

AIエージェント trace監査

生のprompt、回答、文書、顧客identity、API keyを受け取らず、ボトルネック証拠と同一caseの改善前後を再現可能にします。

synthetic local proofです。顧客案件、本番benchmark、security認証、削減保証ではありません。

問題

AI agentはretrieval、model、retry、toolが重なると、どこが遅く高価なのか分かりにくくなります。trace証拠なしの助言は別の層を最適化する危険があります。

boundary-first方式

- allowlist型JSONL metadata schema
- unknown content field・secretらしい値を拒否
- 標準libraryのoffline analyzer
- aggregateとhashed signatureだけを出力
- step latency/error/retry/token/cost
- 透明で再実行可能なimpact順位
- before/afterは同一case IDが必須
- 不明点と因果claimの限界を明記

synthetic observation

− 65.3%

P95 trace latency

− 57.3%

cost metadata

− 57.2%

input tokens

metric	baseline	optimized
matched cases	5	5
P50 / P95 latency	4,700 / 8,440 ms	2,570 / 2,932 ms
error / retry	5.88% / 1	0% / 0
input tokens	56,500	24,200
提供cost metadata	USD 1.070	USD 0.457

固定監査で納品できるもの

検証済metadata inventory、trace/step map、証拠付き上位bottleneck、優先順位plan、matched replay cases、別途許可された1改善の比較、保守可能な文書runbookです。

主張できないもの

aggregate traceだけでは、root cause、利用者影響、回答の事実性、security/compliance、本番因果、将来性能、金額削減を証明できません。実案件でもこの境界を維持します。