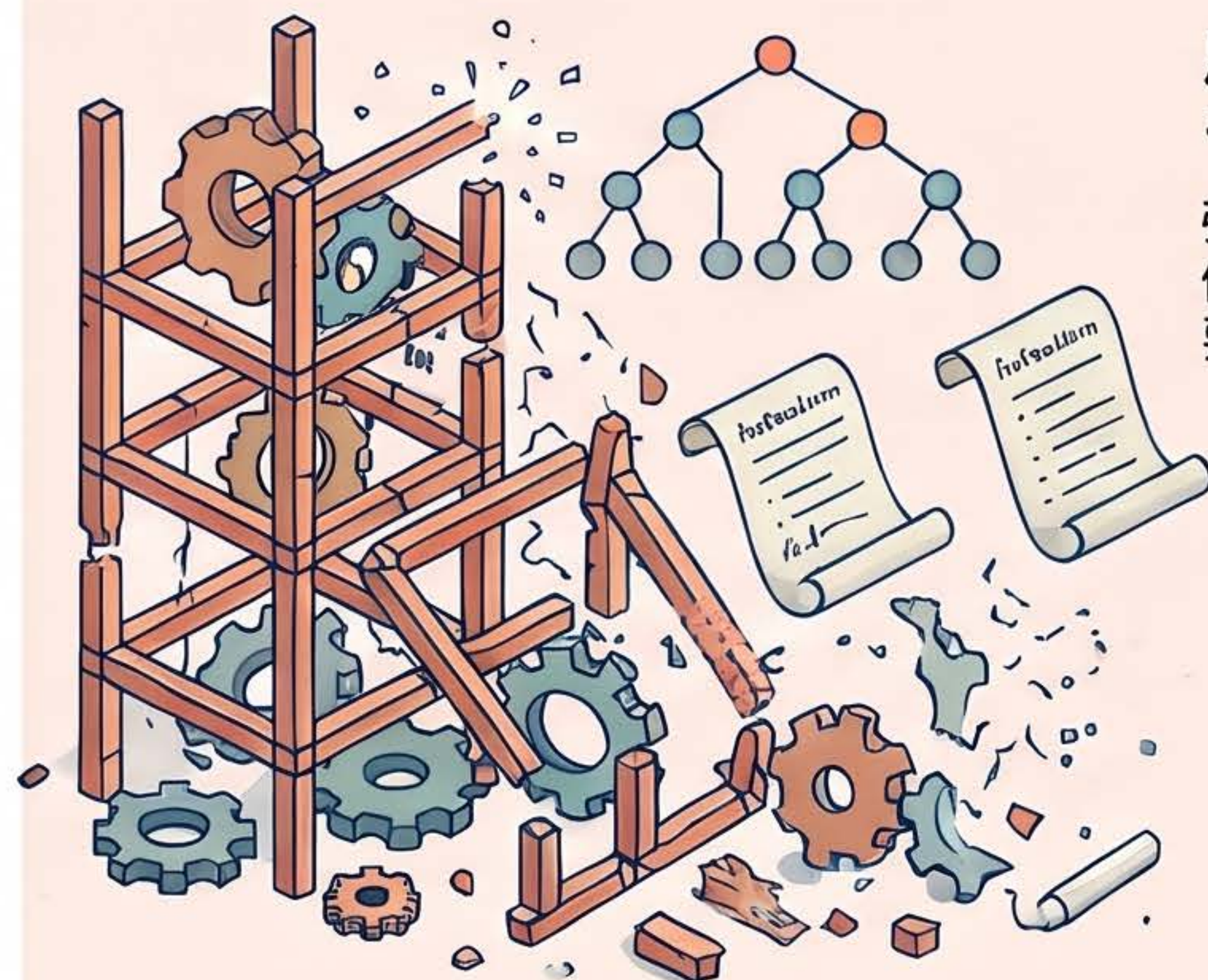


AI高性能化時代のエージェント設計：ハーネスは「不要」ではなく「役割転換」へ

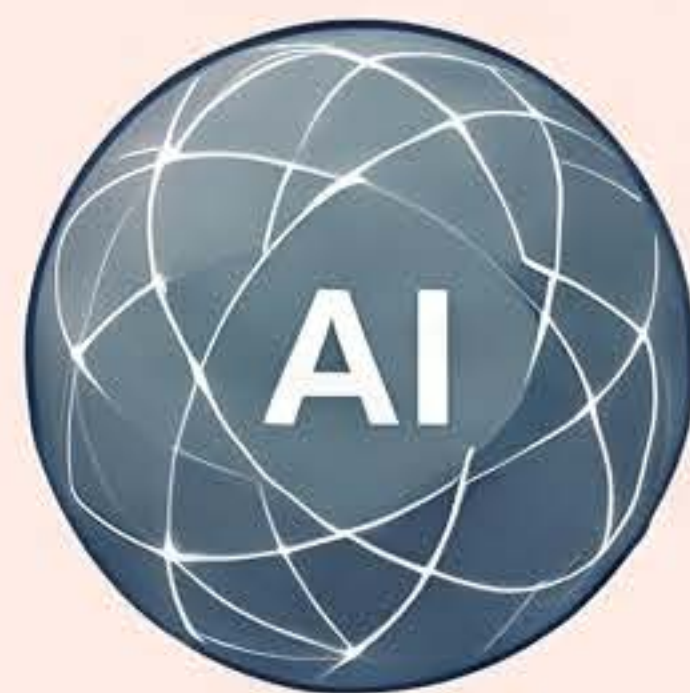
強いAIモデルの登場により、エージェントにおけるハーネスの役割が「思考の補助」から「実行・安全統制」へと変化している本質を伝える。

衰退する「知能補助型」ハーネス (Scaffolding)



思考・計画・ツール選択は
モデル内部へ吸収

強いモデル自身が自律的に探索やエラー
修正を行うため、手書きの計画指示は不要化。



複雑な既存ハーネスを単一エージェント (Malena) が凌駕

相同な計算予算下で、複雑な検索木や複数エージェント構成を統計的に上回る。

統計的に下回る

凌駕



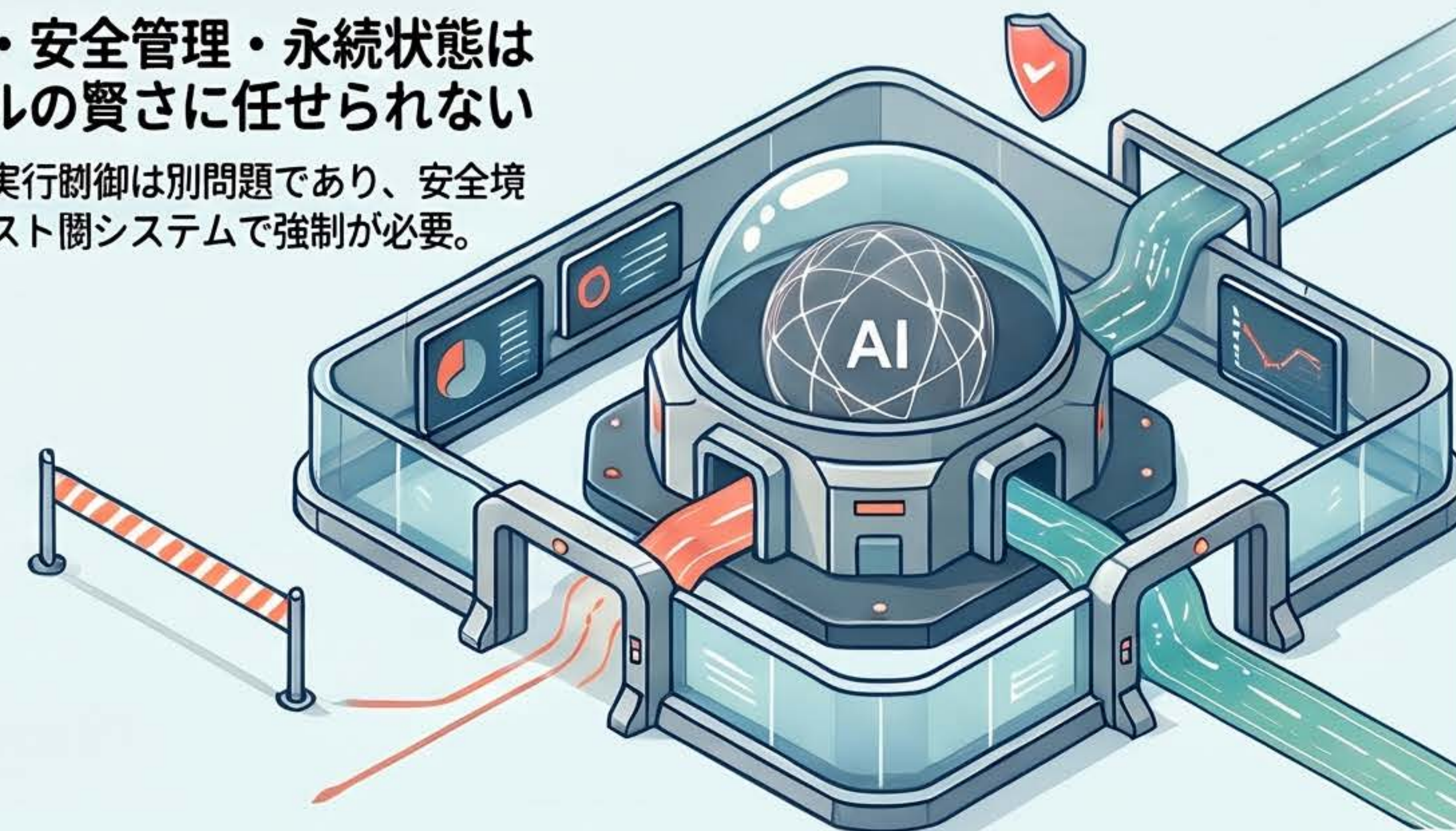
複雑な既存ハーネス
(Multiple Agents, Search Trees)

単一エージェント
(Malena)

重要性を増す「実行・統制型」ハーネス (System Harness)

権限・安全管理・永続状態は
モデルの賢さに任せられない

認識と実行制御は別問題であり、安全境界は
ホスト側システムで強制が必要。



ホスト側ハーネス (CapScope) で不正実行が47回から3回へ激減

権限管理をプロンプトではなく実行層で強制することで攻撃実行を極小化。



不正実行: 47回

プロンプト管理のみ



不正実行: 3回へ激減

ホスト側ハーネス (CapScope) で実行層強制