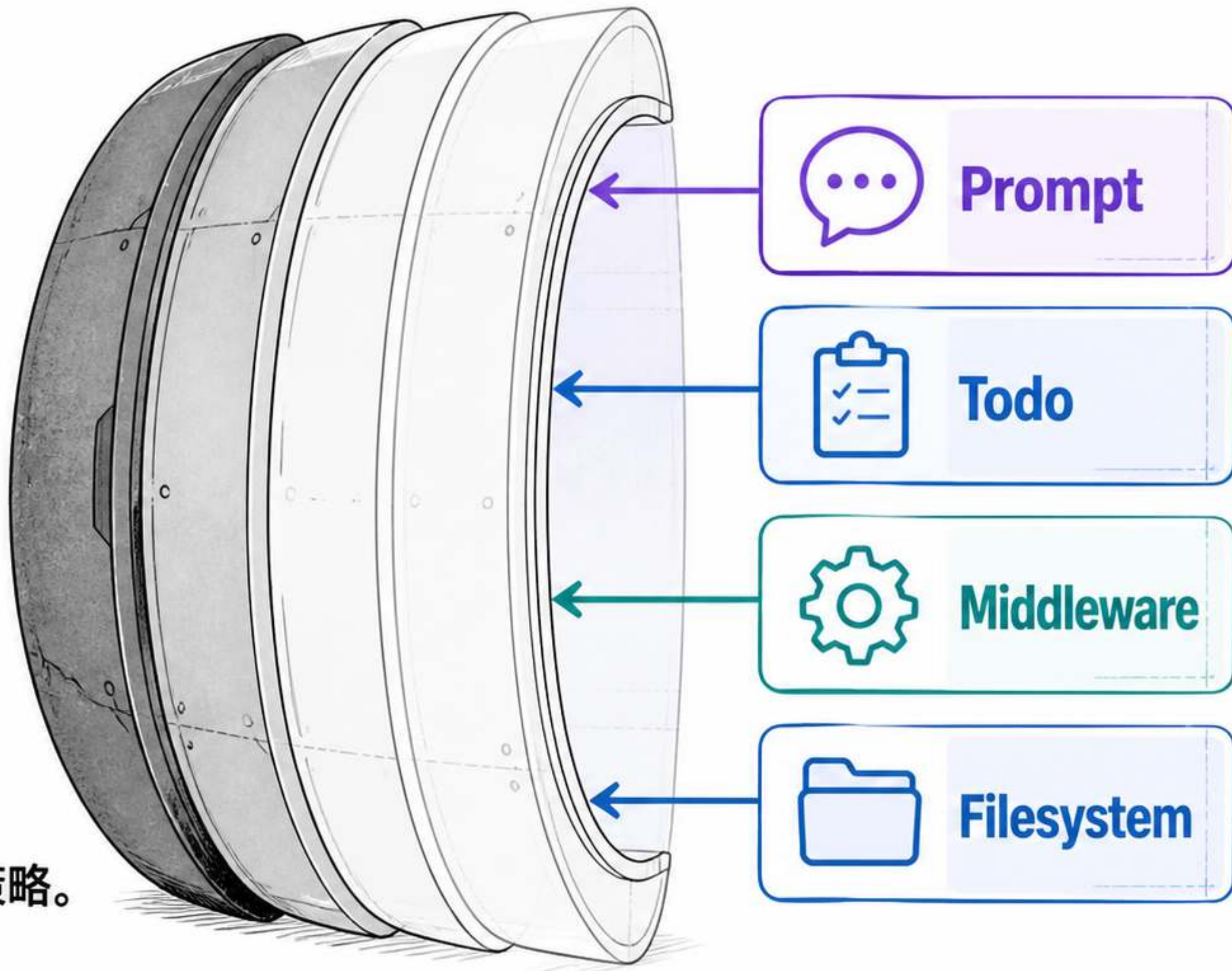


Deep Agents v0.7

更轻、更透明、更可配置的 Harness

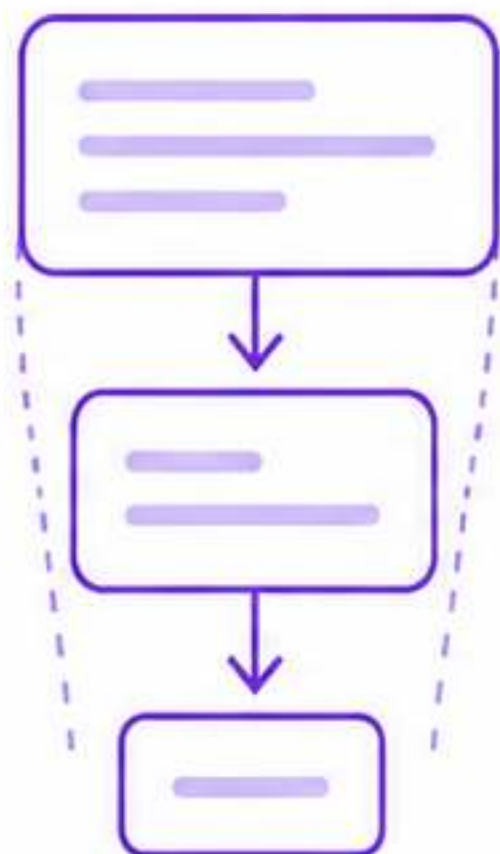
默认层变薄之后，应用必须主动选择策略。



先按 changelog 看全貌，再判断升级影响

1 组优化、4 组功能、3 组破坏性变化

Optimizations



- ✓ 空基础 Prompt
- ✓ 精简工具描述
- ✓ 基础输入 -65%

Features



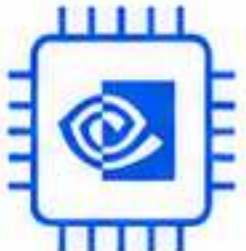
Middleware 覆盖



文件工具

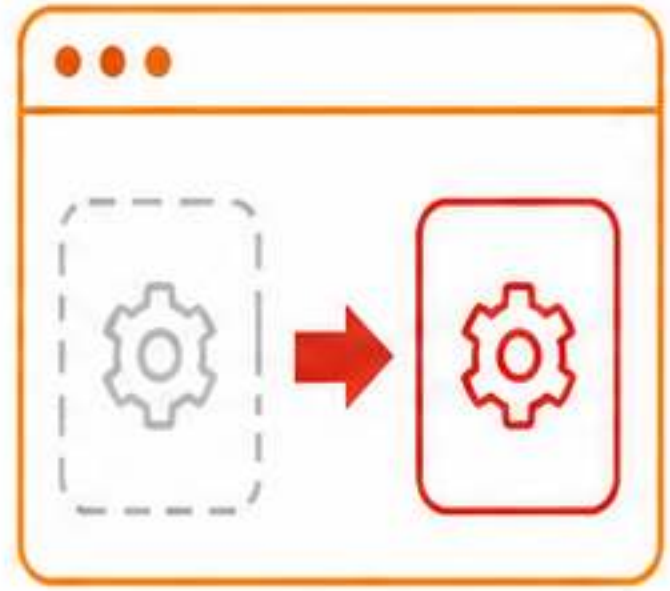


Prompt Caching



NVIDIA 支持

Breaking changes



- ! Todo 改为可选
- ! Backend 兼容层移除
- ! 输出格式变化

基础输入减少 65%， 不等于总成本下降 65%

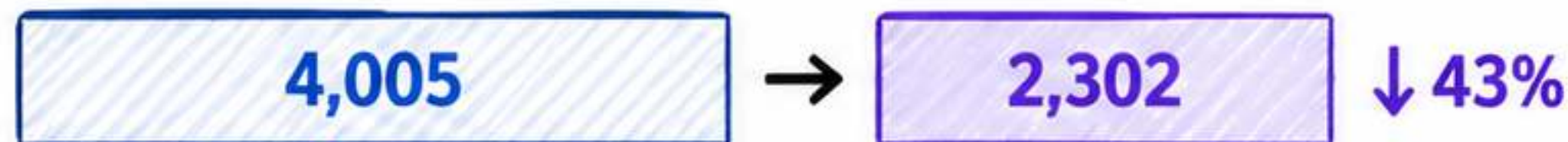
65%

☆ Source: Deep Agents v0.7 blog / PR #5009

5,395 → 1,895 Token: 默认简单回合基础输入



4,005 → 2,302 Token: 工具描述, 下降 43%

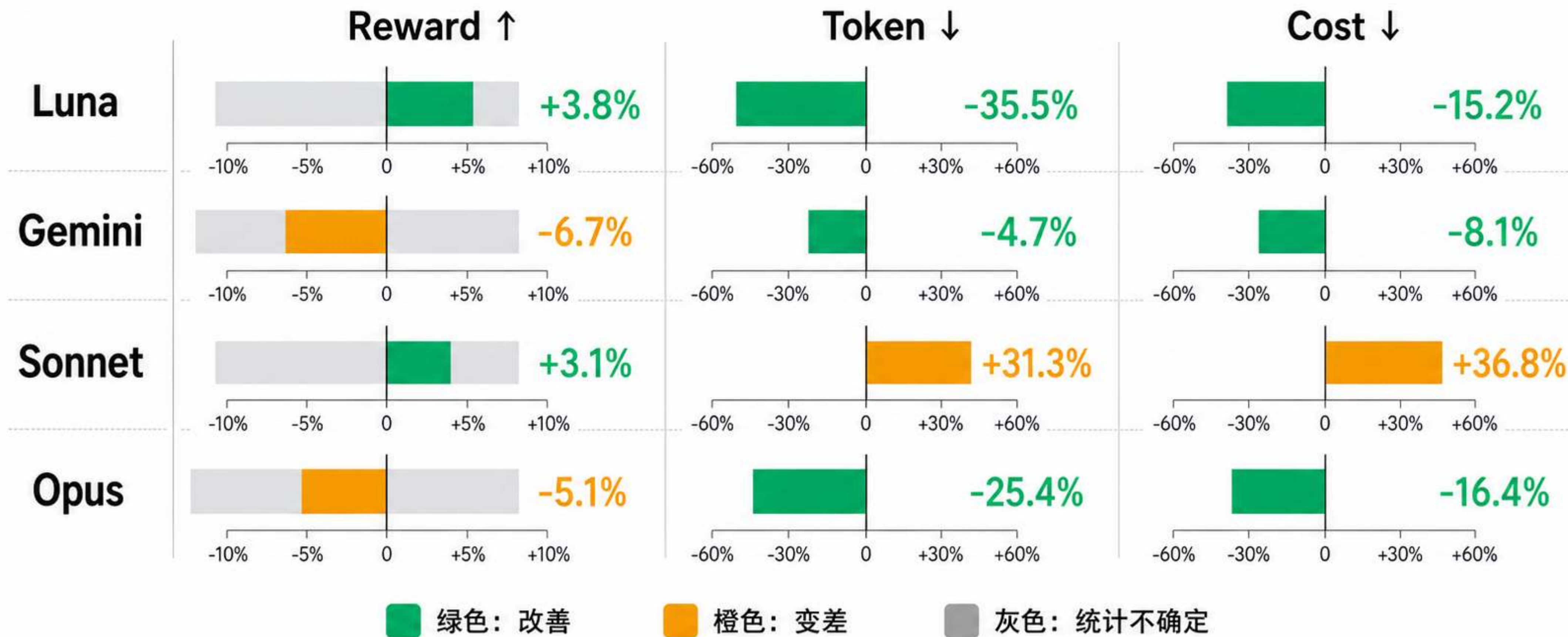


实际账单仍受历史、工具结果、
子 Agent、摘要和重试影响



好坏不能只看正负号，要先看指标方向

Reward ↑ 越高越好；Token ↓、Cost ↓ 越低越好



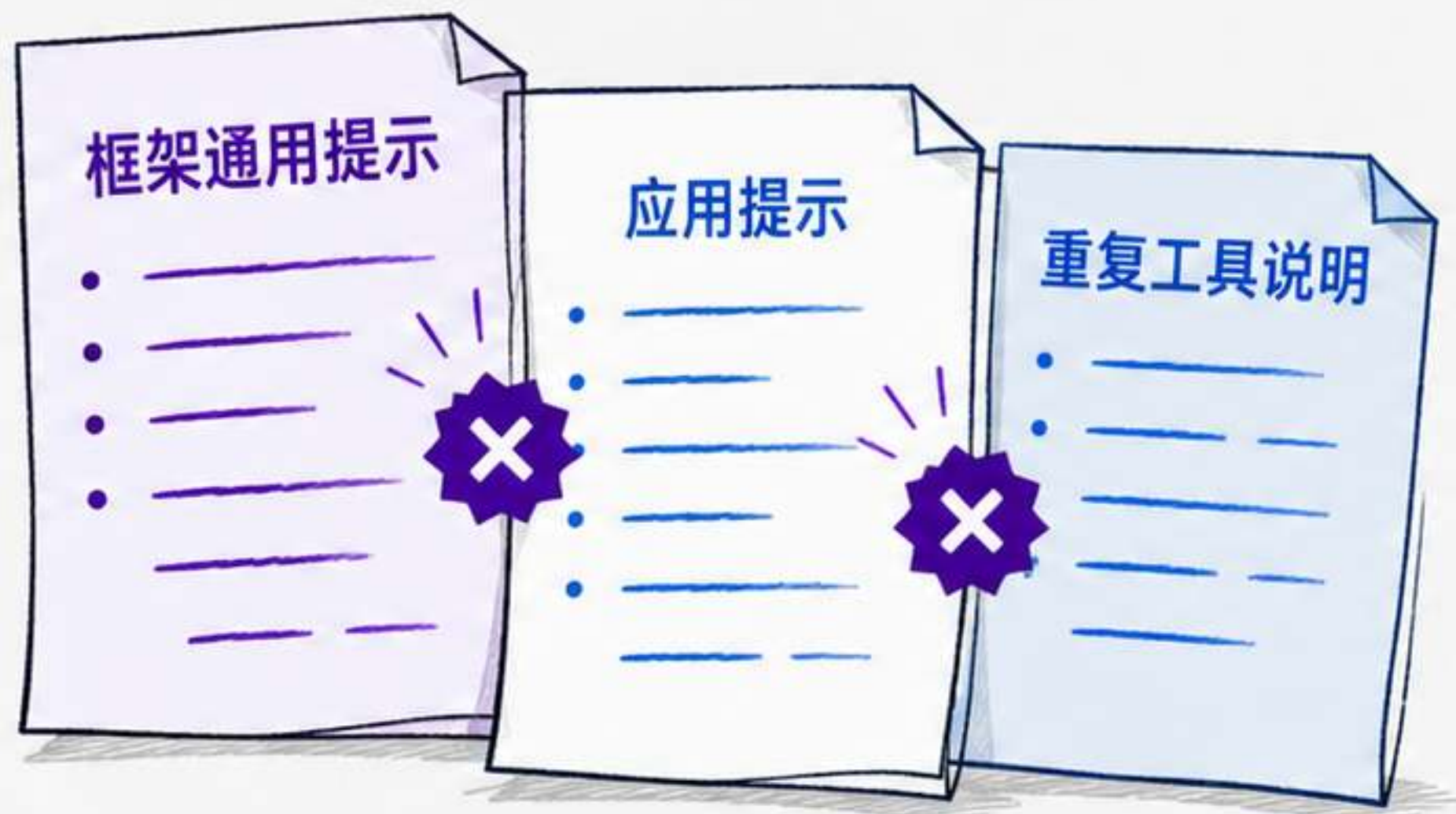
所有 Reward 置信区间均跨零；颜色只表示方向，不表示统计显著

Source: Deep Agents PR #5009

Prompt 变空，让业务指令真正“说了算”

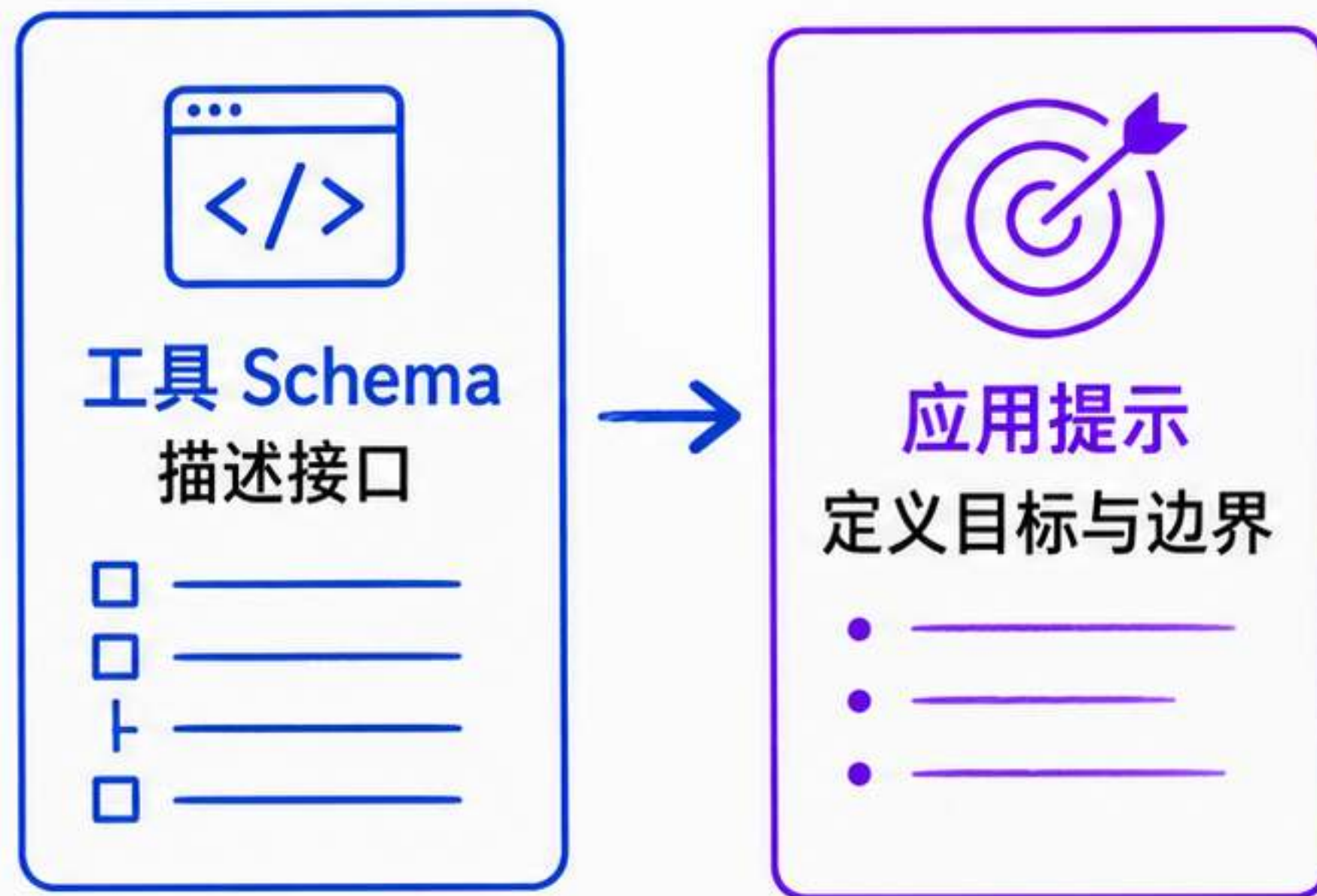
接口负责描述能力，系统提示词只保留业务约束。

旧： 框架通用提示 + 应用提示 + 重复工具说明



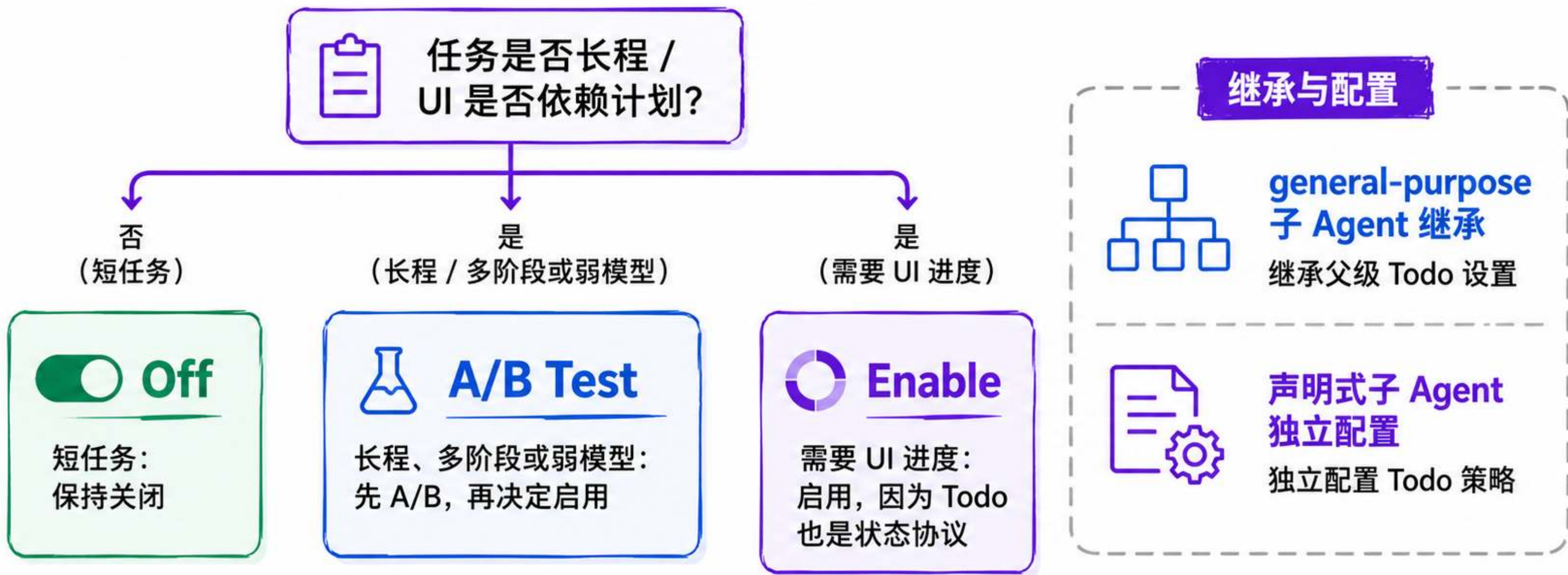
不要为了怀旧把旧长 Prompt 整段复制回来

新： 工具 Schema 描述接口，应用提示定义目标与边界



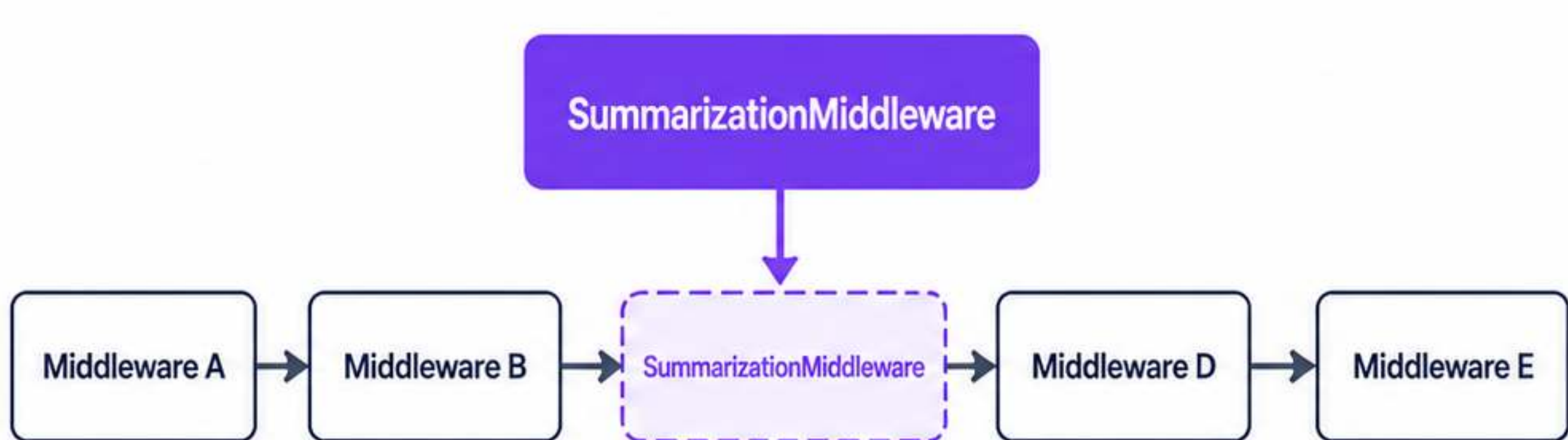
Todo 从默认税费，变成任务策略

没有普遍准确率收益，不代表所有场景都该关闭。



Middleware 可以替换，但不会帮你合并配置

同名实例在原位置替换；替换的是完整对象，不是单个字段。



同名：保持栈顺序，原位替换



不同名：插入核心层之后、尾部能力之前



Backend、阈值、提示词与权限状态都要重新审查



主 Agent 覆盖会影响默认 general-purpose 子 Agent



替换 $\neq$ 合并



Backend



trigger



keep



summary_prompt

文件工具不只变快，工具面和返回语义也变了

新能力、边界控制与迁移风险必须一起看



1 delete

新增默认工具
可递归删除目录

⚠ 高风险副作用



2 write_file

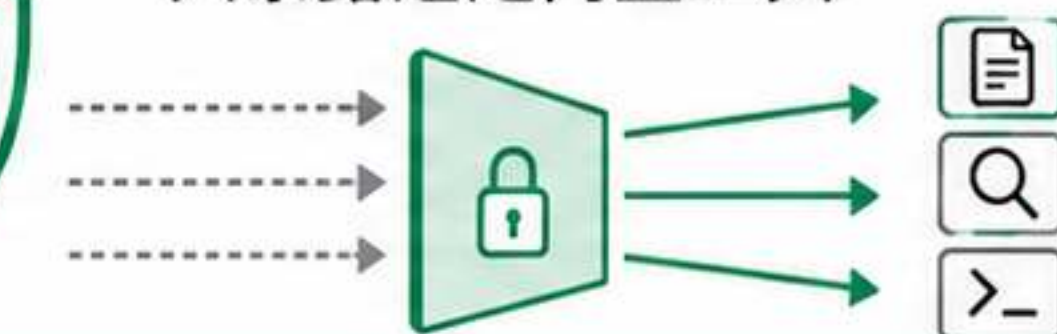
目标存在时完整覆盖

⚠ 先确认再写入



3 tool allowlist

只暴露选定内置工具



🛡 收窄工具面



4 read_file

总行数 + 剩余行数
下一 offset

ℹ 更丰富的分页信息



5 grep / glob

partial results
truncated 标记

ℹ 大结果集更可控



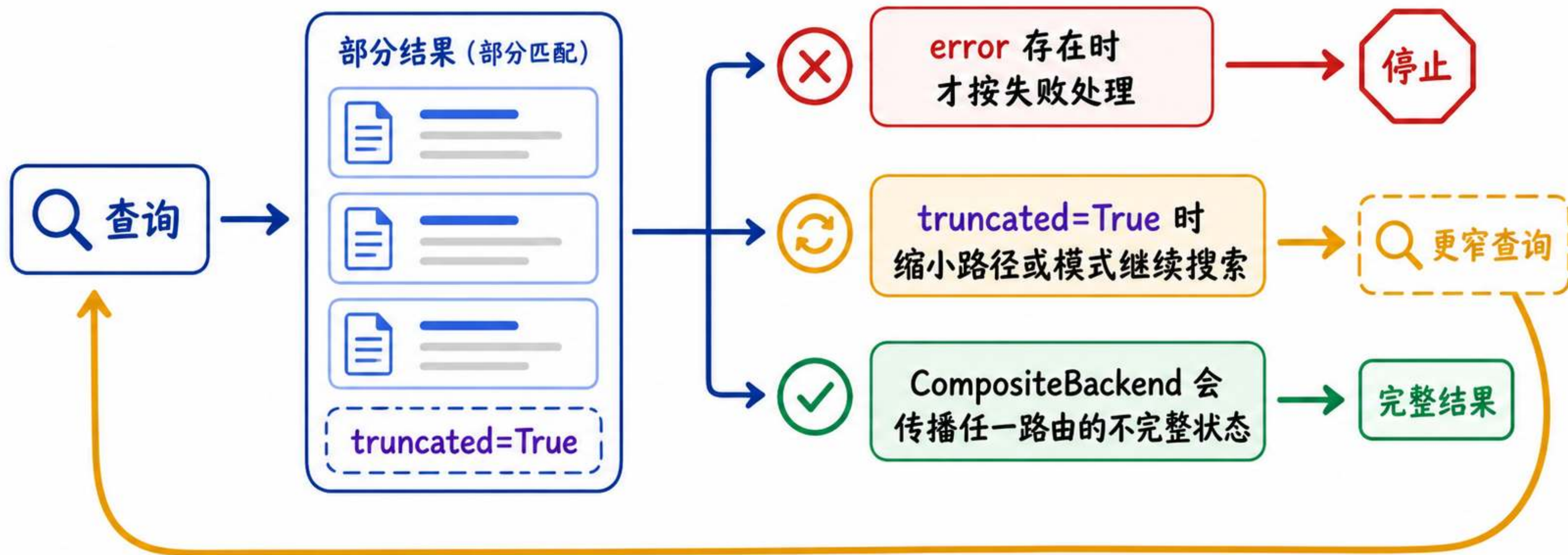
6 grep

最多 1,000 个匹配
流式输出 + context lines

ℹ 更快首屏与上下文

没有报错，不代表搜索结果完整

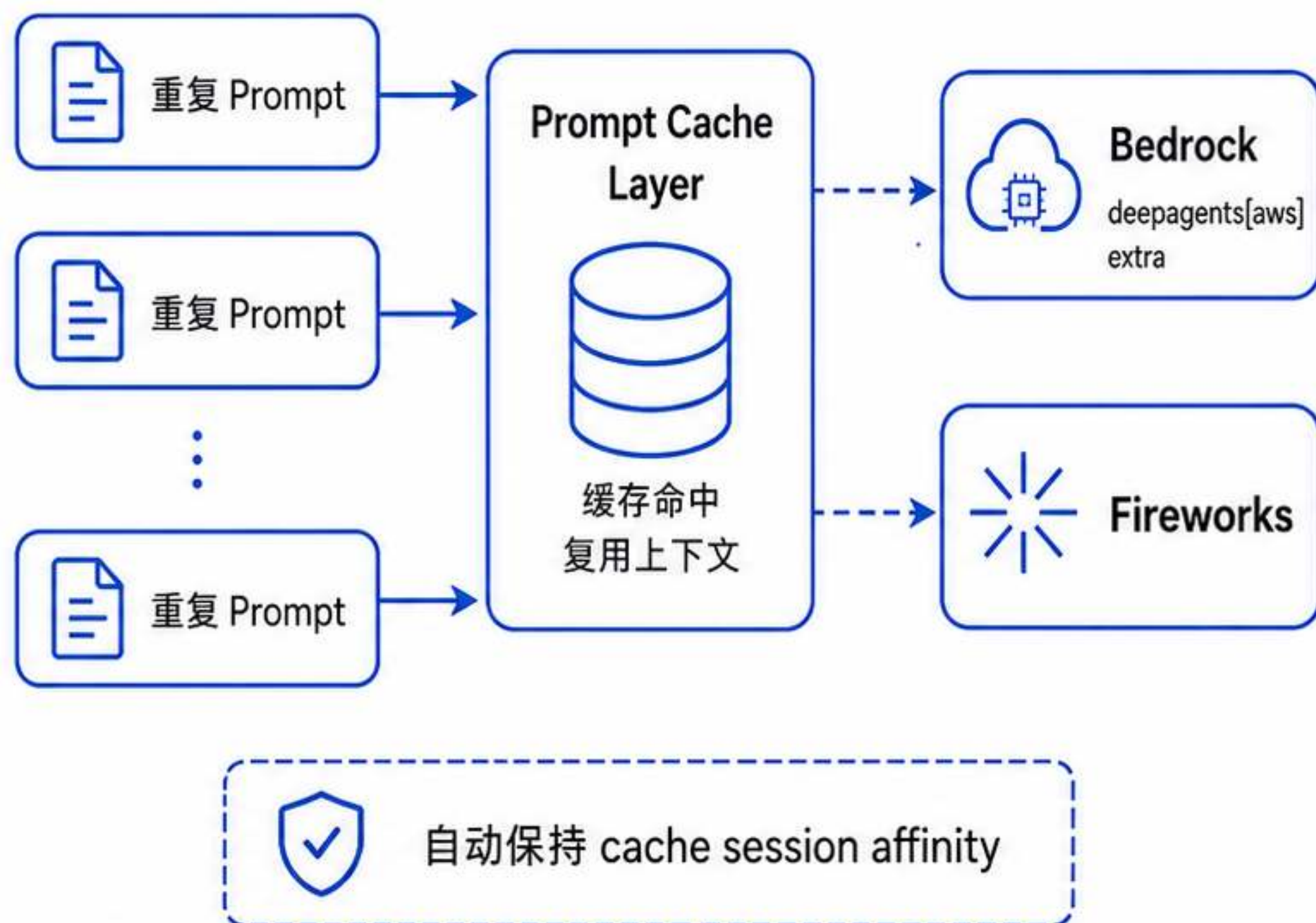
truncated 是“有效但不完整”，不是失败，也不是全集。



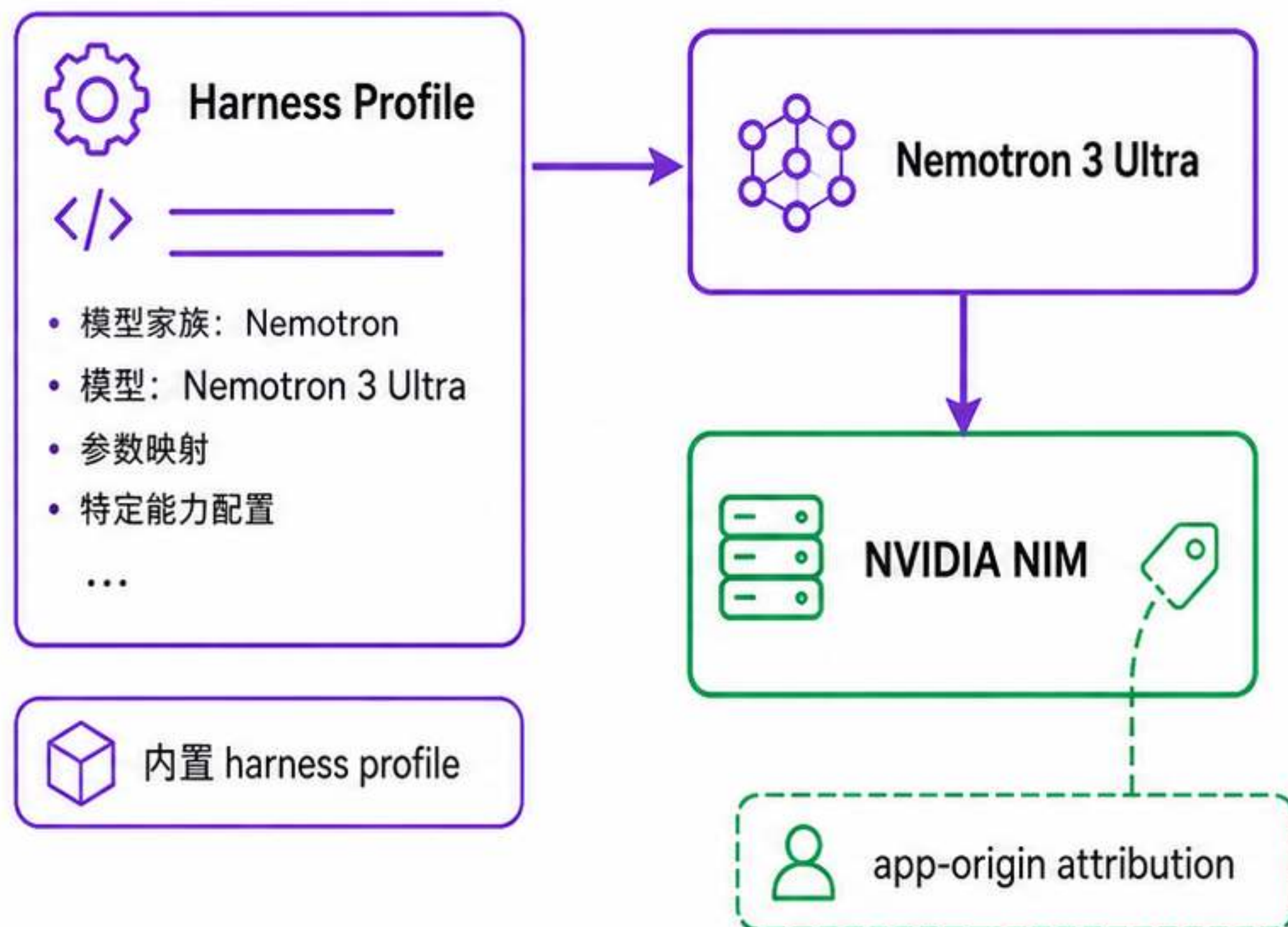
v0.7 还补齐了两类 Provider 能力

Prompt Caching 降低重复上下文成本, Harness Profile 适配特定模型

More Prompt Caching



NVIDIA Support



真正阻塞升级的是兼容层与输出约定

Todo 只是默认值变化；Backend 和解析逻辑会直接报错或悄悄读错

v0.6 (旧)



v0.7 (新)

1	✗ Backend factory	→	✓ BackendProtocol 实例
2	✗ StoreBackend	→	✓ 显式 namespace
3	✗ BackendFactory / BACKEND_TYPES / FileFormat / Unset		✗ 移除
4	✗ list[str] FileData.content	→	✓ 新文件使用 string ; 旧数据下次写入时转换
5	✗ [] / 固定宽度 cat -n gutter	→	✓ No files found / 新 read_file 格式



不要把 ToolMessage 文本当稳定机器协议

Source: v0.7.0 changelog /
PR #4541 / PR #4561

七步迁移，把升级变成可验证实验

日常安装跟随当前 0.7.x；只有复现首发行为才固定 0.7.0



⚠️ ==0.7.0：仅用于首发行为复现



迁移路径
从基线到放量的清晰路线



版本/安装动作
明确的安装与提交物



验证检查点
可度量的产出与验收标准



注意事项
关键限制与使用边界

是否恢复旧默认？先看应用依赖什么

不按怀旧程度决策，按任务、界面和风险边界决策。

应用特征	推荐行动	风险类别
 短任务、无进度 UI	→ Todo 保持关闭	 效率决策
 长程任务或弱模型	→ Todo A/B 后显式启用	 行为决策
 解析原始文件输出	→ 输出迁移列为阻塞项	 兼容性决策
 暴露真实文件系统	→ 审查 delete、覆盖、权限与 HITL	 安全决策
 使用 StoreBackend	→ 显式 namespace 与租户隔离测试	 行为决策

默认层保持轻量，策略由应用选择， 结果由评测验证

