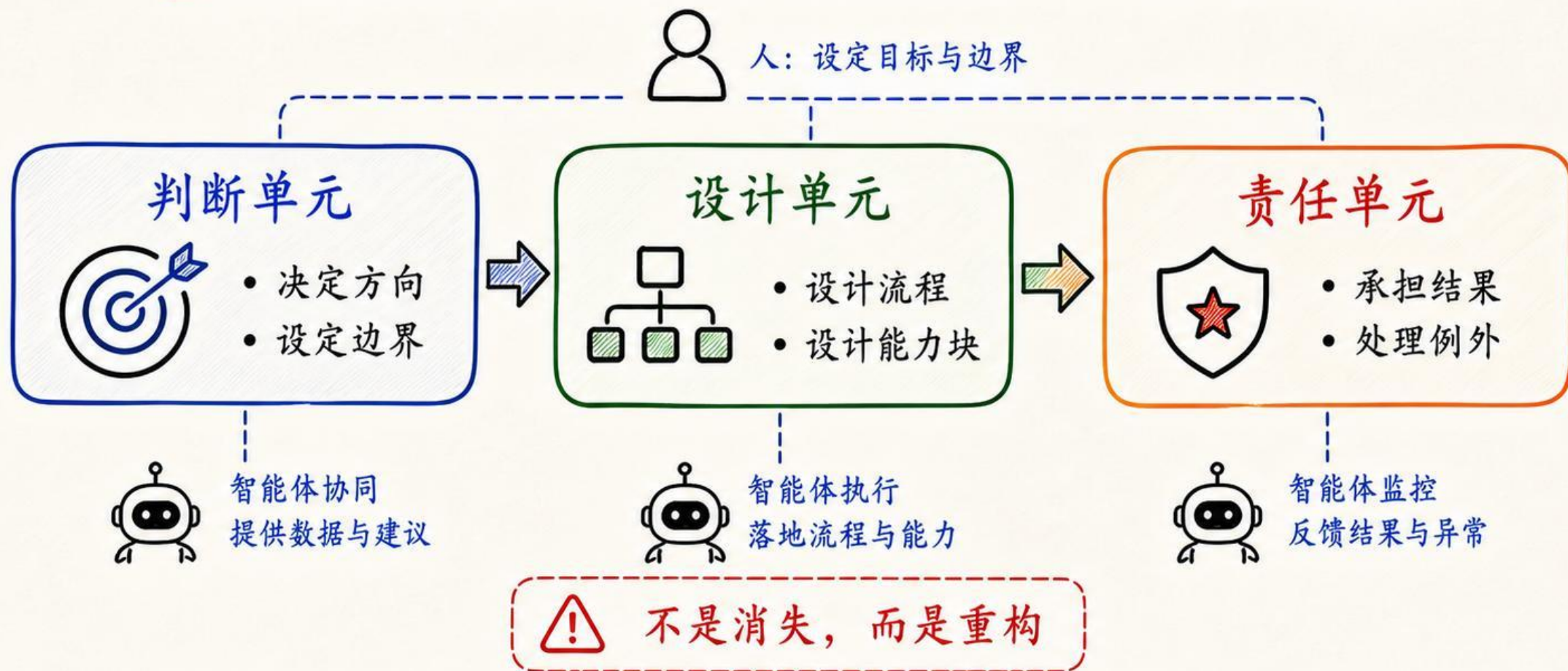
② 人转向流程设计和质量审核。



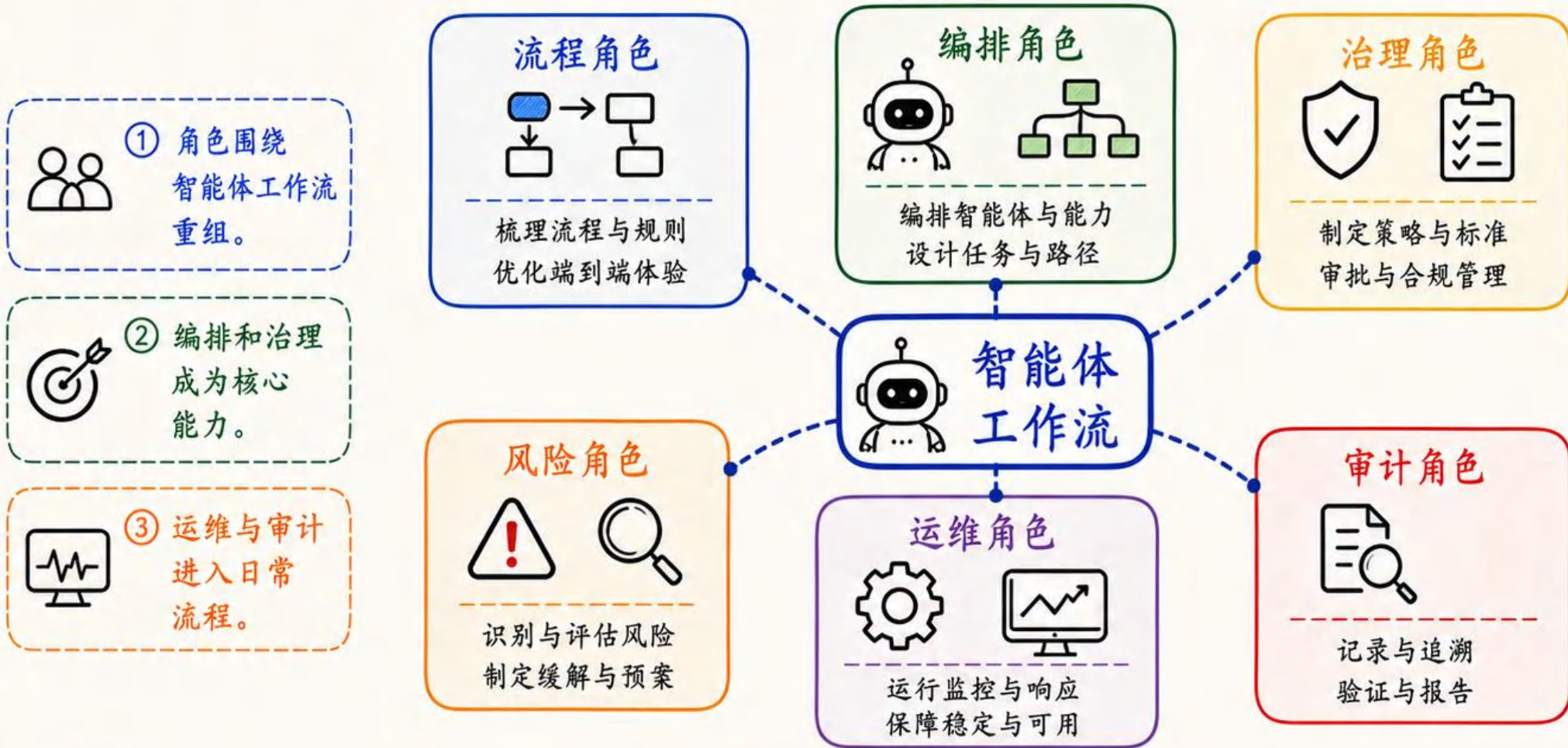
③ 复杂例外仍需人负责。



A2A 不会简单消灭岗位，
而是把岗位改造成**判断单元**、**设计单元**和**责任单元**



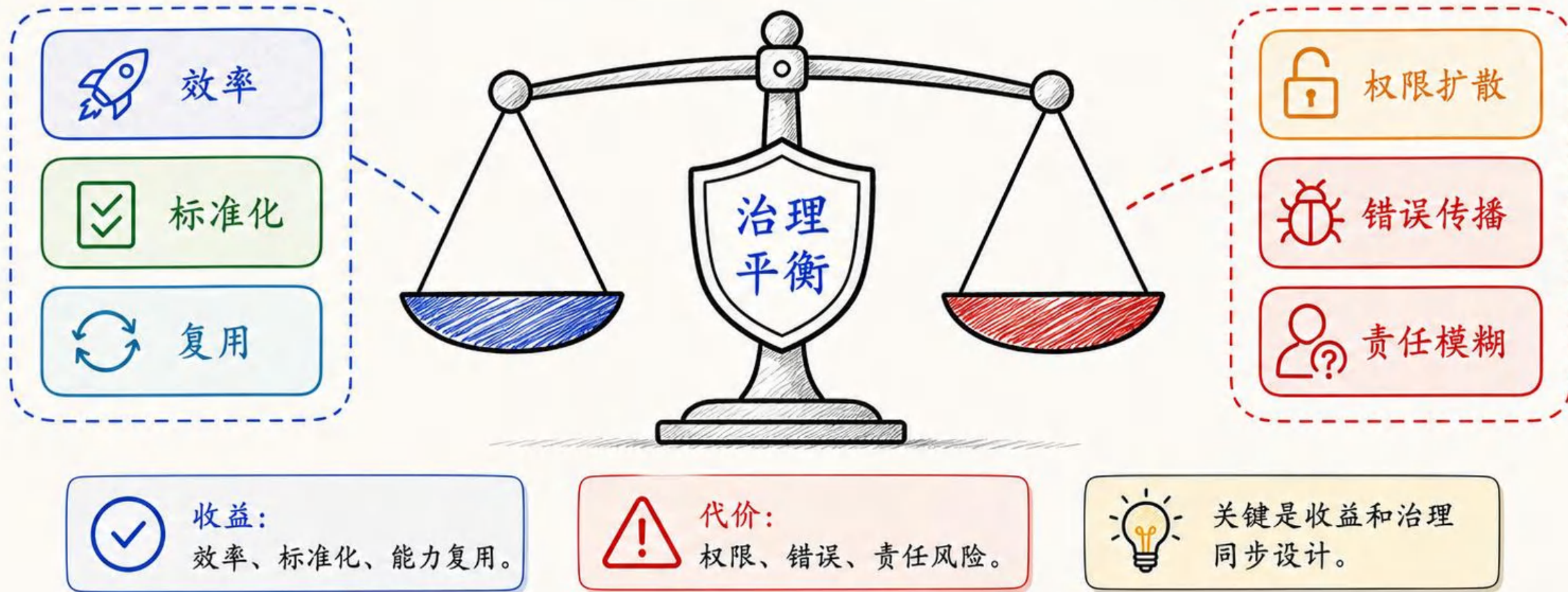
新组织角色将围绕流程、编排、治理、风险、运维和审计形成



- ① 角色围绕智能体工作流重组。
- ② 编排和治理成为核心能力。
- ③ 运维与审计进入日常流程。



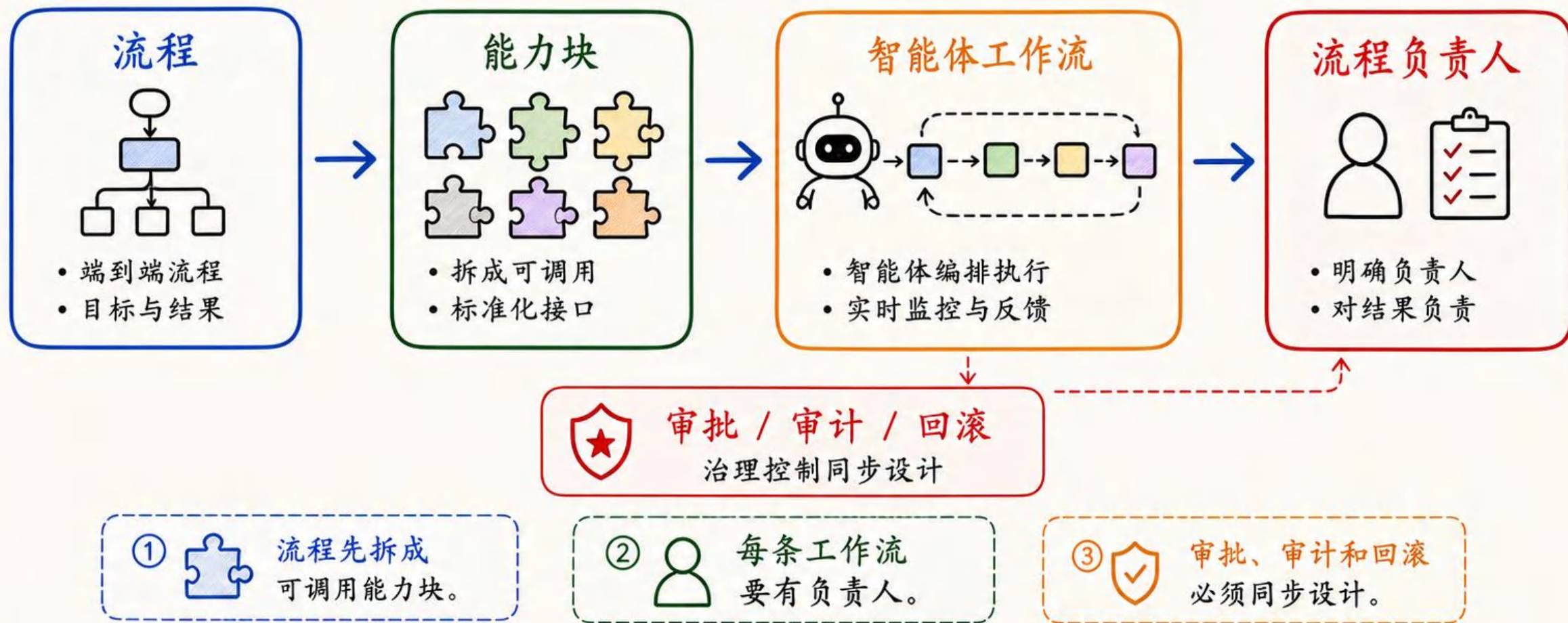
A2A 带来效率、标准化、创新和能力复用，
也带来权限扩散、错误传播和责任模糊



优先选择高频、规则明确、跨系统协作多、人工交接成本高的流程



把流程拆成能力块，并为每条智能体 workflow 设置明确流程负责人



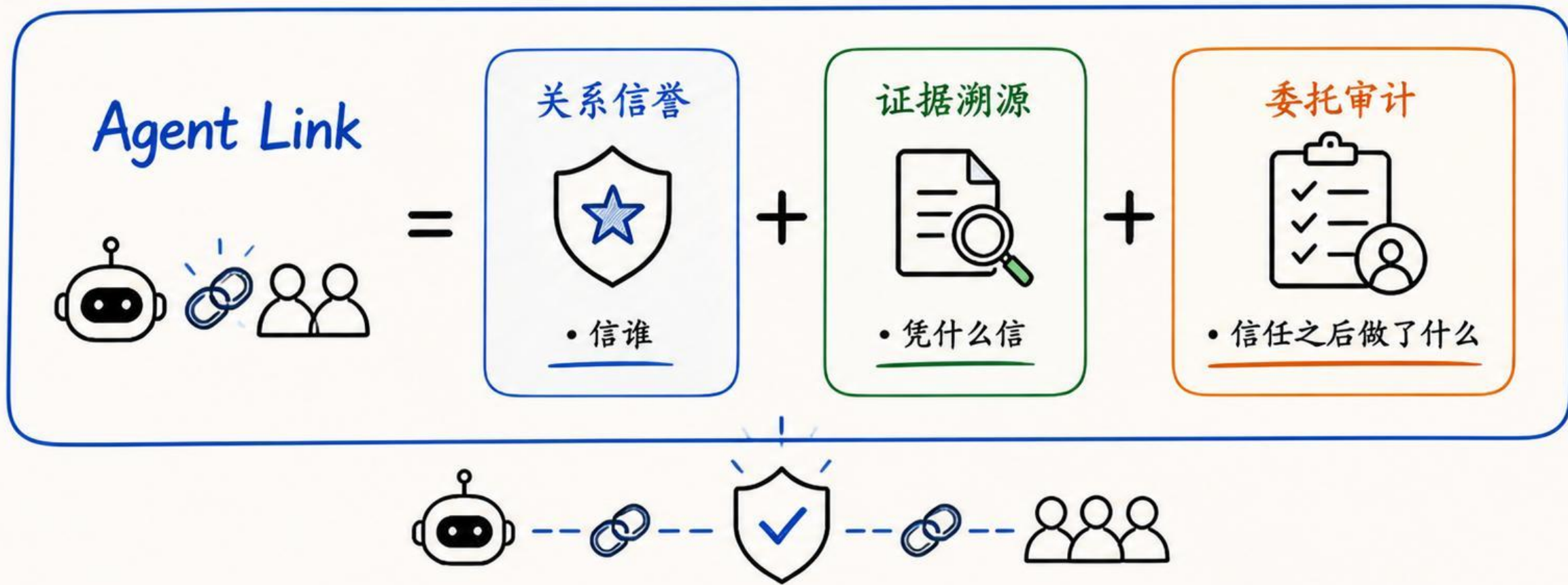
本报告**原创概念**用于形成**可复用**的研究框架，
不代表行业既有**标准**



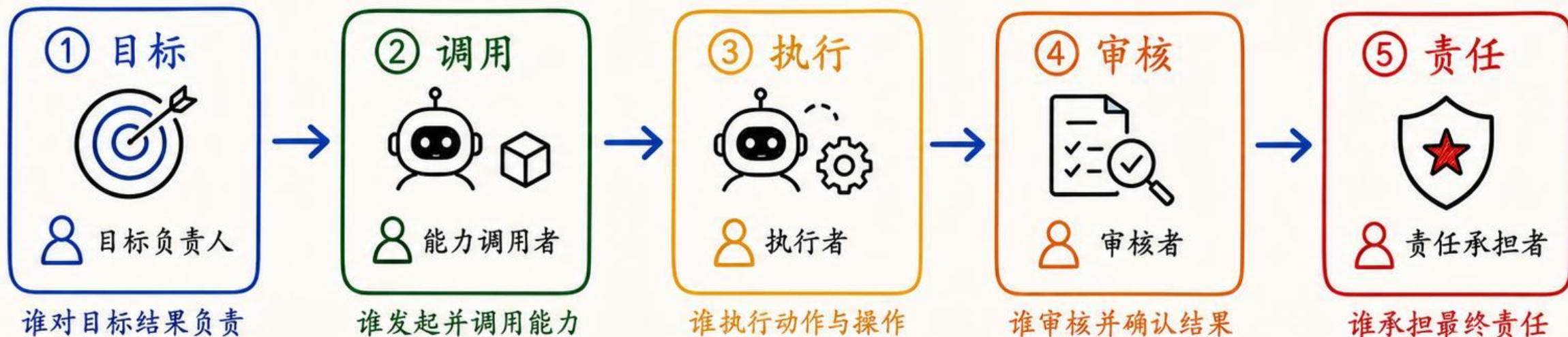
- ✓ 用于研究和讨论。
- ✓ 不等同于行业标准。
- ✓ 帮助复用分析框架。



智能体链接 = 关系信誉 + 证据溯源 + 委托审计



智能体责任图记录谁负责目标、
谁调用能力、谁执行动作、谁审核结果、谁承担责任



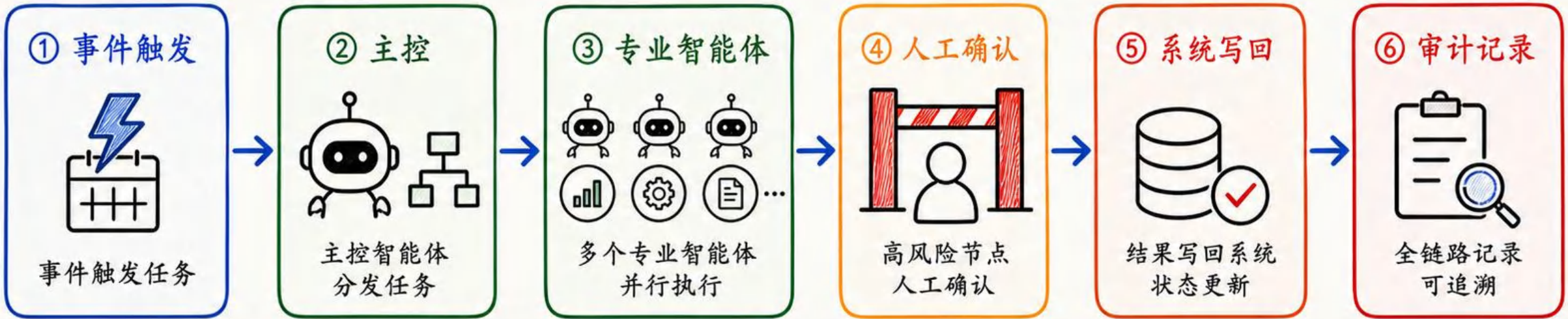
不能匿名负责



责任链清晰可追溯，才能保障安全、合规和可追责。



智能体任务链由事件触发、主控智能体、
多专业智能体、人类确认节点、系统写回和审计记录构成



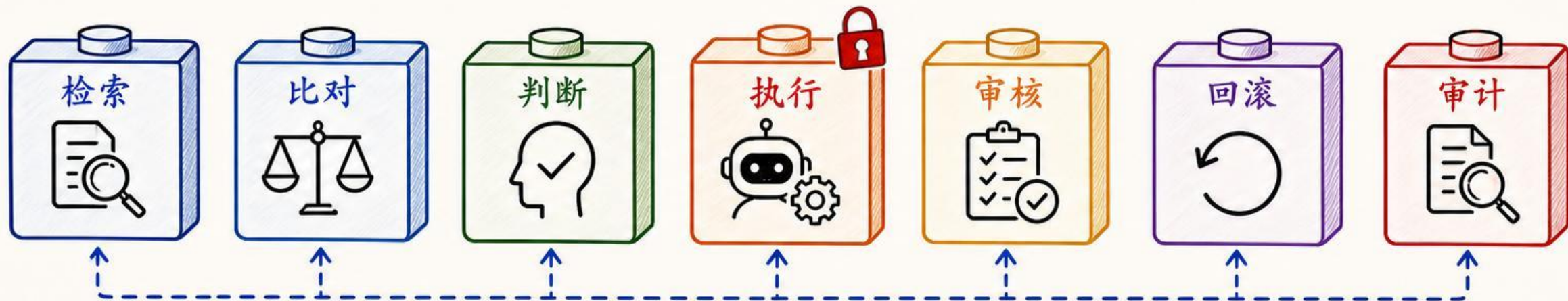
① 事件触发任务。

② 主控分发给专业智能体。

③ 高风险节点进入人工确认和审计。



Agent 能力块是检索、比对、判断、执行、审核、回滚、审计等可复用能力单元



可发现、可调用、可组合、可编排

① 能力块可以被发现和调用。

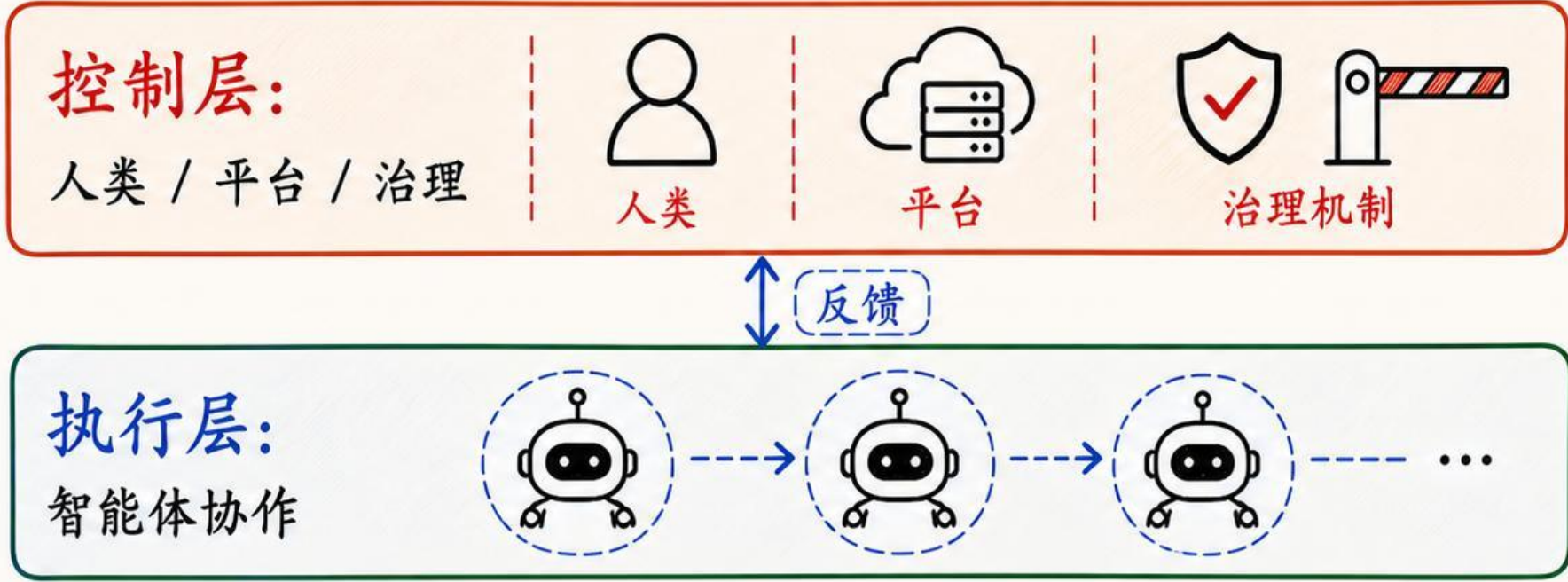
② 多个能力块组成任务链。

③ 高风险能力块必须被治理。





双层组织由智能体自动协作的**执行层**与
人类、平台和治理机制构成的**控制层**组成



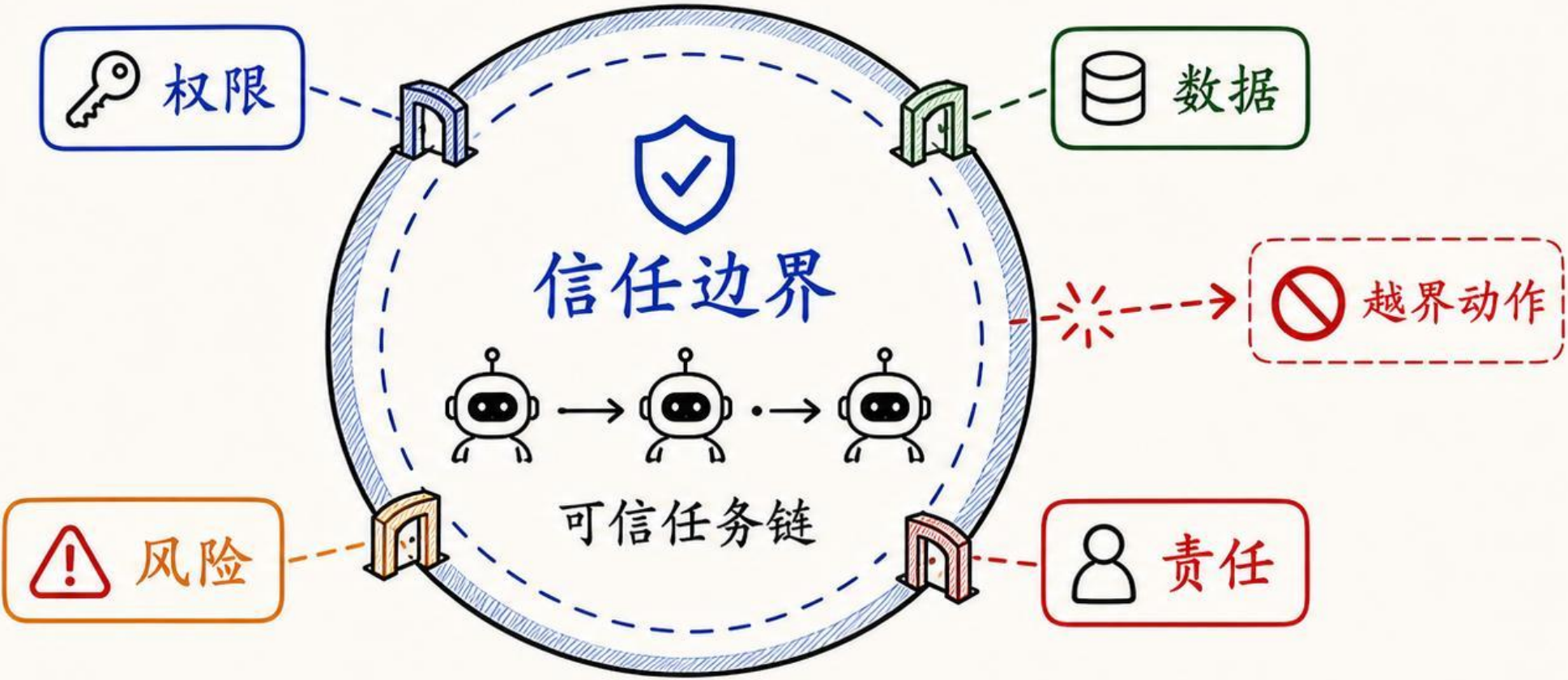
① 执行层：
智能体自动协作。

② 控制层：
人类、平台、治理机制。

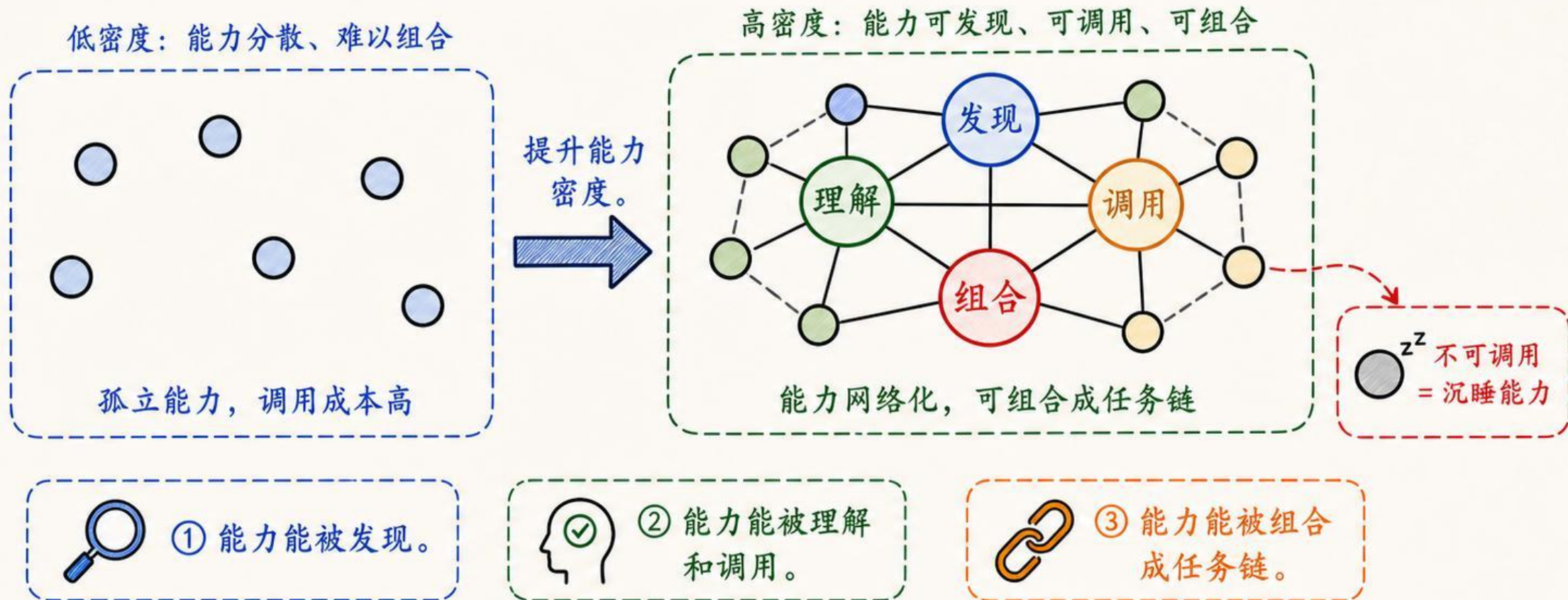
③ 两层
必须同时设计。



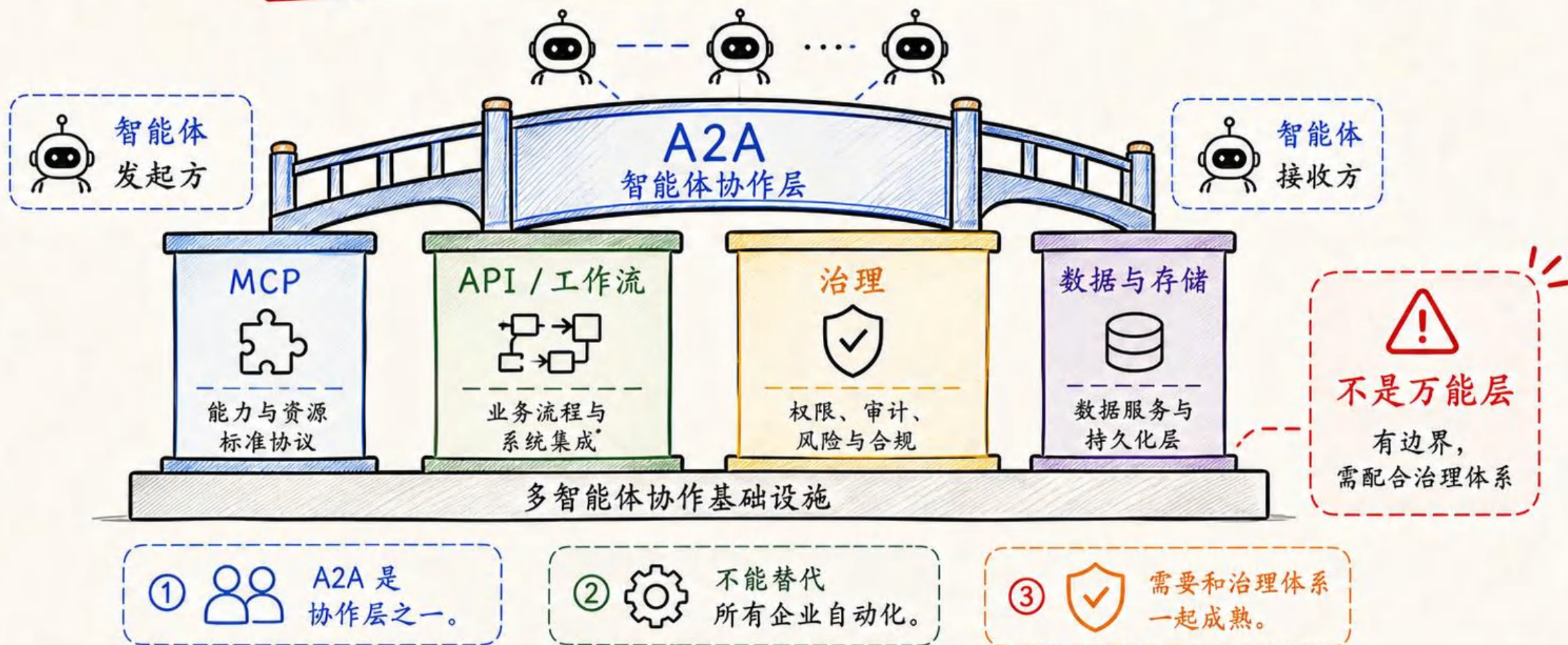
Trust 边界 / 信任边界围绕
权限、数据、风险和责任人链划定



可调用能力密度衡量组织中能够被智能体发现、理解、调用和组合的有效能力丰富程度

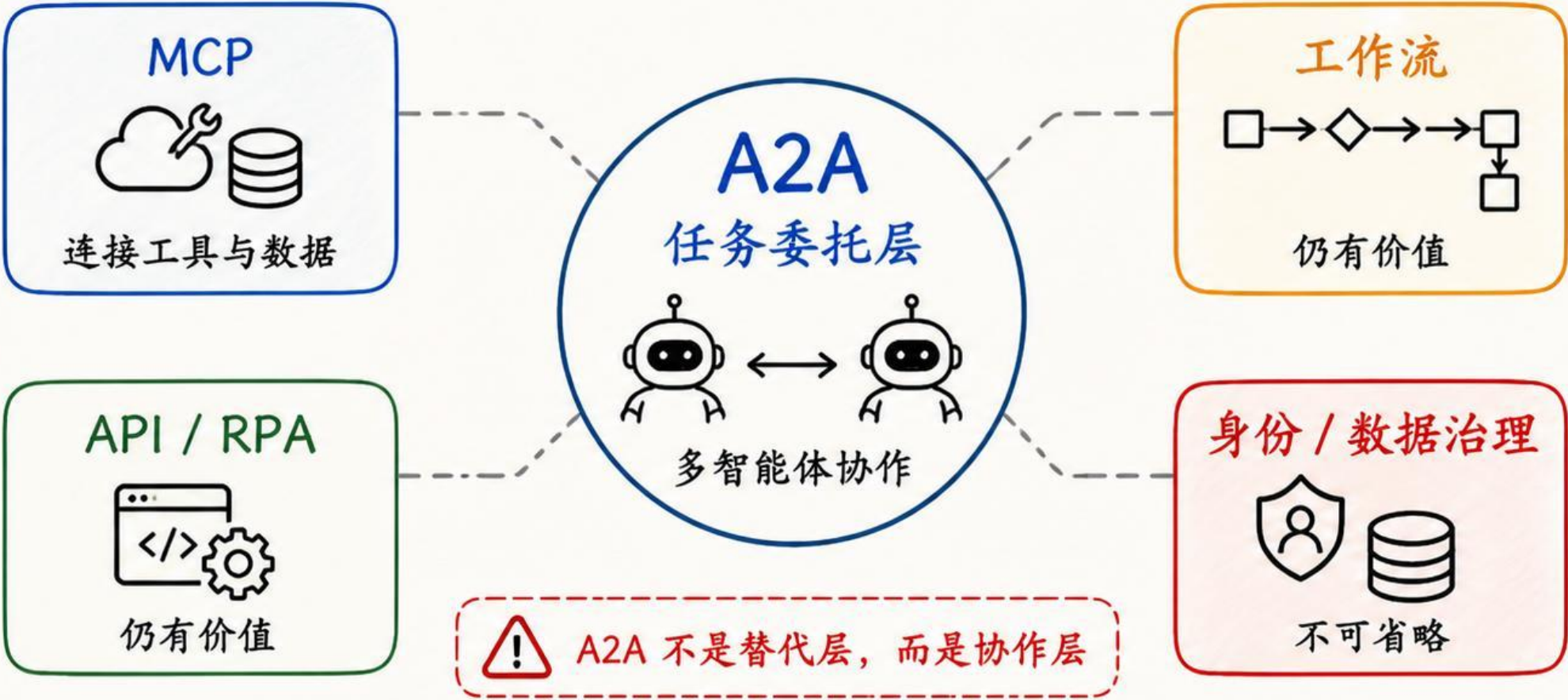


A2A 会成为多智能体协作的重要基础设施之一，但不是万能答案

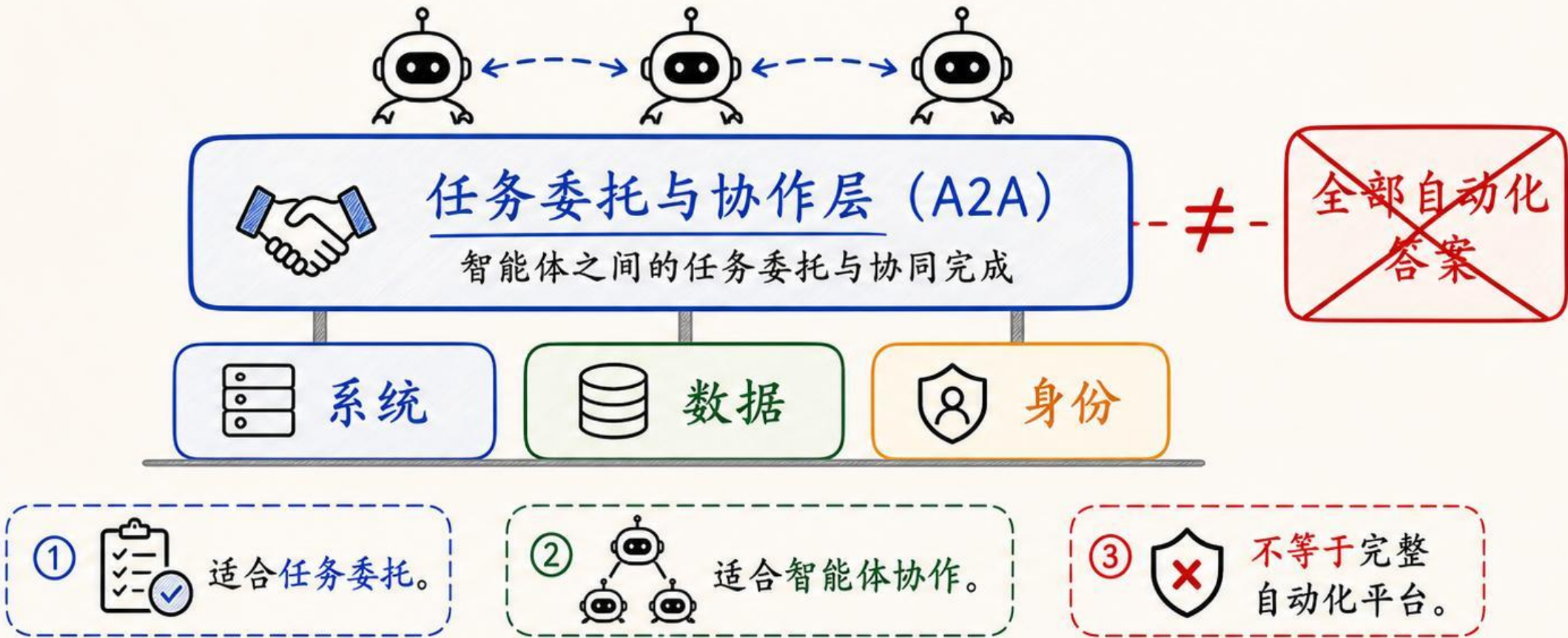




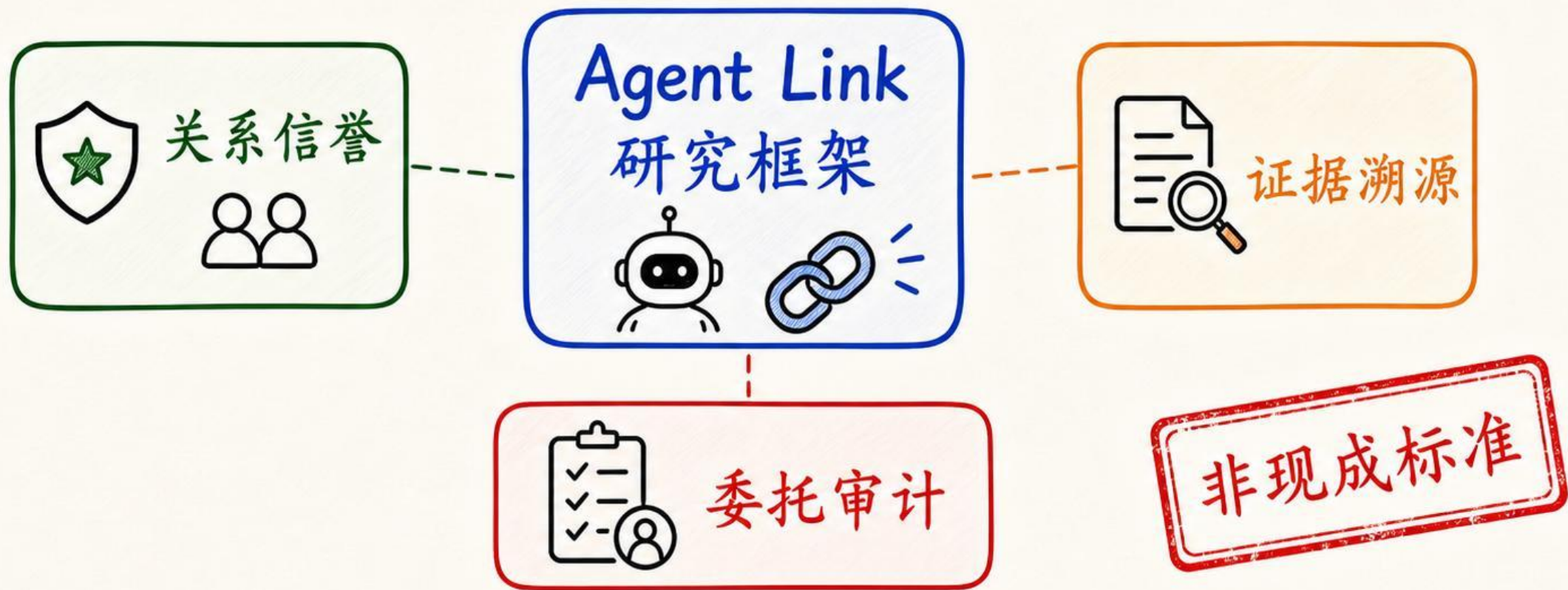
A2A 不是万能编排层



A2A 更适合被理解为任务委托与协作层，
而不是企业自动化的全部答案

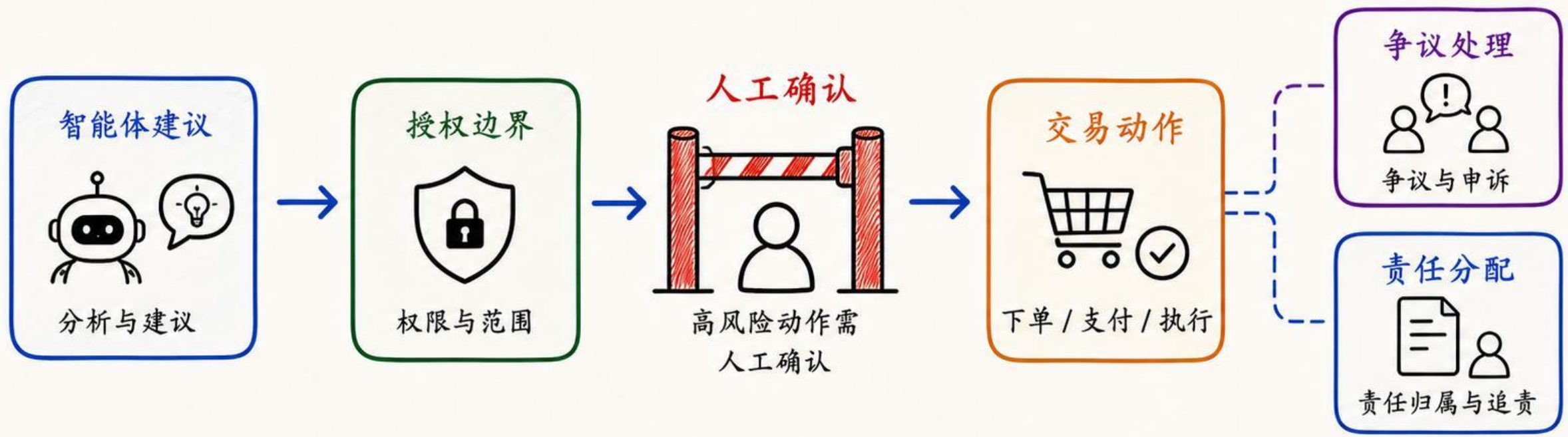


Agent Link 是本文提出的研究框架，
不是现成行业标准





Agentic Commerce 不等于完全自动交易



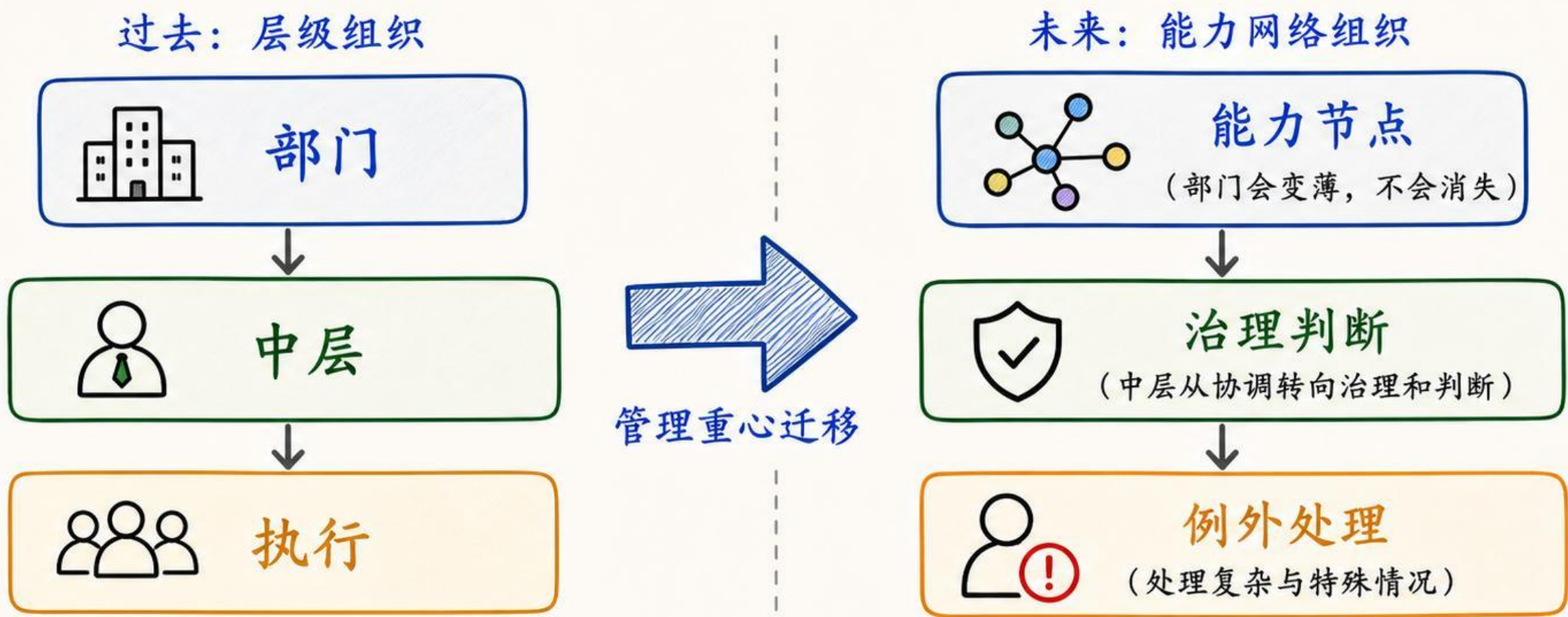
① 授权边界必须清楚。

② 高风险动作需要人工确认。

③ 争议处理和责任分配不能省略。



A2A 不会简单消灭部门和中层



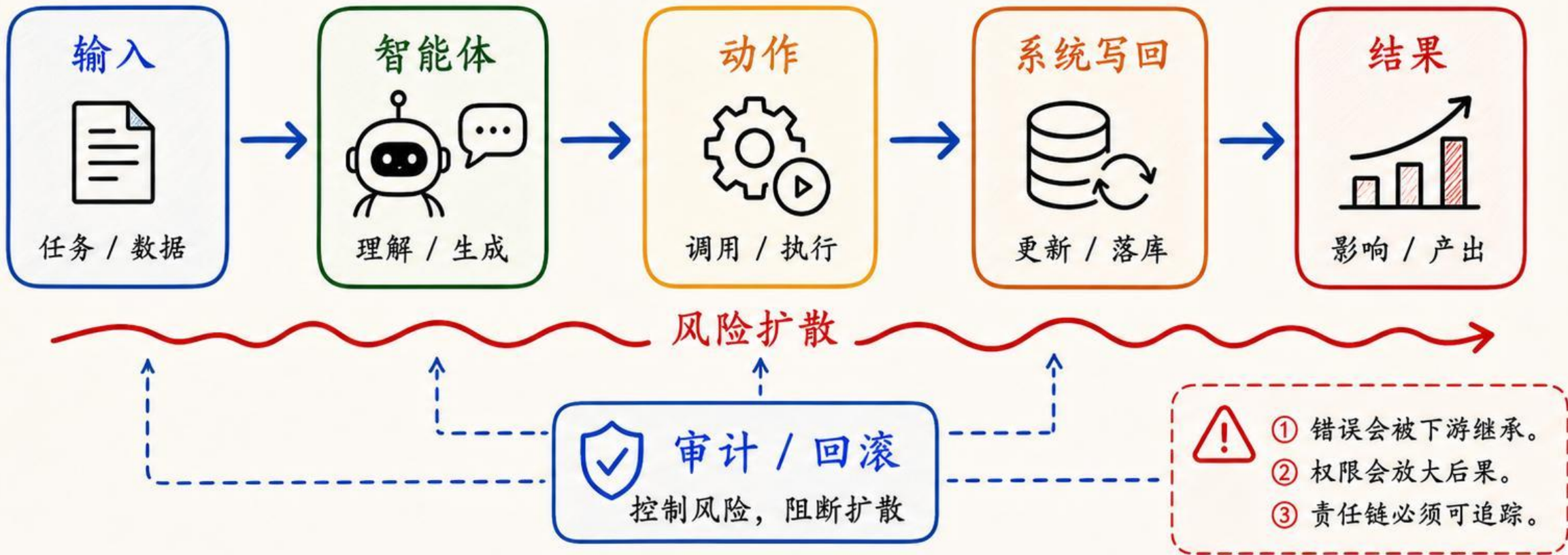
效率提升必须以可控委托为前提



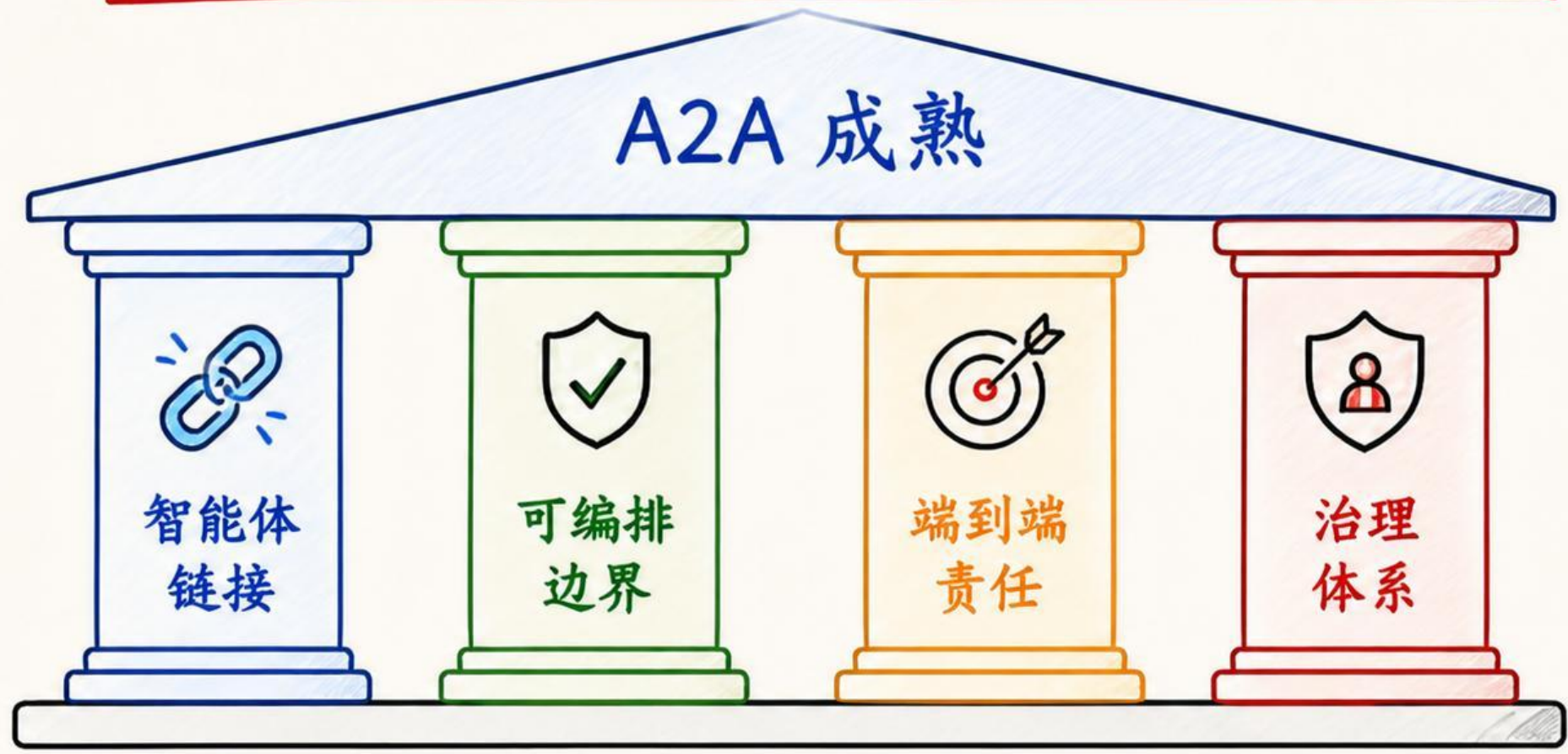
⚠ 没有治理边界，委托越多，风险越大



A2A 的风险不是智能体会不会回答错，
而是**错误、权限和责任**会沿任务链**扩散**



A2A 的成熟取决于智能体链接、可编排边界、端到端责任设计和治理体系是否共同成熟

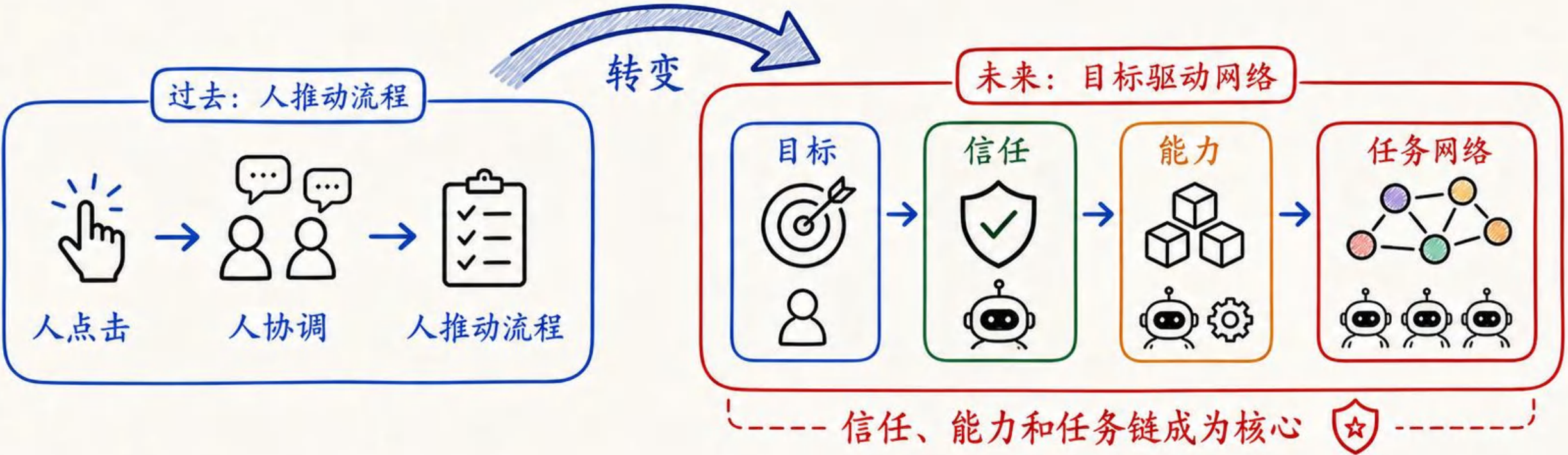


A2A 最值得研究的地方，
不在于让智能体互相聊天，

而在于推动任务委托、能力调用、商业交易和组织流程
重新被设计

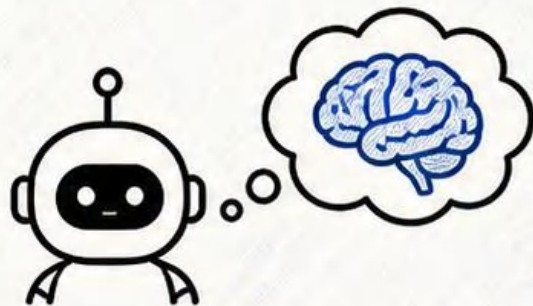


A2A 将让商业系统和组织运行逻辑
从人点击、人协调、人推动流程，
转向人设定目标、智能体建立信任、
智能体调用能力、任务在能力网络中流转



模型大小决定智能体理解任务的能力，
权限边界决定智能体造成后果的能力，
编排画像决定智能体能否进入企业级协作网络

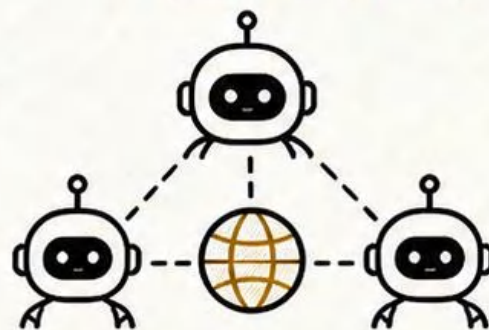
模型大小
→ 理解任务



权限边界
→ 控制后果



编排画像
→ 企业协作



★ 智能体治理的三条主线 ★

