智能体与逻辑链阈值

政府违规信息智能核查系统

判定与输出规范 (v1)

1. 总体目标

为“严苛智能体（Strict Agent）”与“宽松智能体（Lenient Agent）”建立统一的判定逻辑、证据等级体系与输出格式，使两者在不同容忍度下协同工作、互不冲突。

2. 核心概念

2.1 智能体类型

智能体	判定阈值	目标	特点
宽松智能体 (Lenient)	E1 及以上（有直接矛盾证据）	高精度、低误报	“疑罪从无”，不报线索，仅确凿违规
严苛智能体 (Strict)	E0 及以上（存在可疑线索）	高召回、早发现	“召回优先”，线索级也可上报

2.2 冲突类型

1. **状态矛盾 (Binary Contradiction)** – 两条互斥事实并存。
2. **流程/资格矛盾 (Process/Eligibility)** – 缺失前置条件或资质冲突。
3. **数值/会计不一致 (Quantitative)** – 数值或财务逻辑不符。
4. **时间/因果矛盾 (Temporal/Causal)** – 时间顺序反常。
5. **聚合/网络异常 (Aggregate/Network)** – 聚合或宏观统计偏离常识。

2.3 证据等级

等级	名称	定义	示例
E0	线索 (Cue)	异常迹象，尚无可追溯证据	登记缺失、流程异常
E1	直接矛盾 (Direct Conflict)	至少一对互斥事实，可追溯	死亡后仍发养老金
E2	多源印证 (Corroborated Conflict)	两个以上独立数据源印证	民政+社保+税务同证实
E3	高确信 (High Confidence)	多月重复、时序正确、无例外	多期重复违规且无修正记录

3. 判定逻辑

宽松智能体

- **阈值：E1**
- 达到 E1+ → `confirmed`
- 未达 E1 → `no-conclusion`（附调查摘要）

严苛智能体

- **阈值：E0**
- 达到 E0 → `suspected`
- 达到 E1+ → `confirmed`
- 必须附“人工复核建议”或“下一步取证路径”

4. 搜索与预算策略

- **Join 深度上限**：4
- **数据源上限**：6
- **执行时间限制**：60 秒
- **扫描行数上限**：2,000,000
- **终止条件**：
  - 达到阈值立即停止；

- 超出预算仍未达阈值 → 返回“调查覆盖声明（search_summary）”并结束。

5. 统一输出格式（JSON Schema）

```
{
  "verdict": "no-conclusion | suspected | confirmed",
  "evidence_level": "E0 | E1 | E2 | E3",
  "conflict_type": "binary | process | quantitative | temporal | aggregate",
  "risk_score": 0.0,
  "evidence_chain": [
    {"source": "string", "entity_key": "string", "fact": "string", "time": "string"}
  ],
  "search_summary": {
    "sources_scanned": 0,
    "joins_attempted": ["A→B on key"],
    "coverage_gaps": ["string"],
    "dq_flags": ["string"],
    "runtime_budget": {"elapsed_sec": 0, "row_scanned": 0, "limit_hit": false}
  },
  "next_actions": ["string"],
  "explanations": "string",
  "child_cases": [ {"...嵌套个体案件..."} ]
}
```

6. 示例案例

案例一：状态矛盾（死亡仍领养老金）

- **冲突类型**：Binary Contradiction
- **证据等级**：E2
- **宽松体输出**：`confirmed (E2)`
- **严苛体输出**：`confirmed (E3)`
- **解释**：民政与社保系统均显示互斥事实，多期重复。

案例二：流程矛盾（发放但无登记）

- **冲突类型**：Process
- **证据等级**：E0
- **宽松体输出**：`no-conclusion`（附 coverage 声明）
- **严苛体输出**：`suspected (E0)`（附复核建议）
- **解释**：缺登记记录但可能跨区或延迟入库。

案例三：聚合异常（虚拟地址集群低参保）

- **冲突类型**：Aggregate
- **证据等级**：E0 集群线索 + E1 个体矛盾
- **宽松体输出**：集群 `no-conclusion (E0)`，个体 `confirmed (E1)`
- **严苛体输出**：集群 `suspected (E0)` + 嵌套 `child_cases[E1]`
- **解释**：多企业共享虚拟地址、社保人数异常，个别企业高开票零参保。

---

7. 指标与评估

指标	宽松体	严苛体
主要目标	Precision	Recall
容错	高精度、可放弃	高召回、可冗余
输出要求	证据链完整	复核建议明确
适用场景	最终核查、通报	初筛、预警

---

8. 落地建议

- **配置化阈值**：`operating_point = {lenient: E1, strict: E0}`
- **统一日志格式**：记录 prompt → plan → sql → error → fix → evidence_level
- **图谱化证据链**：节点=事实，边=关联，支持溯源展示
- **UI 展示**：卡片式输出 `verdict + badge(evidence_level) + explanation + next_actions`

---

9. 总结

> 宽松智能体：只判“实锤” → 未达E1即不结论。
>
> 严苛智能体：线索即上报 → 从E0开始报疑似。
>
> 两者共享统一证据等级体系与输出格式，区别仅在阈值与话术。