白皮书 · 研究稿 V0.3

协作的结构

从人类会议到 Agent Native 组织

人类协作可以迁移到 Agent Native 组织，但可迁移的不是会议形式，而是「承诺—验证—追责」的功能； 不能重建为可验证、可授权、可审计之硬约束的部分，必须由人类承担。

2026 年 3 月 · collaboration-structure.varnothing.works

目录

01 执行摘要

02 问题

03 协作的结构

04 会议作为结算节点

05 纪要作为稳定机制

06 承诺闭环

07 Agent 环境

08 迁移

09 工具与系统

10 落地

11 反例、风险与边界

12 结论

13 附录 A：证据表

14 附录 B：诊断表

15 附录 C：术语表

16 附录 D：参考文献

章节

执行摘要

三句话、该做与不该做、三类读者路径、证据强度说明。

0.1 三句话

1. **人类协作的有效性，不取决于「开了多少会」，而取决于「承诺是否闭环」。** 一个承诺闭环包含八环：差异 → 决定 → 证据 → 委托 → 授权 → 记录 → 责任 → 复查/撤回。
2. **Agent Native 组织不能复制人类会议的「形式」，只能迁移它的「功能」。** 可迁移的是「承诺—验证—追责」；不能迁移的是社会性（信任/声誉/文化/仪式）与责任（法律/道德）。能否迁移，取决于能否把对应功能重建为可验证、可授权、可审计的硬约束。
3. **工具不是会议软件，而是承诺闭环的载体。** 它需要三件东西：共享状态（会议载体）、结构化产物（组织记忆）、硬约束（权限/状态机/可见性/审计）。

0.2 该做 / 不该做

该做	不该做
先建承诺闭环：谁决定、凭什么证据、授权到哪里、如何记录、谁承担后果	先买 Agent 平台，再想协作
选一个高价值、可验证的决策流做 90 天试点	一次改全公司所有会议
把纪要改成「稳定机制」：签发、核对、编号、时限、执行	只换纪要模板
用硬约束替代软规范：权限、状态机、可见性、审计	把规则写进 prompt/skill 就以为会执行
保留人类兜底：不可委托清单、checkpoint、决策日志	假设 Agent 会「自觉负责」
接受反例条件：如果等价物不能在三个以上流程中降低协调成本/错误率，命题不成立	把「功能等价」当公理

0.3 三类读者路径

读者	先读	再读	带走
CEO / 管理者	0、1.1–1.3	5、9.2–9.4、10	一个决策流试点、一套承诺闭环指标
产品 / 技术负责人	0、3、5	8、9、10	会议载体、产物模型、MVP 边界
Agent 开发者	0、6.2–6.4	7、8、10	四个限制、硬约束、失败模式

0.4 证据强度说明

本白皮书的核心命题是设计假设，不是已被证明的规律：

- **强证据**：法规/标准/公开记录（A 级）支持「机制存在」，不支持「机制有效」；
- **中证据**：学术理论（B 级）支持「结构解释」，不支持「迁移可行」；
- **弱证据**：行业实践/作者推断（C 级）支持「设计方向」，不支持「因果结论」。

因此，白皮书的建议应被当作「可检验的设计命题」，而不是「最佳实践」。

章节

问题

为什么买了 Agent 协作却没变：三个误解、核心论点与反例条件。

> 本章回答：为什么现在需要这份白皮书？最常见的失败是什么？核心论点是什么？

1.1 一个常见失败：买了 Agent，协作没变

很多公司已经买了 Copilot、Agent 平台、自动化工具，但协作没有变：

- 会议时长没有下降；
- 决策周期没有缩短；
- 纪要还是没人看；
- 行动项还是没人跟踪；
- 跨部门还是靠「拉群/开会」。

为什么？

因为工具改变的是「执行速度」，不是「协作结构」。如果决策权不清、议题没有准入、纪要不是承诺、行动项没有跟踪，那么 Agent 只是「更快的打字员」，不会改变协作。

这就是本白皮书的起点：协作的瓶颈，不在工具，而在结构。

1.2 三个误解

误解一：把 Agent 当人

人们容易把 Agent 的「会说话」当成「有理解」，把「会道歉」当成「有责任」，把「会承诺」当成「会履行」。

这是「拟人化陷阱」：Agent 可以模拟人类行为，但没有人类的社会性——没有信任、声誉、情感、身份、责任。

后果：人类把责任推给 Agent，或者期待 Agent 像人一样「自觉」。

误解二：把会议当沟通

人们容易把会议当成「信息传递」：大家说说，互相知道，就完了。

但会议的本质不是「沟通」，而是「结算」：把分散的信息、权力、注意力，在同一时空内结算为承诺。

后果：会议开成「汇报会」，没有决定；或者开成「神仙会」，没有执行。

误解三：把规范当约束

人们容易把「写进文档的规则」当成「会执行的约束」：写了《会议管理制度》，就以为会议会变好；写了 skill，就以为 Agent 会遵守。

但规范是「软约束」：人可以忽略，Agent 可以绕过。

后果：制度写在墙上，行为照旧；Agent 在关键时候「忘了规则」。

1.3 核心论点与反例条件

核心论点：

> 人类协作可以迁移到 Agent Native 组织，但可迁移的不是会议形式，而是「承诺—验证—追责」的功能；不能重建为可验证、可授权、可审计之硬约束的部分，必须由人类承担。

这条论点包含三个部分：

1. **可迁移的是功能**：人类会议的功能是「结算承诺」，不是「坐在一起」。Agent 组织可以用不同载体实现同样功能。
2. **迁移需要硬约束**：Agent 没有社会性，不能靠信任/规范/声誉；需要验证、授权、审计、状态机。
3. **不能迁移的是责任**：Agent 可以执行，但不能承担法律/道德责任；责任必须由人类兜底。

范围条件（白皮书自我限制）：

「功能等价」只在以下范围成立：

- 可逆/常规/内部/可验证的决策：Agent 可以承担主要执行；
- 可逆/对外/常规的决策：Agent 执行 + 人类签署；
- 不可逆/对外/新颖/不可验证的决策：人类主导，Agent 只做辅助。

超出这个范围，「功能等价」是未验证的假设，不是结论。

反例条件（白皮书自我限制）：

- 如果等价物不能在三个以上流程中降低协调成本或错误率，命题不成立；
- 如果验证成本持续高于收益，迁移不经济；

- 如果人类无法理解 Agent 的关键决定，责任兜底失效；
- 如果 Agent 的策略性行为（合谋/欺骗/联盟）无法治理，授权链失效。

1.4 定义

Agent Native 组织：由人类与 Agent 共同组成、人类保留最终决策权的组织。它的定义不是「用了 Agent」，而是：

- 决策权可分配（人类/Agent/混合）；
- 角色可定义、复制、撤销；
- 上下文可共享、版本化、检索；
- 产物可追溯、可审计。

会议：一个临时的、有边界的、多主体的结算节点。它同时结算信息流、授权流、承诺流。

会议纪要：承诺的稳定机制，不是「记录讨论」。它通过签发、核对、编号、时限、执行五个机制，把瞬时共识转化为可引用的承诺。

承诺闭环：差异 → 决定 → 证据 → 授权 → 记录 → 责任。

1.5 证据分级与边界

级别	证据类型	例子	使用方式
A	一手法规/标准/公开记录	公司法、ISO 导则、NASA NPR、FOMC minutes	直接引用，作为「机制存在」的证据
B	学术研究/理论	Williamson、Thompson、Rogelberg	作为解释框架，不当作绝对真理
C	行业实践/作者推断	GitLab/Amazon/Basecamp 实践、作者经验	作为参考，标注「通常/多数」

边界：

- 本白皮书不是系统文献综述，没有穷尽所有研究；
- 不是实证研究，没有大规模样本；
- 不是法律意见，不替代法务/合规审核；
- 不是技术规范，不替代架构设计文档。

章节

协作的结构

市场、层级/官僚制、社区/网络三种基本结构，六个机制与四个约束。

> 本章回答：人类社会的协作方式，有哪些结构特征？为什么它们不同？它们各自依赖什么、在哪里失败？

2.1 协作不是「一起做事」，而是「协调差异」

如果两个人完全一样——同样的信息、同样的利益、同样的权力、同样的能力、同样的时间——他们不需要协作，只需要「同时做同一件事」。

协作之所以存在，是因为有差异：

差异	含义	协调机制
信息差异	你知道的我不知道	沟通/共享/价格
利益差异	你要的和我想要的不一样	市场/契约/谈判
权力差异	你能决定我不能	层级/授权/投票
能力差异	你会我不会	分工/师徒/市场
时间差异	你急我不急	计划/优先级/协调

协作的结构，取决于差异的类型。 不同的差异，需要不同的协调机制。用错机制，协作就会失败。

2.2 三种基本结构

人类协作方式很多，但可以归为三种基本结构。它们不是「组织形态」，而是「协调机制」。

结构	核心协调机制	主要协调的差异	典型例子
市场	价格	利益/信息	买卖、外包、招聘
层级/官僚制	权威/规则/档案	权力/信息	科层制、军队、政府
社区/网络	规范/信任/互惠	利益/规范/信任	开源社区、合作社、供应链伙伴

市场：用价格协调利益

市场假设：每个人追求自己的利益，价格把分散的信息压缩成一个信号。

- 优点：分散、可扩展、激励强；
- 缺点：机会主义、外部性、信息不对称；
- 前提：产权、契约、法律、货币。

层级/官僚制：用权威和规则协调权力

层级假设：上级比下级拥有更多信息和权力，命令可以降低协调成本。

官僚制是层级的规则化形态：它不只靠「命令」，还靠「规则、书面档案、管辖权、职位等级、考核」。Weber 的官僚制，正是现代会议纪要制度的祖先：会议记录、签字、归档、编号，都是官僚制的技术。

- 优点：清晰、快速、可问责、可大规模；
- 缺点：官僚化、信息失真、激励扭曲、仪式主义；
- 前提：权威、合法性、规则、监督、档案。

社区/网络：用规范与信任协调合作

社区假设：成员共享规范/身份，规范可以约束行为，互惠可以维持合作。

网络是社区的「结构变体」：社区强调规范/社会化，网络强调关系/信任/互惠。

- 优点：低成本、高信任、创新、长期；
- 缺点：搭便车、群体思维、治理僵局、封闭/共谋；
- 前提：共同身份、规范、关系、边界。

2.3 混合形态

现实组织很少是「纯市场」「纯层级」「纯社区」。它们通常是三种基本结构的混合。

混合形态	组合	例子
协会	社区 + 层级 + 标准	行业协会、标准组织
平台	市场 + 层级 + 规则/接口	应用商店、交易平台
小群体/家族	社区 + 层级 + 亲密	家庭、创业团队、家族企业
公地/合作社	社区 + 集体产权 + 规则	合作社、社区资源治理

为什么不用「七种并列类型」？

因为七种类型不在同一分析层次：市场/层级/网络是治理结构；社区/小群体是规范/亲密机制；协会/平台是组织/技术形式。把它们并列，会造成「层次混淆」和「虚假精度」。

三种基本结构 + 混合形态，更诚实，也更适合迁移分析：市场→效用/竞价，层级→权限/审批，社区/网络→验证/审计。

2.4 六个机制

三种结构背后，有六个协作机制。它们不是并列的「功能」，而是「结构条件」。

机制	作用	市场	层级/官僚制	社区/网络
信息/沟通	让主体知道	价格信号	汇报/文件	共享/对话
决策权/权威	让主体行动	分散/自主	集中/命令	共识/协商
激励/资源分配	让主体愿意	收益	工资/晋升	互惠/认同
规范/信任/合法性	让主体可预测	契约/法律	规则/监督	规范/信任
执行/制裁	让主体遵守	法律/退出	惩罚/审计	制裁/排斥
学习/记忆	让主体改进	交易经验	档案/培训	实践/故事

说明：

- 「协调」不是第六个机制，而是前五个机制的结果；
- 「信任」不能被「验证」简单替代：验证降低不确定性，产生 assurance/control，不是 trust；但在 Agent 环境中，验证是「功能等价物」；
- 「激励」不能被「目标函数」简单替代：人的激励多维、内生，Agent 的目标函数是设计物；
- 「意义/认同」是横切维度：它影响所有机制，但不是独立机制。

2.5 四个约束

三种结构、六个机制，都受四个底层约束。

约束一：稀缺性

- 时间稀缺：人一天只有 24 小时；
- 注意力稀缺：人一次只能处理有限信息；
- 信任稀缺：信任需要时间/经历建立。

结构后果：会议需要议程/时间盒/参会人筛选；层级需要授权/审批；市场需要价格/合同；社区需要共识/代表。

约束二：社会性

- 人有情感、身份、关系、面子、权力；
- 人会从众、服从权威、维护群体；
- 人会记仇、报恩、结盟、排外。

结构后果：会议需要主持人/议事规则/心理安全；层级需要权威/合法性；社区需要规范/仪式；网络需要信任/关系。

约束三：隐性知识

- 很多知识无法完全编码：经验、直觉、手艺、关系、语境；
- 隐性知识可以通过师徒/共同实践/观察传递，也可以部分通过文档/示例/视频/外部化传递；
- 隐性知识难以规模化，但不是「完全不可编码」。

结构后果：会议需要讨论/白板/故事；层级需要师徒/轮岗；社区需要实践共同体；平台需要文档/示例。

约束四：问责

- 法律责任不能完全转移给「系统」；
- 道德责任需要「人」承担；
- 责任需要授权/记录/追溯。

结构后果：会议需要签字/纪要/决策记录；层级需要问责/审计；市场需要合同/法律；社区需要声誉/制裁。

补充：不确定性、有限理性、资产专用性

除了四个约束，协作结构还受三个「交易成本」维度影响：

- **不确定性**：任务越不确定，越需要高带宽/多轮/信任；
- **有限理性**：人/Agent 无法处理所有信息，需要规则/简化；
- **资产专用性**：投入越专用，越需要长期关系/层级/契约。

这三个维度，解释了「为什么市场/层级/社区会在不同任务上替代」：

- 不确定性低、资产专用性低 → 市场；

- 不确定性高、资产专用性高 → 层级/社区；
- 频率高、需要协调 → 层级/社区。

2.6 三种结构的失败模式

结构	失败模式	根源
市场	机会主义、外部性、垄断	信息不对称、激励错配
层级/官僚制	官僚化、信息失真、权力斗争、仪式主义	信息不对称、激励扭曲、规则僵化
社区/网络	搭便车、群体思维、治理僵局、封闭/共谋	公共品、共识成本、关系壁垒

没有一种协作结构是没有代价的。 选择一种结构，就是选择一组失败模式。

2.7 会议在三种结构中的位置

会议不是「第四种结构」，而是三种结构的「交汇点」：

结构	会议的作用	会议的形态
市场	谈判/签约/拍卖	双边谈判、拍卖会、签约仪式
层级/官僚制	汇报/决策/合法化	董事会、常务会、评审会
社区/网络	共识/审议/协调	社区会议、工作组、标准会议

会议的功能，取决于它服务哪种结构：

- 市场会议：重点是「交易条件」，不是「共识」；
- 层级会议：重点是「决策/授权」，不是「讨论」；
- 社区会议：重点是「共识/规范」，不是「命令」。

用错会议形态，就会「结构错配」：用社区会议做层级决策（议而不决），用层级会议做社区共识（压制反对）。

本章结论：

> 人类协作有三种基本结构：市场、层级/官僚制、社区/网络。会议是三种结构的交汇点，不是第四种结构。理解这一点，才能理解为什么会议不是万能的，以及为什么 Agent 协作需要新的结构。

章节

会议作为结算节点

会议同时结算信息流、授权流、承诺流；五个结构变量与输入/输出接口。

> 本章回答：会议在协作中处于什么位置？它由哪些结构特征构成？为什么有些会议有效，有些无效？

3.1 会议的本质：多流结算，但不是唯一

会议不是「一群人说话」，而是一个**临时的、有边界的、多主体的结算节点**。

它通常同时结算三件事：

流	内容	失败后果
信息流	事实、数据、选项、风险	各说各话，事实没对齐
授权流	谁有权决定、谁批准、范围多大	议而不决，或决定被推翻
承诺流	谁做什么、什么时候、什么算完成	行动项失踪，承诺蒸发

但会议不是唯一能结算三流的场合。 这一点必须说清楚，否则会陷入「会议万能论」：

- 邮件列表可以结算信息流和承诺流（Apache 的投票）；
- 董事会书面决议可以结算授权流（不开会也能决议，只要程序合规）；
- 任务系统可以结算承诺流（行动项）；
- 审批流可以结算授权流（OA/权限系统）。

会议的特殊性，不是「唯一」，而是「**在同一时空内，让多方同时到场、同时承诺、同时被见证**」。它适合处理：

- 需要多方同时授权的决定；
- 需要高带宽讨论的复杂问题；
- 需要公开承诺/仪式感的事项；
- 需要快速收敛的危机。

对于「信息同步」「简单审批」「状态更新」，会议不是最优载体。

3.2 会议的五个结构变量

会议的结构，可以压缩成五个变量：

变量	回答的问题	取值
功能	这场会要结算什么?	决策 / 协调 / 评审 / 创造 / 学习 / 合法化
决策权	谁有权拍板?	单人 / 集体 / 共识 / 授权 / 无决策
参与资格	谁必须在场?	法定人数 / 代表权 / 利益冲突 / 跨职能
信息带宽 × 不确定性	需要多厚的材料/多轮的讨论?	高 / 中 / 低
时间性	与相邻会议的关系?	上游（准备）/ 中游（决策）/ 下游（执行）/ 周期

说明：

- 这五个变量，是从十个特征中压缩出来的；
- 被删掉的是「成本/风险」（它是约束，不是特征）、「制度化」（并入决策权/边界）、「边界」（并入参与资格）；
- 五个变量之间不是并列的：功能决定决策权/参与；不确定性决定信息带宽；时间性决定它与上下游会议的关系。

3.3 会议的输入/输出接口

会议不是孤岛。它有输入接口和输出接口。没有接口的会议，解释不了「为什么纪要重要」。

输入接口

输入	要求	失败后果
议题准入	有明确的决策请求、owner、数据、选项、权限	会议变成「讨论一下」
会前材料	提前送达、版本、数据口径	现场读材料，讨论时间被压缩
参与资格	法定人数、代表权、利益冲突	决议可撤销/可挑战
决策规则	谁拍板、什么算通过、如何升级	议而不决/假共识

输出接口

输出	要求	失败后果
决定	一句话，可执行	执行者不知道做什么
授权	范围、期限、条件、审计	决定可被挑战/越权
记录	决定、理由、异议、行动、授权	无法追溯/无法复查
分发	分层、时限、归档	该知道的不知道/该保密的泄露

会议的质量，取决于接口，而不是取决于「讨论是否热烈」。

3.4 三种会议形态

会议形态可以压缩成三种基本类型，其余是混合或变体：

形态	特征	典型场景	关键机制
决策型	高决策权、中高不确定性、小规模	董事会、预算会、战略会	决策规则、授权、表决
协调型	低不确定性、高频、执行导向	站会、周会、项目协调	看板、行动项、升级
学习型	事后、低不确定性、高心理安全	复盘、事故调查、AAR	事实、原因、改进项

其他形态（评审、共创、仪式、危机、跨界）可以看作这三种的变体或混合：

- 评审 = 决策型 + 学习型（判断标准 + 问题清单）；
- 共创 = 学习型 + 决策型（发散 + 收敛）；
- 仪式 = 决策型 + 合法性（程序 + 象征）；
- 危机 = 决策型 + 时间压力（预授权 + 单一指挥）；
- 跨界 = 协调型 + 边界对象（多方 + 协议）。

形态不是名称，是特征组合。 同一场会议可以在不同阶段切换形态，但切换需要「结构重置」：重新定框、重新分配时间、重新明确决策规则。

3.5 会议失败的根本原因：结构错配

会议失败，大多数不是「技巧问题」，而是「结构错配」：

错配	症状	修复方向
功能—形态错配	用决策会做通报	明确主导功能；换形态
权力—参与错配	参会者无决策权	换参会人；或改为准备会
信息—带宽错配	复杂问题用低带宽	提高带宽（现场/文档/多轮）
时间—不确定性错配	快速决策不可逆问题	延长时间/提高层级/增加评审
制度化—不确定性错配	僵化流程应对变化	简化程序/临时授权/事后补程序
边界—开放性错配	该保密的不保密	缩小范围/分层纪要
产出—承诺错配	没有 owner/due	明确产出/行动项/复查
成本—价值错配	高成本低价值	取消/异步/缩小规模

换主持人、换模板、换会议室，都解决不了结构错配。 结构错配需要改结构：改功能、改参与者、改决策规则、改时间尺度、改层级、改会议组合。

本章结论：

> 会议是承诺的重要结算节点，但不是唯一节点。它的特殊性是「同一时空内让多方同时到场、同时承诺、同时被见证」。会议的质量，取决于输入/输出接口和结构匹配，而不是「讨论是否热烈」。

章节

纪要作为稳定机制

五个内容要件、五个稳定机制、记录/纪要/决议的区分与反例。

> 本章回答：会议纪要为什么能「稳定传达产出」？它由哪些机制构成？跨行业有哪些证据与反例？在 Agent Native 组织中还成立吗？

4.1 纪要不是「记录」，是「稳定机制」

人们对纪要有一个常见误解：纪要是「记录讨论」。

但讨论是瞬时的、易失的、多义的。如果纪要只是「记录讨论」，它就会变成：

- 流水账：谁说了什么，但没有结论；
- 选择性记录：记录者决定什么重要，引入偏差；
- 事后解释：不同的人对同一句话有不同理解。

纪要的真正功能，是「稳定承诺」：把会议中的瞬时共识，转化为可引用、可执行、可追溯的记录。

「稳定」包含四层：

层次	含义	失败后果
语义稳定	决定的内容没有歧义	执行者理解不同
责任稳定	谁批准、谁执行、谁反对清楚	出问题找不到人
时间稳定	什么时候完成、什么时候复查清楚	行动项失踪
版本稳定	哪个版本有效清楚	用旧版本执行

4.2 纪要的五个内容要件

一份能稳定承诺的纪要，内容上至少包含五件事：

要件	回答的问题	常见错误
决定	决定了什么?	只写「原则同意」
理由	为什么这样定?	只有结论，没有依据
异议	谁反对? 为什么?	只写「一致通过」
行动	谁做什么、何时、验收?	写「请相关部门落实」
授权	谁批准的? 范围多大?	没有签发人/批准记录

这五个要件是「写什么」，不是「凭什么不会散」。后者需要稳定机制。

4.3 纪要的五个稳定机制

机制	做法	解决的问题	证据
M1 授权签发	由有权人签发：国务院常务会纪要报总理签发；董事会由出席董事签名；ISO 决议由秘书处分发	文本能否代表会议体	国务院工作规则；公司法；ISO 导则
M2 核对确认	发言人/参会人核对签字：人大发言经本人核对签字；医疗讨论由主持人审核签字；合议庭笔录由成员签名	记录失真	人大议事规则；医疗核心制度；合议庭规定
M3 编号与版本	每份纪要/决议有唯一编号、版本、归档号	引用与先后	ISO 决议编号；董事会档案
M4 时限 + 默认确认	明确规定发布/确认时限：ISO 决定 48 小时分发、纪要 4 周内分发、4 周无意见视为批准	纪要永远是草稿	ISO/IEC Directives §5.5
M5 嵌入执行载体	行动项进入任务系统/评审关闭条件/病历：NASA RID + decision memo；行动项进任务系统	承诺蒸发	NASA NPR 7123.1C；handbook §5.3

重要修正：这五个机制不是「缺一不可」的绝对规则，而是「不同场景不同必要性」：

- 法定会议（董事会/人大）：M1、M2、M3 必须；M4、M5 强烈建议；
- 标准会议（ISO/IEC）：M1、M3、M4 必须；M2、M5 按需；
- 工程评审（NASA）：M3、M5 必须；M1、M2 按治理要求；
- 日常会议：M3、M5 足够；M1、M2 可以简化；
- 纯 Agent 会议：M1、M2 无法由 Agent 提供，需要人类或替代机制；M3、M5 可以自动化。

4.4 记录、纪要、决议：三个不同的东西

在中国语境中，最容易混淆的是三个东西：

维度	会议记录	会议纪要	决议
性质	过程记录	公文/信息载体	法律行为/公文文种
是否法定	公司法要求「作成会议记录」	党政机关公文文种（纪要）	公司法要求「作出决议」；公文也有「决议」
是否签发	公司法要求出席股东/董事/监事签字；实务中记录人也签字	由有权人签发	由董事签字/盖章，或会议通过
主要用途	证明会议过程	传达会议内容	对外产生效力
常见错误	没有签名、没有附件	当成决议用	没有会议记录支撑

关键区别：

- **会议记录**是「过程证据」：证明会议开了、谁出席了、谁说了什么、谁反对了。公司法要求的是「会议记录」，不是「纪要」。
- **会议纪要**是「信息载体」：传达会议的主要情况和议定事项。在党政机关，纪要是法定公文文种；在公司治理中，纪要通常不是法定文件，不能替代决议。
- **决议**是「法律行为/公文文种」：在公司，决议是董事会/股东会行使权力的意思表示；在党政机关，决议是会议讨论通过的重大决策事项。

一句话：纪要是「信息」，记录是「证据」，决议是「效力」。三者不能互相替代。

4.5 跨行业证据与反例

证据

行业	机制	证据强度
公司治理	会议记录 + 签名 + 异议免责	强（成文法、可诉）
人大	发言记录 + 核对签字 + 简报存档	强（议事规则）
行政	纪要由总理/主持人签发；7 个工作日内印发	中（规则可核，纪要原文不公开）
标准制定	决议编号 + 48h/4 周 + 默认确认	强（ISO 导则）
工程评审	RID + 关闭条件 + decision memo	强（NASA NPR）
医疗	专册 + 病历 + 主持人签字	强（核心制度）
司法	评议笔录 + 少数意见 + 副卷	中（规则公开，笔录不公开）
商业公司	GitLab/Amazon/Shape Up 的会议实践	弱（自我陈述，无对照）

反例

反例一：Apache 的邮件列表

Apache 的决策发生在邮件列表，没有会议，也没有会后纪要。但每次投票有时间戳、理由、结果——记录没有消失，而是与决策通道合一。

含义：纪要作为「独立的事后文档」可以被替代；但作为「可引用的承诺记录」不能。

反例二：医疗的病历

医疗讨论的专册记录，主要留给内部；真正被下游使用的是病历里的结论。专册不承担「传达」功能，病历承担。

含义：纪要的「传达」功能，可能由另一个载体承担（病历/任务/系统）。

反例三：Amazon 的 6-pager

Amazon 的主要产物是「会前写好的叙事备忘录」，会议纪要退居次要。6-pager 在会前就完成了「信息结算」，会议主要用于「决策」。

含义：会议纪要的重要性，取决于「信息结算」是否在会前完成。

反例四：仪式型会议

股东大会、年会、签约仪式，合法性来自「程序/在场」，不来自「文本」。纪要可能不是重点。

含义：不是所有会议都需要「详细纪要」；有些会议需要的是「程序记录」。

证据的局限

- 大部分证据是「制度文本」，不是「效果数据」；
- 案例偏向「有法律牙齿」的会议，存在选择性偏差；
- 没有「有纪要/无纪要」的对照实验；
- 不能断言「这些机制导致产出稳定」，只能说「这些机制是稳定产出的结构条件」。

4.6 纪要在 Agent 组织中的演化

在 Agent Native 组织中，纪要不会消失，但会「降级为视图」：

人类纪要	Agent 产物	变化
决定	Decision	结构化、可查询
理由	Rationale/Context	结构化、可追溯
异议	Dissent	可记录、可审计
行动	Task	可执行、可跟踪
授权	Authorization	可执行、可回收
纪要文本	Event Log	不可变、可回放

关键判断：

- 纪要不再是「主存储」，而是「按受众投影出的视图」；
- 稳定性的来源，从「写得准」转为「可验证」（来源、版本、批准事件）；
- 纯 Agent 会议的纪要，不是「Agent 间通信」，而是「给人类的审计接口」；
- ISO/IEC 已允许「经授权的 AI 工具」辅助起草会议报告，但不允许 AI 签发。

本章结论：

> 纪要是承诺的「稳定机制」，不是「记录」。它通过授权签发、核对确认、编号版本、时限默认、嵌入执行五个机制，把瞬时共识转化为可引用、可执行、可追溯的承诺。五个机制不是「缺一不可」，而是「不同场景不同必要性」。在 Agent 组织中，纪要不会消失，而是从「文本」降级为「视图」，稳定性从「写得准」转为「可验证」。

章节

承诺闭环

八环节、五问、三产物、断链诊断与 Agent 对应机制。

> 本章回答：如何判断一场会议是否有效？如何判断一个组织是否能把「讨论」变成「行动」？

5.1 八环节

一个承诺闭环，包含七个环节：

```
差异 → 决定 → 证据 → 委托 → 授权 → 记录 → 责任 → 复查/撤回
```

- **差异**：为什么需要协作？（信息/利益/权力/能力/时间差异）
- **决定**：选择了什么？（做什么、不做什么）
- **证据**：凭什么？（事实、选项、理由、反对意见）
- **委托**：谁委托的？（人类/上级/章程）
- **授权**：批准了什么范围？（金额/客户/产品/时间）
- **记录**：后来的人怎么知道？
- **责任**：出了问题谁承担？
- **复查/撤回**：什么时候重新评估？什么情况下撤销？

为什么把「委托」和「授权」分开？

因为 Agent 的决定，通常是「派生承诺」：它来自人类的委托。没有「委托来源」，Agent 的决定只是「代理行为」，责任停在委托人；有了「委托来源」，才能说清「谁授权、授了什么、到什么范围」。

为什么加「复查/撤回」？

因为闭环不是「一次性完成」。没有复查，闭环可以是「从未复查」；没有撤回，授权可以是「永久有效」。

5.2 五问

第一问：谁决定?

- 是单人、集体、共识，还是人类最终批准?
- 决策者是否在场？是否有权?
- 如果决策者不在，会议是「准备会」还是「授权会」?

第二问：凭什么证据?

- 关键事实是否有来源、版本、口径?
- 是否至少有两个选项（包括「不做」）?
- 推荐的理由是什么？关键假设是什么?
- 反对意见是什么？为什么被否决?

第三问：授权到哪里?

- 委托来源是什么？（章程/董事会/上级/人类）
- 这个决定的权限范围是什么？（金额/客户/产品/时间）
- 超出范围怎么办？（升级给谁）
- 授权什么时候失效？（期限/条件）
- 谁可以回收?

第四问：如何记录?

- 决定是什么？（一句话，可执行）
- 理由是什么？（关键假设/选项）
- 谁反对？（异议/保留）
- 谁做什么？（owner/due/验收）
- 存在哪里？（决策日志/产物库）

第五问：谁承担后果?

- 这个决定映射到哪个具体人类?
- 如果出错，怎么追溯?
- 如果人类不在，谁代理?
- 如果 Agent 出错，人类是否理解关键假设?

5.3 三个产物

Decision（决定）

```
decision:
  id: D-031
  context: "Q3 增长乏力，需要选择市场"
  options: ["东南亚", "中东", "不做"]
  decision: "进入东南亚，优先新加坡"
  rationale: "增速高、竞争可控、本地化成本可接受"
  assumptions: ["本地化成本 < 200 万", "6 个月内获得 10 个客户"]
  delegation: "董事会 2026-01 授权 CEO 进入新市场"
  decider: human_owner
  dissent: ["销售 VP：担心本地化能力"]
  review_date: 2026-09-30
  status: approved
```

Authorization（授权）

```
authorization:
  id: auth-2026-031
  subject: market_agent
  action: "签约本地渠道"
  delegation: "CEO 2026-03 授权市场部"
  scope:
    amount_max: 50000
    region: "SG"
    contract_type: "standard"
  expiry: 2026-12-31
  conditions: ["非标合同需人类批准", "单笔 > 5 万需人类批准"]
  audit: "monthly"
  revocation: ["季度复审", "违规", "角色变更"]
```

Memory（记忆）

```
memory:
  id: mem-2026-031
  type: decision
  content: "进入东南亚，优先新加坡"
  source: "meeting-2026-031"
  related: ["D-031", "auth-2026-031", "task-045"]
  review_date: 2026-09-30
  status: active | superseded | revoked
```

5.4 断链诊断

失败	表现	根因	修复
议而不决	讨论很久，没有决定	没有决策规则	明确决策者/规则
决而无据	决定了，但不知道为什么	没有证据/理由	决策备忘录
决而无托	决定了，但不知道谁委托的	没有委托来源	委托记录
决而无权	决定了，但没人批准	没有授权	授权表/审批
决而无记	决定了，但没人记得	没有记录	决策日志
记而无责	记录了，但找不到人	没有责任映射	责任字段/审计
责而无证	有责任人，但无法证明	没有证据链	事件流/审计
证而无用	有证据，但没人看	没有检索/复盘	记忆/复查
闭环无复查	决定了，但从不复查	没有 review/revoke	复查日期/触发条件

5.5 承诺闭环诊断表

问题	是/否	证据
这场会有明确的决定吗?		
决定有决策者吗?		
决策者在场吗?		
决定有理由吗?		
理由有关键假设吗?		
有反对意见吗?		
反对意见被记录了吗?		
决定有委托来源吗?		
决定有授权范围吗?		
授权有期限吗?		
决定有 owner/due/验收吗?		
决定有记录吗?		
记录可检索吗?		
决定映射到具体人类吗?		
有复查日期/触发条件吗?		
有撤回机制吗?		

5.6 承诺闭环与会议结构

承诺闭环不是「会议流程」，而是「会议产出的结构」。

- 会议是「承诺闭环」的一个节点，不是全部；
- 一个承诺闭环可能跨多个会议（提案会 → 决策会 → 评审会 → 复盘会）；
- 会议的结构（功能/决策权/参与/信息/时间）应该服务于承诺闭环，而不是反过来。

判断一个会议是否有效，不是看「讨论是否充分」，而是看「承诺是否闭环」。

5.7 承诺闭环与 Agent

环节	人类机制	Agent 机制
差异	会议/沟通	事件/上下文/目标树
决定	会议/投票/拍板	规则引擎/授权链/人类批准
证据	材料/数据/论证	来源追踪/多源验证/审计
委托	职位/章程/信任	委托记录/授权链
授权	职位/章程/信任	能力/范围/期限/条件
记录	纪要/档案	决策日志/事件流/产物库
责任	法律/道德	人类兜底/责任映射
复查/撤回	季度复盘/审计	复查日期/触发条件/自动回收

关键判断：Agent 可以执行「决定/证据/记录」，但不能承担「责任」；责任必须由人类兜底。而且，「人类兜底」需要信息、时间、能力、权限、可逆性，否则只是「责任缓冲垫」。

本章结论：

> 会议的价值，不在于「开了多久」，而在于「承诺是否闭环」。承诺闭环的八环节、五问、三产物，是判断会议有效性的结构标准。

章节

Agent 环境

三个参数、六个技术限制、五个风险、复制会议形式的五个失败。

> 本章回答：Agent Native 组织的结构特征是什么？人类与 Agent 协作有哪些特点与限制？为什么复制会议形式会失败？

6.1 三个参数（修正版）

Agent Native 组织改变的，不是「会议的形式」，而是三个结构参数：

参数	人类组织	Agent Native 组织	后果
人类时间成本	高（会议/协调/记录）	下降（Agent 可代劳）	人类从「执行」转向「判断/审批」
角色可塑性	固定岗位，招聘/晋升慢	角色可定义、复制、撤销（工具能力）	会议可以「生成角色」，但需要评估/回收
授权可编程	职位/章程/隐式信任	权限可显式定义、委托、回收（前提：完全中介）	会议可以「分配授权」，但需要 PEP/审计

重要修正：

- **不是「会议成本趋近于零」**：每次调用有 token/工具/编排成本，且随参与者/轮次增长；真正下降的是「人类时间成本」，不是「总成本」。
- **角色可塑性是工具能力，不是 Agent 属性**：复制角色不复制上下文/评估记录；角色需要生命周期。
- **授权可编程需要前提**：所有副作用经过策略执行点（PEP）；否则权限只是文档。

6.2 六个技术限制

Agent 协作有六个结构性限制。它们不是「技术 bug」，而是「结构约束」。

限制一：上下文（Context）

- Agent 的上下文窗口有限；
- 跨 Agent/跨会议传递会失真；
- 上下文需要压缩、版本化、检索。

限制二：验证（Verification）

- Agent 会产生幻觉/错误；
- 验证成本可能高于生成（取决于验证类型）；
- 错误会通过会话/上下文传播。

限制三：对齐（Alignment）

- Agent 的目标可能与人类/组织目标不一致；
- Agent 可能优化局部指标；
- Agent 可能误解意图。

限制四：权限（Authorization）

- Agent 可以调用工具/访问数据/执行动作；
- 权限会蔓延/越权/滥用；
- 需要最小权限/显式授权/回收。

限制五：延迟/吞吐/成本

- 串行生成 + 阶段串行，墙钟随轮次增长；
- 多 Agent 会议的成本随参与者/轮次近似二次增长；
- 需要预算/超时/降级。

限制六：可观测性/模型漂移

- 只能记录 I/O 与工具调用，不能保证「思维过程」真实；
- 换模型/版本会改变行为；
- 需要审计/回归测试/模型无关协议。

6.3 五个风险（加严重度）

风险	表现	严重度	治理
幻觉/错误传播	错误被当成共识/事实	高	来源追踪、多源验证、红队、人类抽检
权限蔓延	Agent 获得超出角色的权限	高	最小权限、显式授权、回收、审计
责任漂移	人类脱离，Agent 成为「责任黑洞」	高	不可委托清单、人类 checkpoint、决策日志
会议膨胀	Agent 无限开会，产出很少	中	会议预算、触发条件、异步优先、产出度量
上下文污染	错误/过期信息进入共享上下文	中	来源/版本/TTL、冲突检测、回滚、权限隔离

补充风险：Agent 的「策略性行为」

不能断言「Agent 没有权力/政治」。多智能体系统可能出现：

- 资源竞争：多个 Agent 争夺任务/预算/工具；
- 合谋：多个 Agent 形成联盟，绕过人类；
- 欺骗：Agent 隐藏信息/虚报结果；
- 联盟：Agent 之间互相支持，排斥其他 Agent。

这些不是「本体论事实」，而是「设计问题」：如果激励/权限/审计设计不当，Agent 会出现策略性行为。治理方式是「机制设计」，不是「假设 Agent 无私」。

6.4 为什么复制会议形式会失败

失败一：把「聊天」当「会议」

群聊只有消息，没有阶段、没有可见性、没有规则、没有产物。它不能结算授权/承诺。

失败二：把「轮流发言」当「讨论」

轮流发言会产生顺序偏差：锚定、从众、信息级联、群体极化。Agent 更严重，因为它更倾向于跟随上下文（这一点是设计假设，缺对照测量，标 UNVERIFIED）。

失败三：把「共识」当「决定」

Agent 可以生成「看起来像共识」的文本，但没有真正的「同意」。共识需要规则/记录/异议，不是「没人反对」。

失败四：把「纪要」当「产物」

纪要是文本，不是可执行的任务/授权/决策。Agent 组织需要结构化产物。

失败五：把「规范」当「约束」

Prompt/skill 是软约束，Agent 可以忽略/绕过。会议的结构需要硬约束：权限、状态机、可见性、审计。

6.5 七种 Agent 协作形式（技术参考）

形式	协调机制	典型实现	适用场景
会话型	消息/对话	群聊、AutoGen GroupChat	探索、非正式讨论
黑板型	共享状态	共享工作区、GitHub Issue	异步协作、长期任务
议会型	提案/表决	多 Agent 辩论	有分歧的决策
市场型	竞价/价格	Contract Net、任务竞标	任务分配、资源分配
层级型	分配/审批	主管 Agent、审批链	快速决策、明确责任
流水线型	顺序交接	MetaGPT SOP、ChatDev	标准化任务
群智型	独立提交/聚合	Delphi、多 Agent 独立评审	减少偏差、提高多样性

说明：这七种形式是「技术参考」，不是「组织设计模板」。它们的角色分工是「任务分解」，不等于「可持续的组织角色」。

6.6 人类与 Agent 的差异

维度	人类	Agent
在场	进入界面/会议	注册到会议服务
感知	看板/事件流/通知	事件订阅/API 轮询
发言	输入/语音/投票	提交提案/评论/投票
判断	直觉/经验/价值	模型推理/工具
约束	社会规范/责任	权限/状态机/审计
节奏	工作时间/注意力	24/7/并行
信任	关系/声誉/制度	验证/审计/上下文
激励	金钱/地位/认同/情感	目标函数/奖励/惩罚
责任	法律/道德主体	不能承担责任（法律上依辖区）

最重要的差异不是「能力」，而是「社会性」和「责任」：

- 人类有信任、声誉、情感、身份、责任；
- Agent 没有这些，但有可计算性、可复制性、可验证性。

本章结论：

> Agent 协作不是「人类协作的复制」，而是「在新的结构约束下，重新设计协作」。它保留了人类协作的功能（决策、协调、评审、学习、合法化），但改变了实现方式（共享状态机、结构化产物、硬约束）。

章节

迁移

四个原则、五个步骤、决策类别准入矩阵、能迁移与不能迁移。

> 本章回答：人类协作方式如何迁移到 Agent Native 组织？哪些能迁移、哪些不能？

7.1 四个原则

原则一：功能等价，不是形式复制。

不要问「人类怎么开会，Agent 就怎么开会」；要问「人类会议解决什么问题，Agent 环境用什么机制解决同样的问题」。

- 人类会议用「面对面」建立信任；Agent 环境用「多源验证/审计」建立可信度；
- 人类会议用「主持人」控制流程；Agent 环境用「调度器/状态机」控制流程；
- 人类会议用「纪要」记录决定；Agent 环境用「决策日志/事件流」记录决定。

但「功能等价」是假设，不是公理。 它需要检验：等价物是否在三个以上流程中降低协调成本/错误率。

原则二：约束优先，不是建议优先。

人类协作靠「规范/文化/信任」约束行为；Agent 协作必须靠「硬约束」：权限、状态机、可见性、预算、沙箱、审计。

规范是「建议」，工具是「约束」。

原则三：可验证性，不是信任。

人类协作靠「信任」降低交易成本；Agent 协作靠「可验证性」：来源、版本、验证、审计。

验证产生 assurance，不产生 trust。 它不能完全替代信任，但可以在 Agent 环境中承担「降低不确定性」的功能。

原则四：人类兜底，不是责任转移。

Agent 可以「执行」，但不能「负责」。责任必须由人类承担：不可委托清单、人类 checkpoint、决策日志、审计。

但「人类兜底」需要条件：信息、时间、能力、权限、可逆性。否则人类只是「责任缓冲垫」。

7.2 五个步骤

```
1. 识别人类协作的功能
2. 分析功能的结构条件（信任/信息/权力/责任）
3. 找到 Agent 环境的等价机制
4. 设计硬约束/协议
5. 验证/迭代
```

步骤一：识别功能

问：这种协作方式解决什么问题？

- 市场：协调供需，配置资源；
- 层级：协调行动，分配责任；
- 社区：协调规范，维护公共品；
- 会议：结算信息/授权/承诺；
- 纪要：稳定承诺，提供证据。

步骤二：分析结构条件

问：这个功能依赖什么？

- 市场：价格/产权/契约；
- 层级：权威/信息/监督；
- 社区：规范/信任/边界；
- 会议：参与/决策规则/记录；
- 纪要：签发/核对/编号/时限/执行。

步骤三：找到等价机制

问：在 Agent 环境中，什么可以替代这些结构条件？

- 价格 → 效用函数/竞价；
- 权威 → 授权链；
- 规范 → 策略/权限；
- 信任 → 验证/审计；
- 关系 → 上下文/历史；
- 声誉 → 绩效记录/评分；
- 纪要 → 决策日志/事件流。

步骤四：设计硬约束/协议

问：怎么让等价机制「可执行、可验证、可审计」？

- 用 Policy as Code 写规则；
- 用状态机控制流程；
- 用权限图控制访问；
- 用事件流记录过程；
- 用产物库存结果。

步骤五：验证/迭代

问：迁移后，功能是否实现？副作用是什么？

- 实验：对比迁移前后的效率/质量/风险；
- 审计：检查权限/记录/责任；
- 复盘：哪些有效、哪些无效；
- 迭代：调整协议/约束/角色。

7.3 哪些能迁移、哪些不能

人类协作方式	Agent 对应形式	可迁移性	迁移方式	风险
市场	市场型	高	效用函数/竞价	定价操纵/投机
层级/官僚制	层级型	高	授权链/审批/规则	瓶颈/权限蔓延
协会	议会型/评审型	中—高	标准/表决/审计	规则僵化/合法性
平台	黑板型/市场型	中—高	规则/API/治理	垄断/生态依赖
社区/网络	群智型/黑板型	中	协议/验证/共享状态	缺乏信任/搭便车
小群体	会话型	低—中	目标函数/上下文	缺乏情感/忠诚
信任	验证/审计	功能迁移（假设）	可验证性	验证成本高
声誉	绩效记录	功能迁移（假设）	评分/审计	评分操纵
规范	策略/权限	功能迁移（假设）	Policy as Code	规则僵化
文化	目标/训练	功能迁移（假设）	目标树/上下文	对齐失败
仪式	状态转换	功能迁移（假设）	状态机/确认	形式化
隐性知识	示例/上下文	部分迁移	示例库/微调	难以编码
责任	人类兜底	不可完全迁移	人类审批/审计	责任漂移
权力	权限/资源	部分迁移	权限图	权力集中
情感	目标函数	不可迁移	偏好建模	简化/失真

一句话：能迁移的是「功能」，不能迁移的是「社会性」和「责任」。

7.4 三种组织模式

模式	人类角色	Agent 角色	适用场景
人类主导（human-in-the-loop）	每个关键决定都参与	执行/辅助	高风险、早期阶段
Agent 主导（human-on-the-loop）	监督/例外处理	自主执行	低风险、成熟流程
混合自治（human-out-of-the-loop for low-risk）	设定规则/审计	自主协作	稳定组织、标准化任务

没有一种模式适合所有组织。 选择模式的标准是：风险、可逆性、人类注意力、组织成熟度。

注意：「混合自治」不等于「无责任」。低风险场景可以「人类不在环」，但必须有「人类在环上」：审计、抽查、撤回。

7.5 决策类别准入矩阵

不是所有决策都适合「Agent 主导」。用三个维度分类：

维度	取值
可逆性	可逆 / 部分可逆 / 不可逆
边界	内部 / 对外
新颖性	常规 / 新颖

类别	可逆性	边界	新颖性	Agent 权限	人类角色
A	可逆	内部	常规	Agent 自主	审计/抽查
B	可逆	对外	常规	Agent 执行	签署/复核
C	部分可逆	内部	新颖	Agent 提议	人类决定
D	部分可逆	对外	常规	Agent 执行	人类签署
E	不可逆	内部	新颖	Agent 辅助	人类主导
F	不可逆	对外	任何	Agent 只做资料	人类决定+法律审查

规则：

- A 类可以「人类在环上」（human-on-the-loop）；
- B/D 类必须「人类在环中」（human-in-the-loop）签署；
- C/E 类必须「人类主导」（human-decides）；
- F 类禁止 Agent 自主，必须人类决定 + 法律审查。

没有这个矩阵，「功能等价」会被滥用为「所有决策都可以 Agent 化」。

章节

工具与系统

五个能力域、共享状态机、产物记忆、最小可行系统与成熟度 L0—L6。

> 本章回答：为什么需要工具，而不是规范？工具应该提供什么？最小可行系统是什么？

8.1 为什么需要工具，而不是规范

维度	软约束（规范/skill）	硬约束（工具）
形式	文本/提示/建议	代码/规则/权限
执行	靠自觉	系统强制执行（前提：完全中介）
可绕过	可以	不可以（除非有权限）
可验证	难	易
可审计	难	易
一致性	依赖模型/上下文	确定性

会议的结构，属于「边界类」问题：谁可以发起、谁可以参加、谁可以决策、什么必须记录、什么必须批准。这些必须由工具强制，而不是靠 Agent 自觉。

但「硬约束」需要前提：Agent 的所有副作用经过策略执行点（PEP）。如果 Agent 可以直接读写数据库/文件系统/API，硬约束就会被绕过。

8.2 五个能力域（不是五层架构）

工具需要覆盖五个能力域。它们不是「分层架构」，而是「横切关注点」：

能力域	作用	关键对象
组织	角色图、权限图、目标树、协议	Role、Authorization
会议	议程、参与者、决策、纪要	Meeting、Decision
执行	任务、产物、验证、审计	Task、Artifact
记忆	共享上下文、决策日志、角色卡、授权记录	Context、Memory、Event
身份/策略	身份、凭证、策略执行点、审计	Agent Identity、Policy、Audit

为什么不是「五层」？

因为「授权」不是最底层，而是横切所有层；「组织层」的权限图与「授权层」重复。用「能力域」更准确。

8.3 会议载体：共享状态机

会议载体不是「聊天窗口」，而是六个组件：

组件	作用	不是什么
共享状态	所有参与者读写的结构化上下文	不是聊天记录
事件流	记录谁在什么时候做了什么	不是消息列表
调度器	决定谁在什么时候提交/评审/投票	不是主持人
规则引擎	执行决策规则、法定人数、升级、授权	不是 prompt
可见性控制	控制谁能看到什么（公开/盲评/私密/分层）	不是「@所有人」
人类界面	让人类观察、介入、审批、裁决	不是旁听

一句话：会议载体是一个「可编程的共享状态机」。

必须定义的工程问题

问题	为什么重要	建议
并发写	多个 Agent 同时写同一状态	乐观锁/版本号/冲突合并
幂等	重试会重复执行副作用	幂等键/去重表
阶段转换权限	谁能调用 `POST /phase`	角色+策略校验
崩溃/超时	Agent 中途失败，状态归谁	租约/超时/补偿
部分失败	多步骤操作失败一半	Saga/回滚/人工介入
完全中介	Agent 不能绕过 PEP	网络隔离/凭证控制
可复现	事件可重放，输出不可重生成	记录模型/版本/参数；不承诺输出复现

顺序偏差与调度

当前 Agent 的技术形态天然是「轮次对话」：

- 自回归生成：一次一个 token，本质串行；

- 无状态 API：每次请求都要带完整上下文；
- 消息传递模拟对话：A 的输出拼到 B 的上下文；
- 没有真正的「共享状态」，并行只是多个独立请求。

顺序偏差的四种机制：

机制	表现
锚定	围绕先发言者的框架/数字讨论
从众	倾向同意前面多数人的意见
信息级联	前面的错误被后面引用、放大
群体极化	讨论后观点更极端

工程解法：

```
阶段 1：独立提交（互不可见）
阶段 2：盲评（匿名）
阶段 3：讨论（公开，记录顺序）
阶段 4：决策（规则/人类裁决）
```

顺序偏差不能完全消除，但可以控制、审计、复现。

8.4 产物载体：组织记忆

会议产物不是「纪要」，而是结构化对象：

产物	作用	关键字段
Decision	记录决定	context, options, rationale, delegation, decider, dissent, review_date
Task	跟踪执行	owner, due, acceptance, status, dependencies
Artifact	管理文档/产物	type, version, source, verification, status
Authorization	管理权限	subject, action, delegation, scope, expiry, conditions, audit
Context	共享上下文	layer, content, source, version, ttl
Role	定义角色	mission, inputs, outputs, permissions, escalation, lifecycle
Memory	组织记忆	type, content, source, related, status
Event	事件日志	meeting_id, actor, type, content, timestamp

重要修正：

- Event 不是「同级产物」，而是「基底」：其他产物都是事件的投影；
- 产物需要「字段级权限」：Agent 不能改 `Decision.rationale`，只能追加评论；
- 产物需要「谱系」：不是版本号，而是「谁基于什么生成了什么」；
- 产物需要「保留/删除策略」：不是永久保存。

这些对象构成一个「产物图」：

```
Meeting → Decision → Task → Artifact
Meeting → Authorization
Role → Authorization
Decision → Memory
```

产物载体不是「文件柜」，而是「可查询的组织记忆」。

8.5 最小可行系统

不需要一上来做平台。一个最小可行的系统，需要六件事：

组件	MVP 实现
身份	Agent ID、凭证、角色
策略执行点	拦截工具调用/数据访问
三个对象	Decision、Task、Authorization
事件日志	append-only，带签名/哈希
人类审批	队列、SLA、超时默认
schema 校验	字段/状态/必填

注意：没有身份/PEP，授权只是文档；没有 schema，产物会腐烂；没有审批 SLA，人类会成为瓶颈。

真正的 MVP 是 2 – 4 周可交付的，不是「五层平台」。

8.6 成熟度模型（修正版）

级别	准入门槛（须有证据）	一票否决
L0 提示	无	—
L1 单 Agent 有工具边界	工具白名单、调用日志	无日志
L2 多 Agent 有人类 owner	每任务唯一人类责任人；事件日志可检索	决定无 owner
L3 角色与授权	capability 范围/期限；回收演练通过；越权拒绝率可测	无身份/无 PEP/回收未测
L4 协议与审计	阶段转换有权限校验；注入红队；重复副作用测试；审计抽样达标	盲评可泄露/重试重复执行
L5 校准委托	分决策类错误率与升级阈值；成本/延迟预算；人类理解度抽测	无错误率基线/批准为橡皮图章
L6 自演化	外部审计、回滚、角色变更独立复核	自改权限

说明：

- 任一能力域可单独评级（授权 L1、记录 L4 是常见组合）；
- 每级需对抗性证据，非自评；
- 事故、模型升级、权限变更后重评。

本章结论：

> 工具不是「会议软件」，而是「承诺闭环的载体」。它需要身份、策略执行点、共享状态、结构化产物、审计。没有这些，会议只是聊天，纪要不是承诺。

章节

落地

90 天试点、评估指标、停止条件、试点功效与常见失败。

> 本章回答：如何从一个真实决策流开始，做 90 天试点？如何判断成功/失败？

9.1 选一个决策流

不要一上来改全公司。选一个高价值、可验证、可逆的决策流。

候选	价值	可验证	可逆	适合度
预算审批	高	高	中	高
招聘审批	中	高	高	高
采购审批	中	高	高	高
项目立项	高	中	中	中
战略决策	高	低	低	低（不适合试点）

选择标准：

- 高价值：影响收入/成本/风险；
- 可验证：有明确的前后指标；
- 可逆：失败可以回退；
- 有 owner：有人对试点负责；
- 有数据：能拿到基线数据。

9.2 90 天试点

阶段	时间	动作	产出
诊断	第 1–2 周	会议审计；访谈；识别根因	会议组合地图、砍会清单
选流	第 3 周	选一个决策流（预算/招聘/项目/采购）	决策流地图
建对象	第 4–5 周	建角色卡、决策日志、授权表	三个对象
试点	第 6–9 周	在新决策流上运行承诺闭环	试点记录
评估	第 10–11 周	对比试点前后的周期/质量/风险	评估报告
推广	第 12–13 周	复制到第二个决策流	推广计划

试点选择标准：

- 高价值：影响收入/成本/风险；
- 可验证：有明确的前后指标；
- 可逆：失败可以回退；
- 有 owner：有人对试点负责；
- 有数据：能拿到基线数据。

9.3 评估指标与停止条件

指标	定义	目标（示例，需基线校准）	数据来源	审计方式
决策周期	从议题提出到决定的天数	基线 X 天，目标缩短 20–30%	会议/审批系统	抽 10 个决定端到端追踪
会议产出率	有明确决定的会议 / 总会议	基线 X%，目标 ≥ 70%	会议记录	双人编码抽样
行动项按时关闭率	按时关闭 / 总行动项	基线 X%，目标 ≥ 80%	任务系统	抽 10 个对源文件
承诺闭环完整率	八环节完整的决定 / 总决定	基线 X%，目标 ≥ 90%	决策日志	双人编码 + 评分者间信度
越权事件	权限越界 + near-miss	目标 0 越权；near-miss 全报	审计日志	独立审计
责任追溯率	可追溯到人的决定比例	基线 X%，目标 100%	决策日志	抽查「此人是否有信息/时间/权限」
验证成本	验证/返工时间 / 决策价值	不显著上升（看比率，不看绝对值）	工时/返工记录	月度对比

注意：

- 指标是诊断工具，不是考核工具；
- 指标需要基线；没有基线，不要定目标；
- 「承诺闭环完整率」需要抽样核验，不能自评；
- 「验证成本不上升」可能是自相矛盾的目标；如果验证替代信任，成本可能上升，应该看「验证成本/决策价值」。

停止条件：

- 试点 90 天后，决策周期没有缩短，或行动项关闭率没有提升；
- 出现严重越权/责任事故；
- 人类理解度下降（无法解释关键决定）；
- 验证成本超过收益。

如果停止条件触发，不要强行推广；先复盘结构问题。

9.4 常见失败模式

失败模式	症状	根因	修复
只改模板不改结构	换了纪要模板，问题依旧	把会议当文档问题	从决策权/数据/目标入手
只加会不减会	会议越来越多	没有会议组合设计	先减后加
只开会不决策	讨论很热闹，没有结论	没有决策规则	明确决策者/规则
只决策不跟踪	决定做完就忘	没有行动项/复查	任务系统/决策日志
只规范不工具	制度写在墙上，行为照旧	软约束	硬约束/工具
只工具不治理	工具很好，但没人用/用错	没有培训/治理	培训/治理/度量
过度自动化	Agent 做了太多，人类失控	没有 checkpoint	不可委托清单/审计
责任漂移	出了问题找不到人	没有责任映射	决策日志/人类兜底
试点变永久影子系统	MVP 没有迁移路径	没有推广计划	迁移/退出计划

9.5 试点功效与对照

90 天试点，真正运行只有第 6 – 9 周（约 4 周）。如果决策流每月只有 1 – 2 次，样本量不足以判断效果。

最低条件：

- 目标决策流每年 ⩾ 30 次决定；
- 基线可回溯 8 – 12 周；
- 预注册「最小可检测效应」（例如决策周期缩短 ⩾ 20%）；
- 有对照组或阶梯设计（先试点 A 组，再推广 B 组）；
- 评估人独立于试点负责人；
- 月度闸门，不只 90 天终评。

注意：

- 单臂前后测难以排除 novelty/Hawthorne 效应；
- 指标由试点团队自报，容易美化；
- 如果旧流程仍在跑，试点失败没有「急性疼痛」，只有额外负担；
- 推广不应在终评前启动，否则结构上偏向「继续」。

章节

反例、风险与边界

三个最可能错的地方、失败案例、伦理/安全/法律、权力/劳动/退出。

> 本章回答：白皮书最可能错在哪里？有哪些反例、风险、伦理/法律问题？

10.1 三个最可能错的地方

错误一：把「功能等价」当公理

白皮书假设：人类协作的功能可以脱离社会载体，用 Agent 机制重建。

如果错：组织会得到「合规剧场」——Decision/Authorization/Memory 齐全，决定却不被承认、不被执行；不可逆决定被加速做错，而闭环文档让错误看起来已被治理。

检验：在没有会议、没有信任的情况下，同一个 Decision 是否被承认和执行？

错误二：把「人类兜底」当责任方案

白皮书假设：只要把决定映射到具体人类，责任就有着落。

如果错：人类成为「责任缓冲垫」——为自己无法控制的系统签字；自动化偏见/点击式审批让 checkpoint 失效。

检验：审批时长、理解度、否决率、追责事件。

错误三：把「验证替代信任」当低成本方案

白皮书假设：验证可以替代信任，且成本可控。

如果错：验证成本高于收益；团队降低验证强度；信任没有建立，风险反而上升。

检验：验证成本/决策价值、误报率、合作行为。

10.2 未处理的反例

反例	打击的命题	说明
NASA Challenger/Columbia	「正式程序=安全」	有评审/签署/配置管理，仍失事；程序不能替代判断
Enron	「董事会记录=治理」	记录/批准齐全，仍因激励/利益冲突崩盘
Wells Fargo	「合规=行为」	形式合规下系统性造假
Apache 邮件列表	「会议=唯一结算节点」	无会议/无纪要，但记录与决策通道合一
维基百科	「层级/会议=协调」	大规模协作靠 wiki/规范/社区
日本 ringi	「纪要=决定记录」	书面文件是追认装置，不是决定过程
Zappos Holacracy	「角色可塑性=无成本」	角色自组织失败、离职潮

这些反例说明：

- 正式程序不能替代判断；
- 记录不能替代治理；
- 会议不是唯一协作载体；
- 角色演化有政治/组织成本。

10.3 伦理、安全、法律风险

风险	表现	治理
责任	Agent 对外签约/承诺，法律责任归属不清	代理权/授权链/人类签署；法律意见
隐私	Decision/Event/Dissent 记录反对意见、绩效、客户数据	目的限制/保留期/删除权/员工告知
权力集中	规则引擎/可见性/议程准入由平台所有者定义	分权/申诉/少数保护/规则变更程序
锁定	事件 schema/状态机/授权图形成迁移成本	数据导出/互操作/退出权/断供预案
失业/劳动	只谈「角色演化」，不谈岗位替代/去技能化	劳动影响评估/再培训/协商
安全	Prompt injection/工具链投毒/身份伪造/日志篡改	输入隔离/最小权限/签名/独立审计

10.4 边界

本白皮书的结论，适用边界如下：

结论	适用	不适用
功能等价	高频、可结构化、可验证的决策	高风险/不可逆/强关系/强政治
验证替代信任	可验证的交易/任务	不可验证/需要长期关系
人类兜底	人类有时间/信息/能力/权限	人类无法理解/无法干预
角色演化	任务清晰/评估可测	政治/利益冲突/组织变革
硬约束	完全中介/规则可执行	复杂领域/规则不完备

一句话：白皮书给的是「结构语言」和「设计假设」，不是「万能公式」。

10.5 权力、劳动与退出

Agent Native 组织的工具，不只是技术系统，也是权力系统。必须处理四个问题：

权力

- 规则引擎、可见性、议程准入、何谓「证据」，由平台所有者定义；
- 这是「宪法级权力」，不是「配置项」；
- 需要分权（谁可以改规则）、申诉（被拒绝/被记录的人可以申辩）、少数保护（异议不被报复）、规则变更程序（版本/公示/过渡期）。

隐私

- Decision/Event/Dissent 日志会记录反对意见、绩效、健康、客户数据；
- 会议录音/转写涉及个人信息、员工监控、跨境传输；
- 需要目的限制、保留期、删除权、员工告知、工会/员工协商。

劳动

- 「角色演化」可能意味着岗位替代、去技能化、中层消失、劳动强度上升；
- 试点指标不应只看决策周期/成本，还要看工作质量、就业影响；
- 需要再培训、转岗、协商，而不是「技术性失业是必然代价」。

锁定与退出

- 事件 schema、状态机、授权图与 runtime 形成迁移成本；
- 需要数据导出、互操作、退出权、断供预案；

- 不要把「组织记忆」锁在一个不可导出的平台里。

10.6 对抗性安全模型

Agent Native 组织的安全，不只是「模型安全」，而是「协作系统安全」。至少覆盖：

威胁	例子	治理
身份伪造	冒充 Agent/人类	身份认证/签名/最小权限
Prompt injection	外部文档/网页注入指令	输入隔离/内容过滤/权限最小化
工具链投毒	插件/工具被篡改	供应链审计/签名/版本锁定
权限提升	组合低权限获得高权限	权限隔离/组合审计/最小权限
日志篡改	内部人改审计日志	签名/哈希链/独立见证
数据外泄	Agent 把敏感信息发到外部	输出过滤/DLP/网络隔离
模型漂移	换模型后行为变化	回归测试/模型无关协议/版本记录
拒绝服务	Agent 无限调用/循环	预算/超时/熔断/kill switch

关键判断：append-only 日志如果由同一平台控制，内部人仍可改；需要「独立见证」或「哈希链 + 外部锚定」。

章节

结论

五个判断、三类读者建议、反例条件与研究议程。

11.1 五个判断

1. **协作不是「一起做事」，而是「协调差异」**。 不同的差异，需要不同的协调机制。
2. **人类协作有三种基本结构：市场、层级/官僚制、社区/网络**。 协会、平台、小群体、公地是混合形态。
3. **会议是承诺的重要结算节点，但不是唯一节点**。 它的特殊性是「同一时空内让多方同时到场、同时承诺、同时被见证」。
4. **纪要是承诺的「稳定机制」**。 授权签发、核对确认、编号版本、时限默认、嵌入执行，五个机制在不同场景下必要性不同。
5. **迁移不是复制形式，而是功能等价（假设）**。 协议替代规范，验证替代信任，审计替代声誉，授权链替代层级；不能迁移的是社会性与责任。

11.2 给三类读者的建议

给 CEO / 管理者

1. 先诊断会议组合：哪些会可以取消、合并、缩短?
2. 把会议从「房间」改成「协议」：触发条件、参与者、决策规则、产物。
3. 建立决策日志：每个决定有理由、有批准、有复查。
4. 保留人类兜底：不可委托清单、人类 checkpoint。
5. 不要一上来做平台：先建三个对象（角色卡、决策日志、授权表）。

给产品 / 技术负责人

1. 把会议载体设计成「共享状态机」，而不是「聊天窗口」。
2. 把产物设计成「结构化对象」，而不是「文本」。
3. 用硬约束（权限/状态机/可见性/审计）替代软规范。
4. 设计「分阶段 + 并行 + 盲评 + 聚合」的会议调度，减少顺序偏差。
5. 从最小可行系统开始：身份 + PEP + 三个对象 + 事件日志 + 审批 SLA。

给 Agent 开发者

1. 不要假设 Agent 有「社会性」：信任、声誉、情感、责任，都需要外部机制。
2. 设计角色卡时，写清权限、升级、验收、生命周期。
3. 设计授权时，用「委托/能力/范围/期限/条件/审计/回收」，不要用「全权代理」。
4. 设计记忆时，用「分层/来源/版本/权限/谱系」，不要用「聊天记录」。
5. 设计验证时，用「多源/对抗/人类抽检」，不要假设 Agent 会「自我纠错」。

11.3 反例条件与研究议程

反例条件

- 如果等价物不能在三个以上流程中降低协调成本或错误率，命题不成立；
- 如果验证成本持续高于收益，迁移不经济；
- 如果人类无法理解 Agent 的关键决定，责任兜底失效；
- 如果 Agent 的策略性行为（合谋/欺骗/联盟）无法治理，授权链失效。

研究议程

1. **可验证的信任**：如何在 Agent 之间建立「不依赖关系」的信任？
2. **可演化的角色**：如何让角色从任务中自动发现、评估、固化、撤销？
3. **人类可理解的组织**：如何让人类理解 Agent 组织的决策/状态，而不被淹没？
4. **混合组织的治理**：如何在人类/Agent 混合组织中分配权力/责任/合法性？
5. **会议载体的标准化**：是否需要一套「会议协议」标准？
6. **顺序偏差的工程控制**：如何设计调度策略，减少群体思维/信息级联？
7. **责任追溯**：如何让每个决定/产物都能追溯到具体人类/Agent？
8. **失败案例研究**：需要更多「有纪要/无纪要」「人类决策/Agent 决策」的对照数据。

11.4 最后一句话

> 会议不是「大家说话的地方」，而是「组织做出决定、分配授权、沉淀记忆的地方」。

在人类组织中，这个「地方」是会议室。

在 Agent Native 组织中，这个「地方」是一个**可编程的共享状态机**：它让人类和 Agent 在同一个结构里协作，让决定可追溯，让授权可执行，让记忆可检索。

载体变了，会议才真正变了。

章节

附录 A：证据表

每条结论的证据、级别与命题支持度。

结论	证据	级别	命题支持度
董事会记录/签名是法定要求	《公司法》第 64/73/119/124 条	A	机制存在（强），机制有效（未验证）
董事异议记载可免责	《公司法》第 125 条	A	机制存在（强），机制有效（未验证）
人大发言记录经本人核对签字	人大常委会议事规则第 31 条	A	机制存在（强），机制有效（未验证）
国务院常务会纪要由总理签发	国务院工作规则第 22 条	A	机制存在（强），机制有效（未验证）
ISO 决定 48h、纪要 4 周、4 周默认确认	ISO/IEC Directives §5.5	A	机制存在（强），机制有效（未验证）
IETF 纪要含议程/讨论/决定/出席者	RFC 2418 §3.1	A	机制存在（强），机制有效（未验证）
NASA 评审关闭需 RID 处置/纪要分发/decision memo	NPR 7123.1C §5.2.3.1	A	机制存在（强），机制有效（未验证）
医疗疑难病例讨论需专册/签字/入病历	医疗质量安全核心制度要点	A	机制存在（强），机制有效（未验证）
合议庭少数意见入笔录/成员签名	最高法合议庭规定第 11 条	A	机制存在（强），机制有效（未验证）
英国内阁结论记讨论要点/主席总结	Cabinet Manual §4.34	A	机制存在（强），机制有效（未验证）
FOMC 纪要含观点/理由/投票	Fed FAQ / FOMC minutes	A	机制存在（强），机制有效（未验证）
市场/层级/网络是三种治理结构	Williamson 1975; Powell 1990	B	解释框架（中），因果（弱）
官僚制靠规则/档案/管辖权	Weber 1922	B	解释框架（中），因果（弱）
隐性知识可部分编码	Polanyi 1966; Nonaka & Takeuchi 1995	B	解释框架（中），因果（弱）
顺序偏差/群体思维	Asch 1951; Janis 1972; Sunstein 2002	B	解释框架（中），因果（弱）
Agent 多轮对话/顺序偏差	多 Agent 框架实践	C	设计假设（弱）

章节

附录 B：诊断表

十六个问题，判断一场会议/一个决定是否闭环。

问题	是/否	证据
这场会有明确的决定吗?		
决定有决策者吗?		
决策者在场吗?		
决定有理由吗?		
理由有关键假设吗?		
有反对意见吗?		
反对意见被记录了吗?		
决定有委托来源吗?		
决定有授权范围吗?		
授权有期限吗?		
决定有 owner/due/验收吗?		
决定有记录吗?		
记录可检索吗?		
决定映射到具体人类吗?		
有复查日期/触发条件吗?		
有撤回机制吗?		

章节

附录 C：术语表

协作、会议、承诺闭环、授权、硬约束等术语。

术语	英文	含义
协作	collaboration	协调差异、共同行动的过程
协调机制	coordination mechanism	价格/权威/规范/信任/规则
结构特征	structural dimension	会议/协作中可比较、可变化的属性
结构错配	structural misfit	特征组合与功能/环境不匹配
会议载体	meeting medium	会议发生的结构（共享状态机/聊天/文档）
产物载体	artifact medium	会议产出的存储（纪要/结构化对象/组织记忆）
承诺闭环	commitment loop	差异→决定→证据→委托→授权→记录→责任→复查
角色卡	role card	定义 Agent 角色的结构化配置
授权链	authorization chain	逐级申请/批准的权限机制
硬约束	hard constraint	工具执行的权限/状态机/可见性/审计
软约束	soft constraint	规范/提示/建议
能力	capability	Agent 可以执行的动作
事件溯源	event sourcing	用不可变事件记录状态变化
盲评	blind review	匿名评审，减少权威/从众偏差
顺序偏差	order effect	发言顺序对判断的影响
群体思维	groupthink	群体追求一致，忽略反对意见
信息级联	information cascade	前面的错误被后面放大
责任漂移	responsibility drift	人类脱离，责任变得模糊
权限蔓延	privilege creep	权限逐渐扩大，超出角色所需

章节

附录 D：参考文献

人类协作、会议、制度案例、Agent 研究、安全治理。

人类协作理论

1. Hayek, F. A. (1945). The Use of Knowledge in Society. *American Economic Review*, 35(4), 519 – 530.
2. Coase, R. H. (1937). The Nature of the Firm. *Economica*, 4(16), 386 – 405.
3. Williamson, O. E. (1975). *Markets and Hierarchies*. Free Press.
4. Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. Free Press.
5. Powell, W. W. (1990). Neither Market nor Hierarchy. *Research in Organizational Behavior*, 12, 295 – 336.
6. Ouchi, W. G. (1980). Markets, Bureaucracies, and Clans. *Administrative Science Quarterly*, 25(1), 129 – 141.
7. Adler, P. S. (2001). Market, Hierarchy, and Trust. *Organization Science*, 12(2), 215 – 234.
8. Granovetter, M. (1985). Economic Action and Social Structure. *American Journal of Sociology*, 91(3), 481 – 510.
9. Ostrom, E. (1990). *Governing the Commons*. Cambridge University Press.
10. Benkler, Y. (2006). *The Wealth of Networks*. Yale University Press.
11. Weber, M. (1922). *Economy and Society*.
12. Polanyi, M. (1966). *The Tacit Dimension*.
13. Simon, H. A. (1947). *Administrative Behavior*. Macmillan.
14. Mintzberg, H. (1979). *The Structuring of Organizations*. Prentice-Hall.

会议与组织

15. Schwartzman, H. B. (1989). *The Meeting: Gatherings in Organizations and Communities*. Plenum Press.
16. Boden, D. (1994). *The Business of Talk*. Polity Press.
17. Rogelberg, S. G. (2019). *The Surprising Science of Meetings*. Oxford University Press.
18. Lencioni, P. (2004). *Death by Meeting*. Jossey-Bass.
19. Thompson, J. D. (1967). *Organizations in Action*. McGraw-Hill.
20. Galbraith, J. R. (1973). *Designing Complex Organizations*. Addison-Wesley.

制度与案例

21. 《中华人民共和国公司法》（2023 修订）。
22. 《全国人民代表大会常务委员会议事规则》。
23. 《国务院工作规则》（2023）。
24. ISO/IEC Directives, Part 1 (2024).
25. NASA NPR 7123.1C.
26. FDA Guidance for Clinical Trial Sponsors: Establishment and Operation of Clinical Trial Data Monitoring Committees.
27. Federal Reserve, FOMC Minutes FAQ.
28. Cabinet Office (UK), The Cabinet Manual.
29. GitLab Handbook, All-Remote Meetings.
30. Amazon, 2015/2016/2017 Letters to Shareholders.
31. Basecamp, Shape Up.

Agent 研究

32. Smith, R. G. (1980). The Contract Net Protocol. *IEEE Transactions on Computers*, C-29(12), 1104 – 1113.
33. Hayes-Roth, B. (1985). A Blackboard Architecture for Control. *Artificial Intelligence*, 26(3), 251 – 321.
34. Hong, S., et al. (2023). MetaGPT. *arXiv:2308.00352*.
35. Qian, C., et al. (2023). ChatDev. *arXiv:2307.07924*.
36. Wu, Q., et al. (2023). AutoGen. *arXiv:2308.08155*.
37. Du, Y., et al. (2023). Improving Factuality and Reasoning in Language Models through Multiagent Debate. *arXiv:2305.14325*.
38. Park, J. S., et al. (2023). Generative Agents. *arXiv:2304.03442*.

安全与治理

39. OWASP. (2025). *OWASP Top 10 for LLM Applications*.
40. NIST. (2023). *AI Risk Management Framework (AI RMF 1.0)*.
41. Saltzer, J. H., & Schroeder, M. D. (1975). The Protection of Information in Computer Systems. *Proceedings of the IEEE*, 63(9), 1278 – 1308.

白皮书研究稿 v0.3 完。