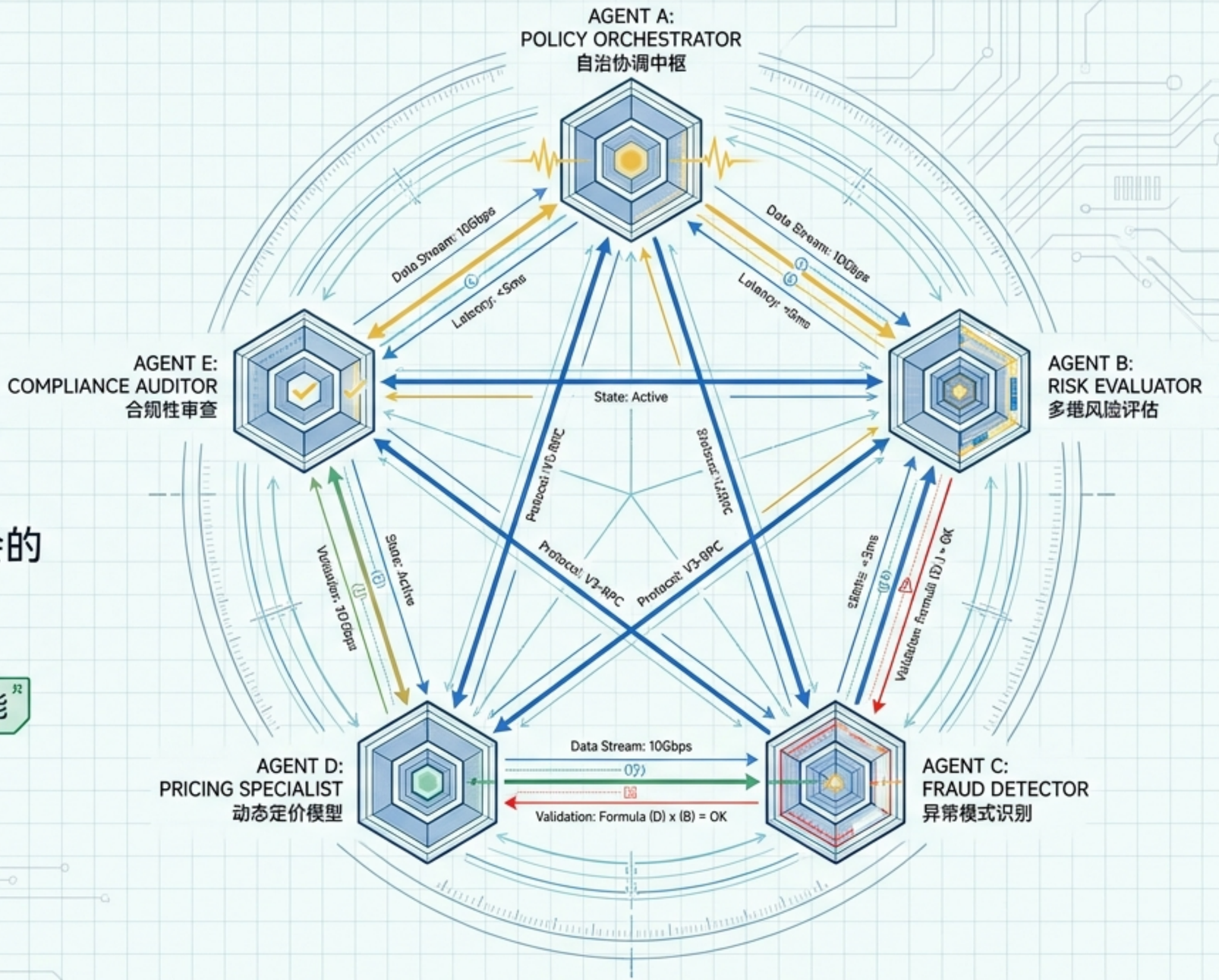
5 个 Agent 的核保辩论会

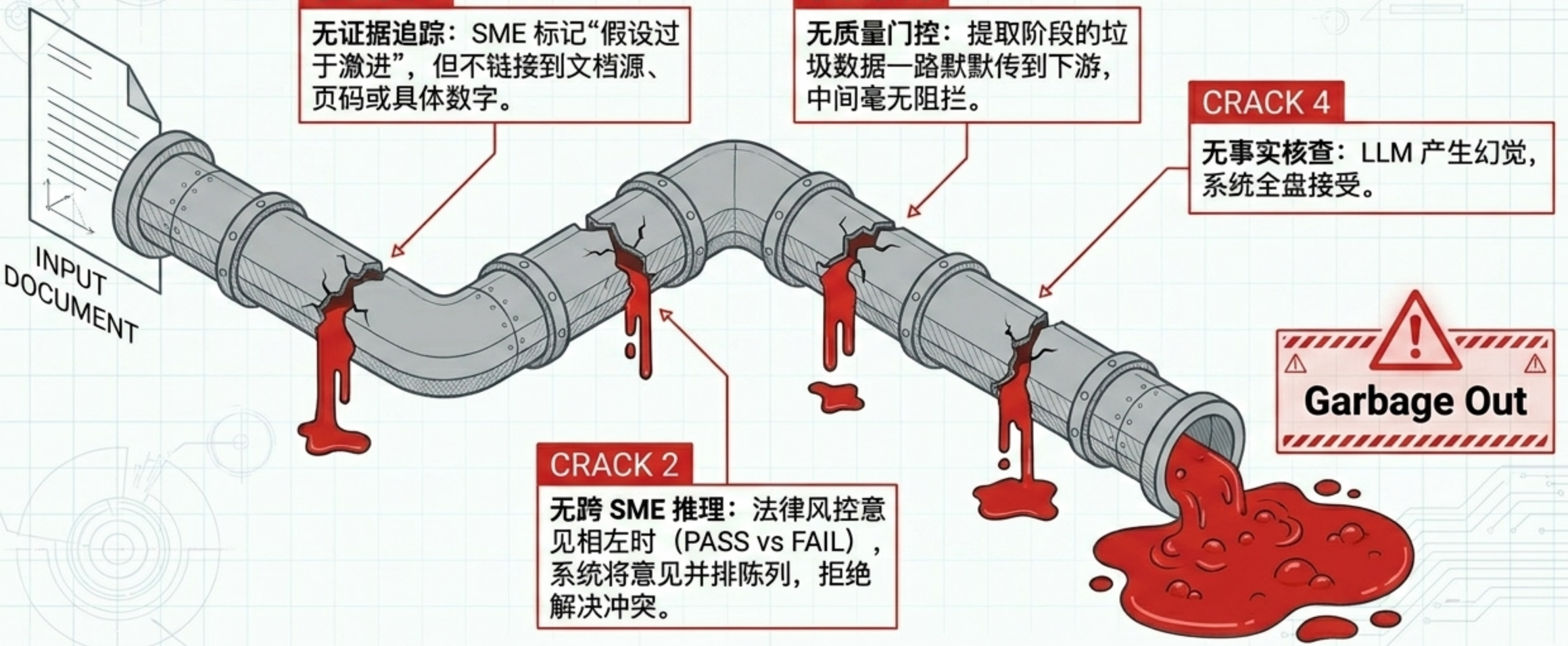
从线性流水线到多智能体自治议会的架构重构验证

架构蓝图 | 系统可靠性 | 成本与效能



V2 架构是一条跑一遍就祈祷的单向流水线

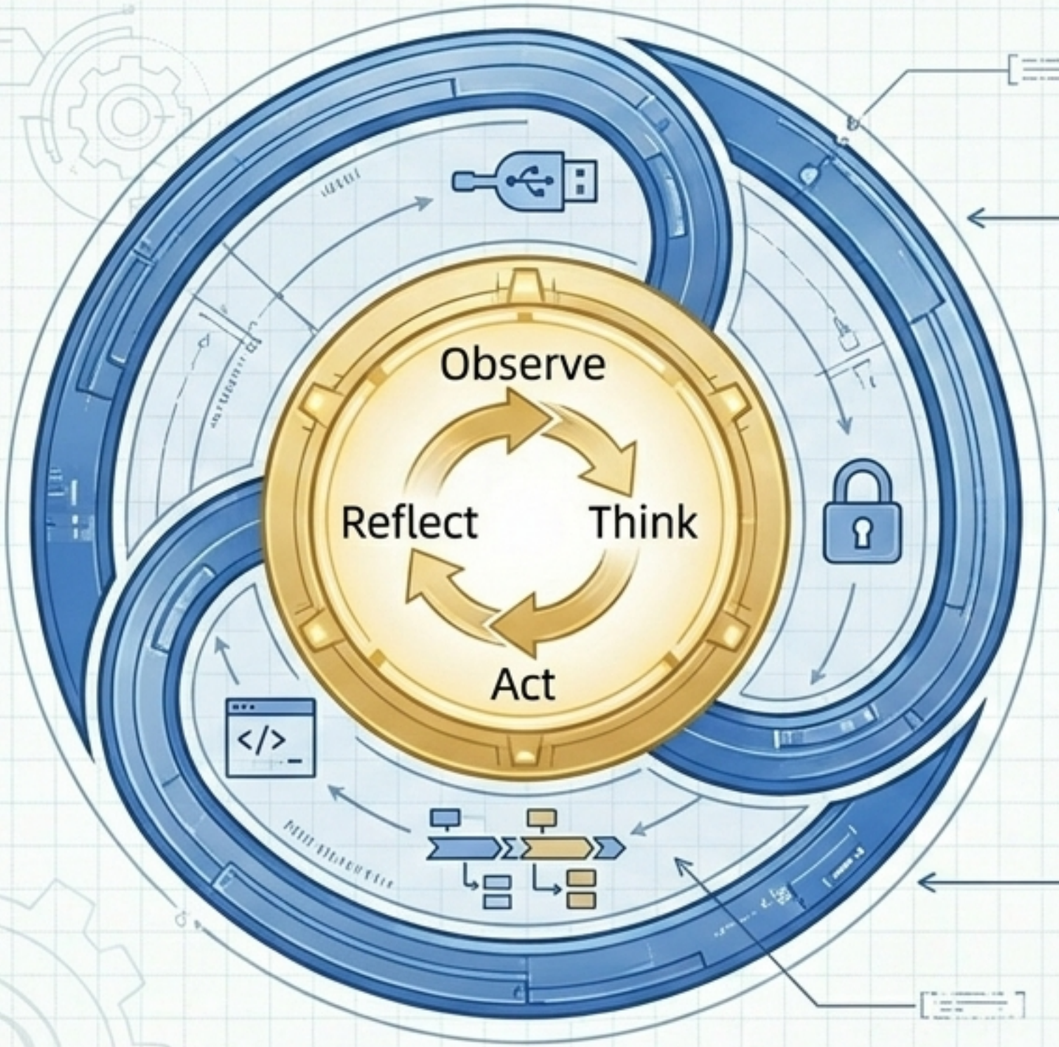
没有检查点的线性系统，必然导致错误在下游无限放大



范式转移：从流水线的盲目拼接，到自治议会的动态纠错

	【V2 架构】单向流水线	【V3 架构】自治议会
核心逻辑	调一次祈祷结果好	Observe-Think-Act-Reflect 迭代
输出形态	仅输出主观观点	必须附带证据与原文切片
意见整合	孤立评估与暴力拼接	结构化辩论与逻辑裁决
内容质控	垃圾进垃圾出（GIGO）	达标质量门控阈值才可产出

底层引擎：驱动所有 Agent 的核心心脏（AgentRuntime）



所有自治 Agent 共享同一个通用执行引擎。

Runtime 插件化

引擎与具体 Agent 无关。任何 BaseAgent 子类都能无缝插入。

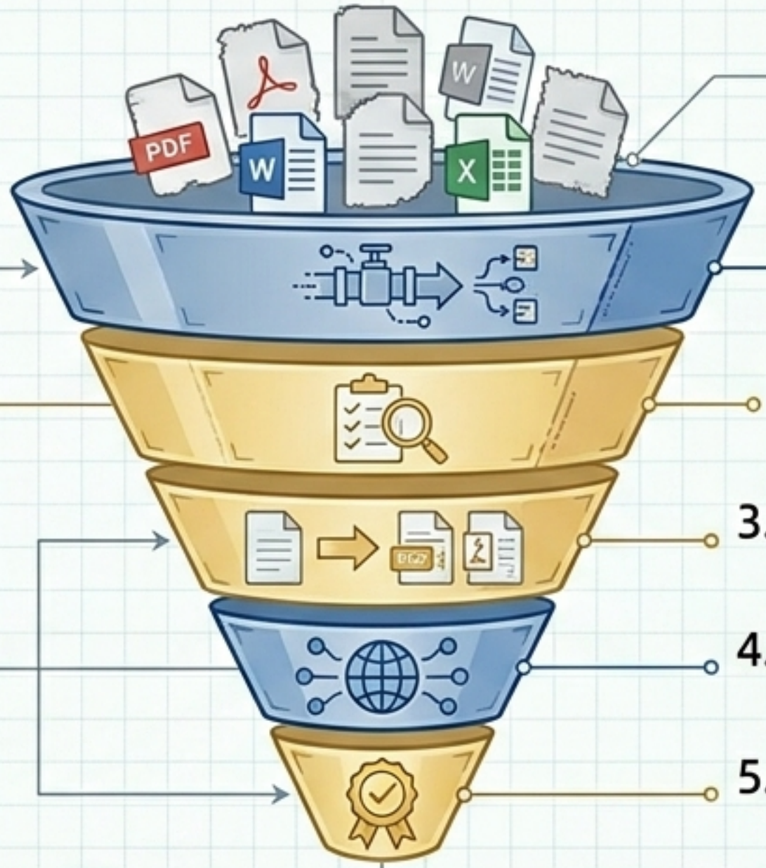
Context 不可变

拒绝突变。每次交互均返回新实例，彻底杜绝别名 Bug。

AgentTrace 全景追踪

成本、Token、耗时、工具 I/O——每一步均被记录。随时审计系统“发生了什么”与“为什么”。

1. 提取 Agent —— 不对就重来的五层过滤网



原始文档输入

1. 批量提取：运行传统管线获取初始 Schema。

2. 反思自评：自检完整性，圈定缺失与异常字段。

2. 反思自评：自检完整性，圈定缺失与异常字段。

3. 针对性补读：重新定向读取特定文档，补齐 Top-5 关键缺失信息。

4. 交叉验证：针对高价值字段，配合网络搜索进行事实核查。

4. 交叉验证：针对高价值字段，配合网络搜索进行事实核查。

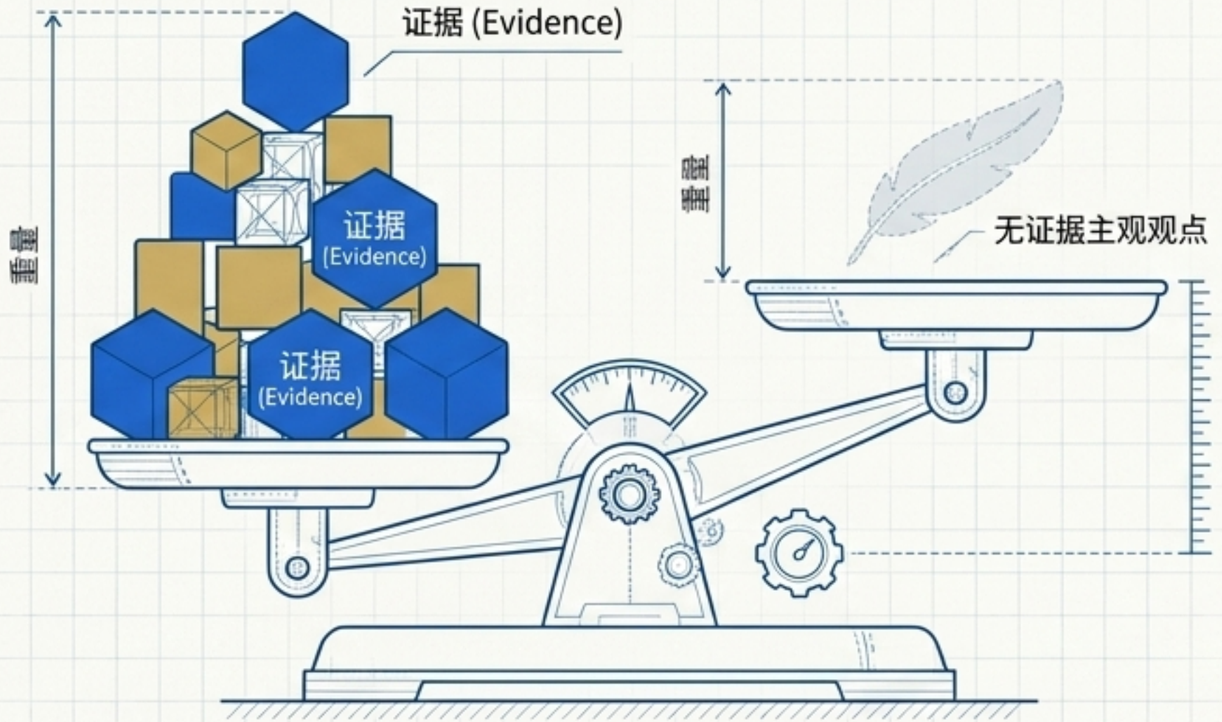
5. 最终反思：达到 80/100 质量门控后才允许最终产出。

质量阈值：80/100

DealSchema

核心模型支撑：gemini-3.1-pro

2. 领域专家 Agent —— 严苛的证据天平与质量公式



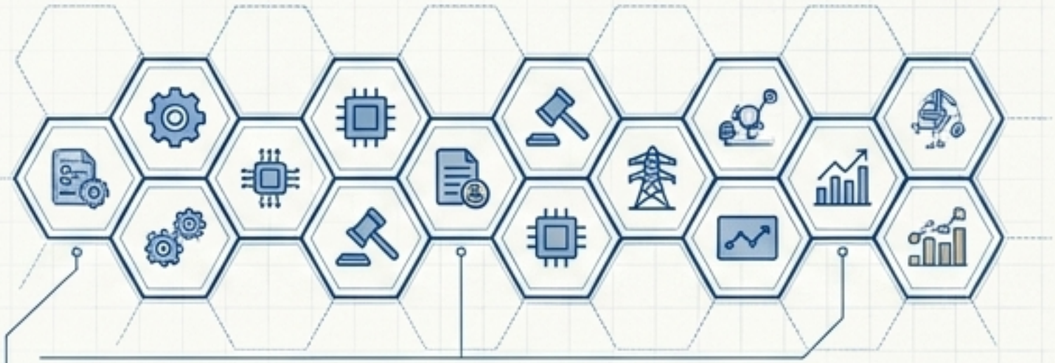
质量分数 = concerns_backed × 40 + strengths_backed × 30
+ confidence × 20 + summary_quality × 10

工作流（4 阶段）：

初始评估 → 检查证据充分率（目标 80%+） → 工具验证辅助 → 最终门控反思

局部域内决策：

每个专家只产出其领域内的决策结论（PASS / PASS_WITH_CONDITIONS / NEEDS_CLARIFICATION / FAIL）。



覆盖 11 大专业领域：包含 GPU 工程师、法律分析师、风控团队、能源基建专家等。不亮证据，系统不予采信。

3. IC 委员会 Agent —— 触发结构化辩论的最终裁判

聚合与综合：编译基线包，将 11 个 SME 的输出转化为连贯的执行摘要。

发现冲突或 CRITICAL 担忧？

YES

结构化辩论（条件触发）

当意见相左时，生成辩论。
双方强制出示证据 -> 逻辑裁决。
针对 CRITICAL 进行逐层质疑（挑战模式）。

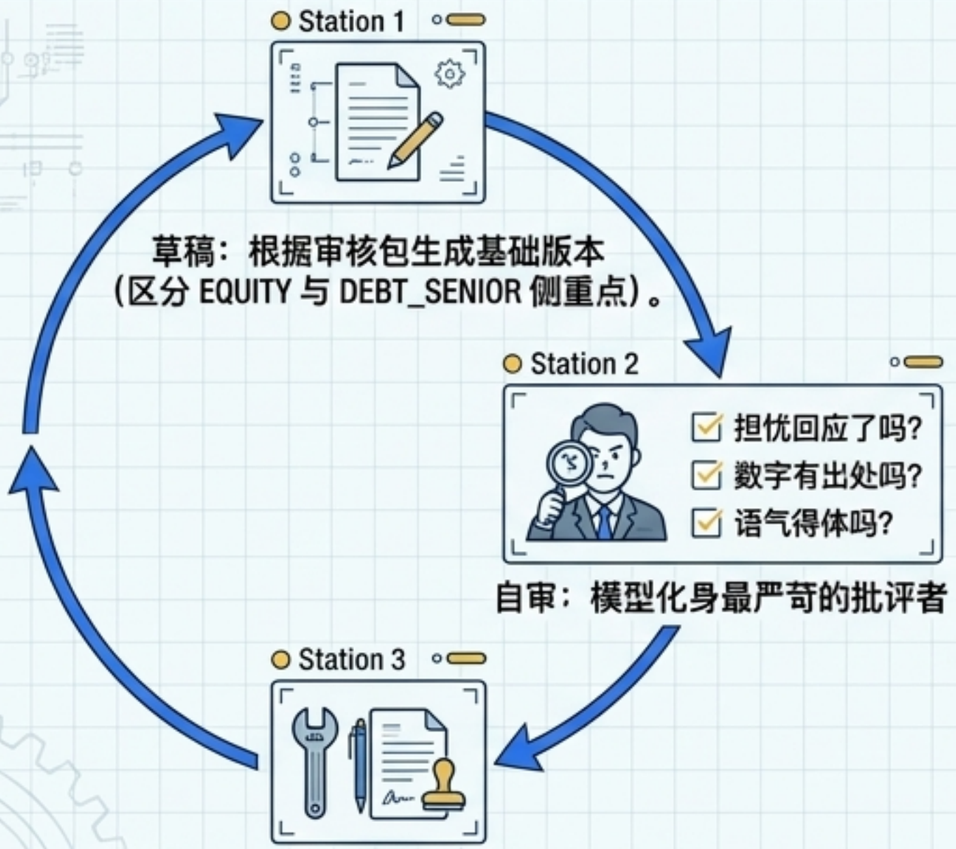
NO

问题合成：摒弃字数匹配，采用语义级别的彻底去重。

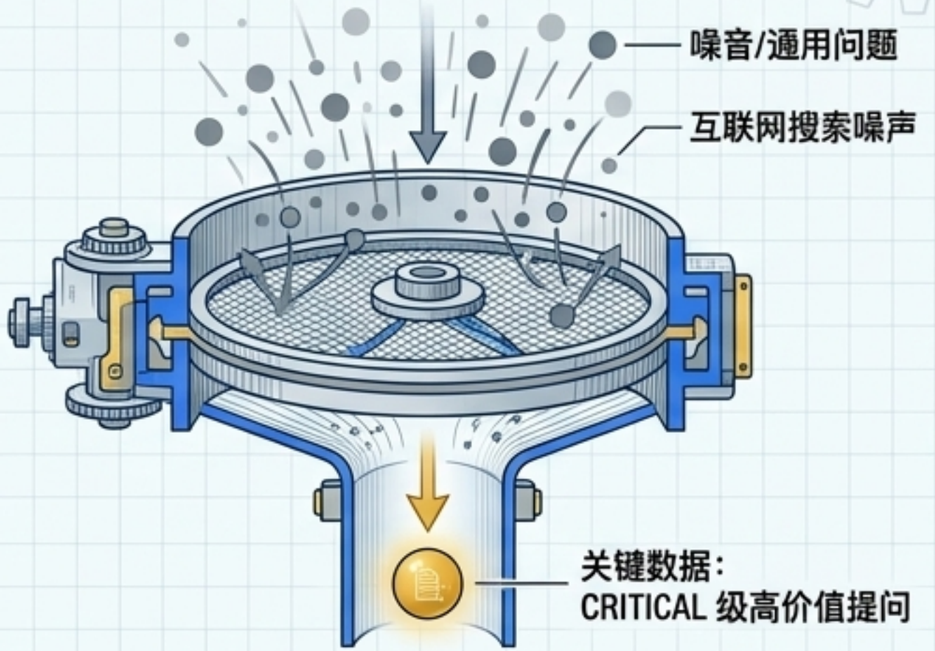
最高系统门槛：**90/100**

4 & 5. 报告与提问 Agent —— 极致的自我批评与噪音过滤

[4] 报告 Agent：自杀式的自我批评闭环



[5] 提问 Agent：互联网级的噪音筛子

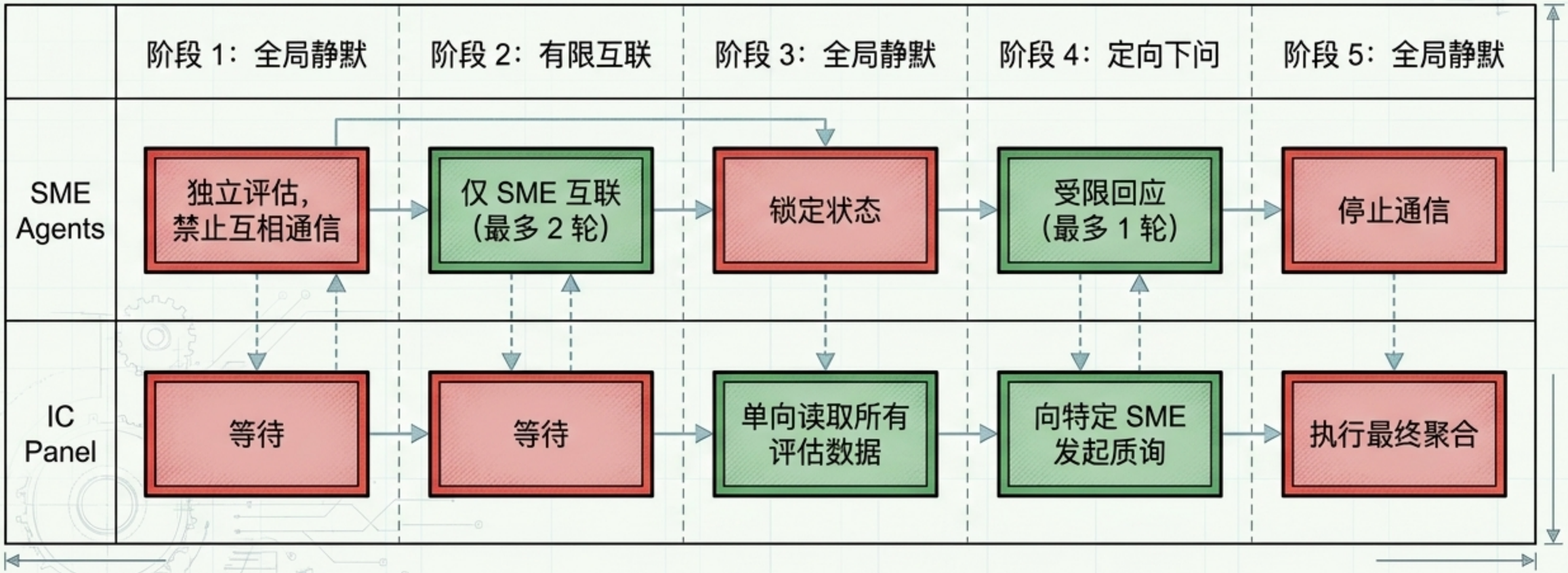


网络过滤：系统自动检索公开网络。凡是能搜到答案的问题，一律剔除。绝不浪费目标运营商的时间。

优先级排序：仅保留针对缺口与冲突的 CRITICAL 级高价值提问。

通信护栏：用严格的时序通道防止系统陷入死循环

不限制通信，结构化消息总线将演变成一锅粥。系统采用 5 阶段静默与唤醒机制：



商业效能测算：极高复杂度背后的极低边际成本

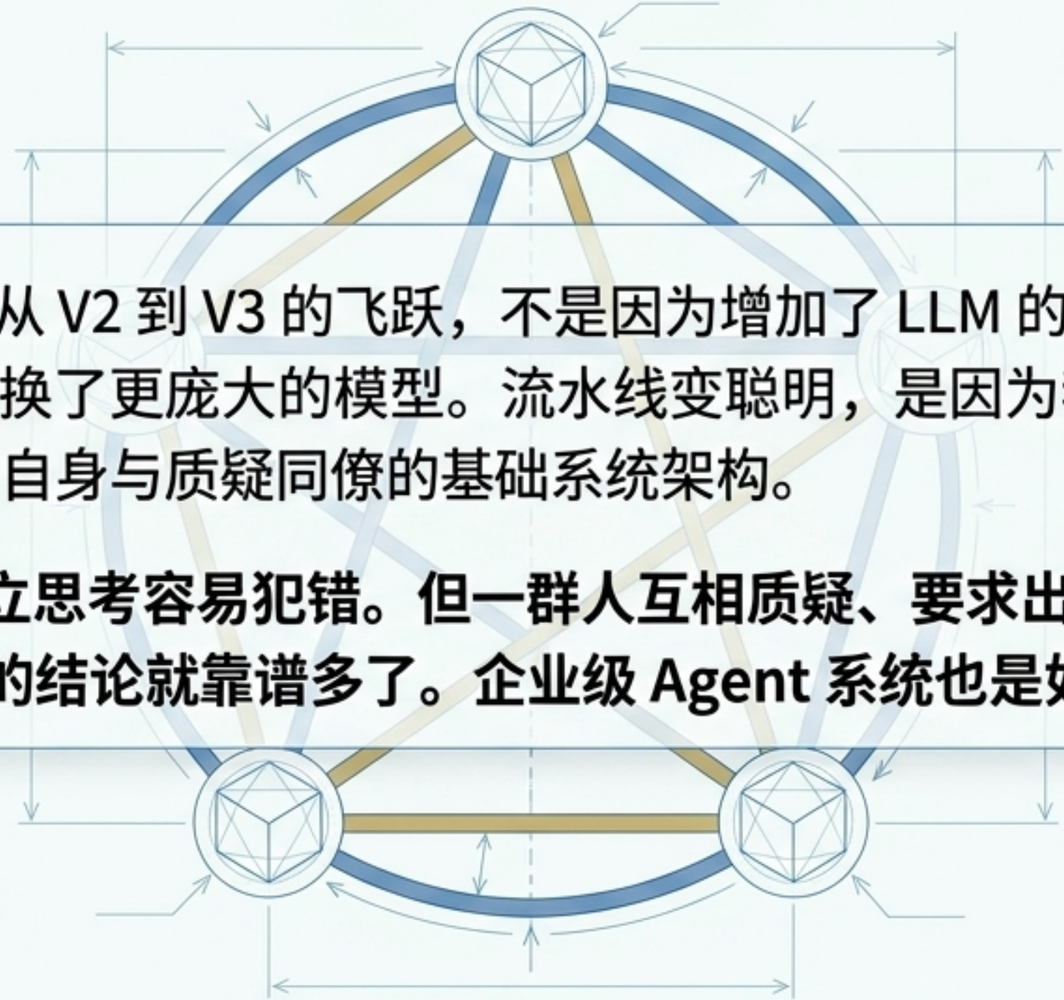
Arkane Cloud 测试案例账单（单次端到端运行）

```
[节点执行] SME Agent（如 GPU 工程师）：75K Tokens | ~5 分钟
[节点执行] IC Panel（投资裁判）：       20K Tokens | ~30 秒
[节点执行] 报告生成（股权类）：         95K Tokens | ~3 分钟
[节点执行] 提问拦截:                    21K Tokens | ~1 分钟
```

总计核心数据：~211K Tokens | 耗时仅约 10 分钟

经济学意义：全流水线跑满 11 个 SME，单笔交易仅需几美元。在确保证据100%可溯源（源文档 -> 物理位置 -> 文本片段）的前提下，成本比人类分析师低几个数量级。

智能的涌现：并非由于模型的堆砌，而是关系的重塑



核心洞察：从 V2 到 V3 的飞跃，不是因为增加了 LLM 的调用次数，也不是单纯切换了更庞大的模型。流水线变聪明，是因为我们赋予了 Agent 怀疑自身与质疑同僚的基础系统架构。

一个人独立思考容易犯错。但一群人互相质疑、要求出示证据——出来的结论就靠谱多了。企业级 Agent 系统也是如此。

自我反思 + 交叉质疑 = 极高的系统级可靠性