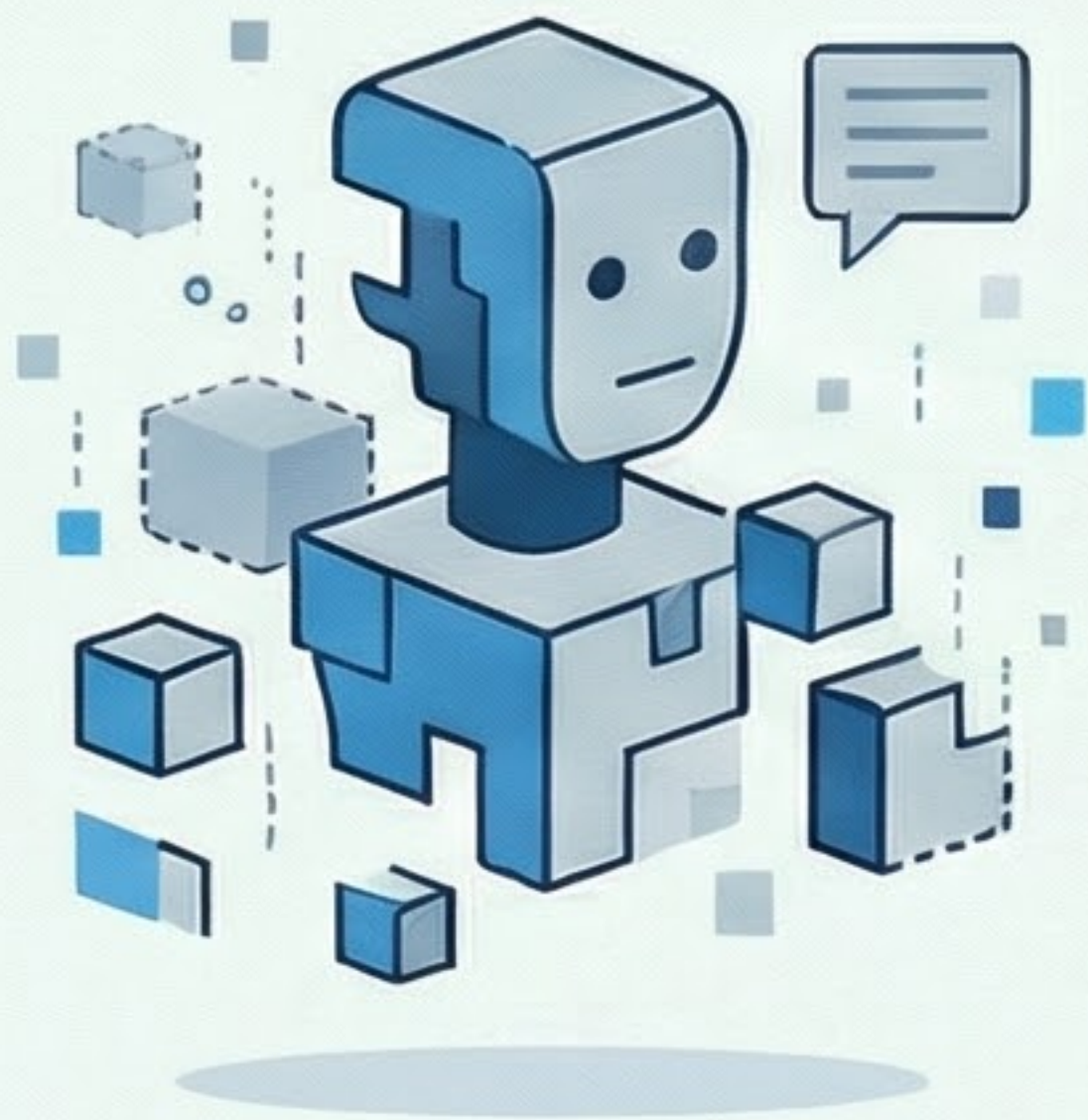


Sophiaフレームワーク：AIに「自己」を与えるSystem 3とは？

現在のAIエージェントは、タスクが終わるごとに記憶がリセットされる「健忘症」という根本的な問題を抱えています。Sophiaフレームワークは、AIにメタ認知能力と長期的なアイデンティティを与える新しい層「System 3」を導入することで、この問題を解決します。

現状のLLMエージェントが抱える「3つの限界」



健忘症的 (Amnesiac)

タスク完了後、内部状態がリセットされ、経験から学ぶことができない。



反応的 (Reactive)

ユーザーからの指示がなければ行動せず、自律的な目標を持たない。



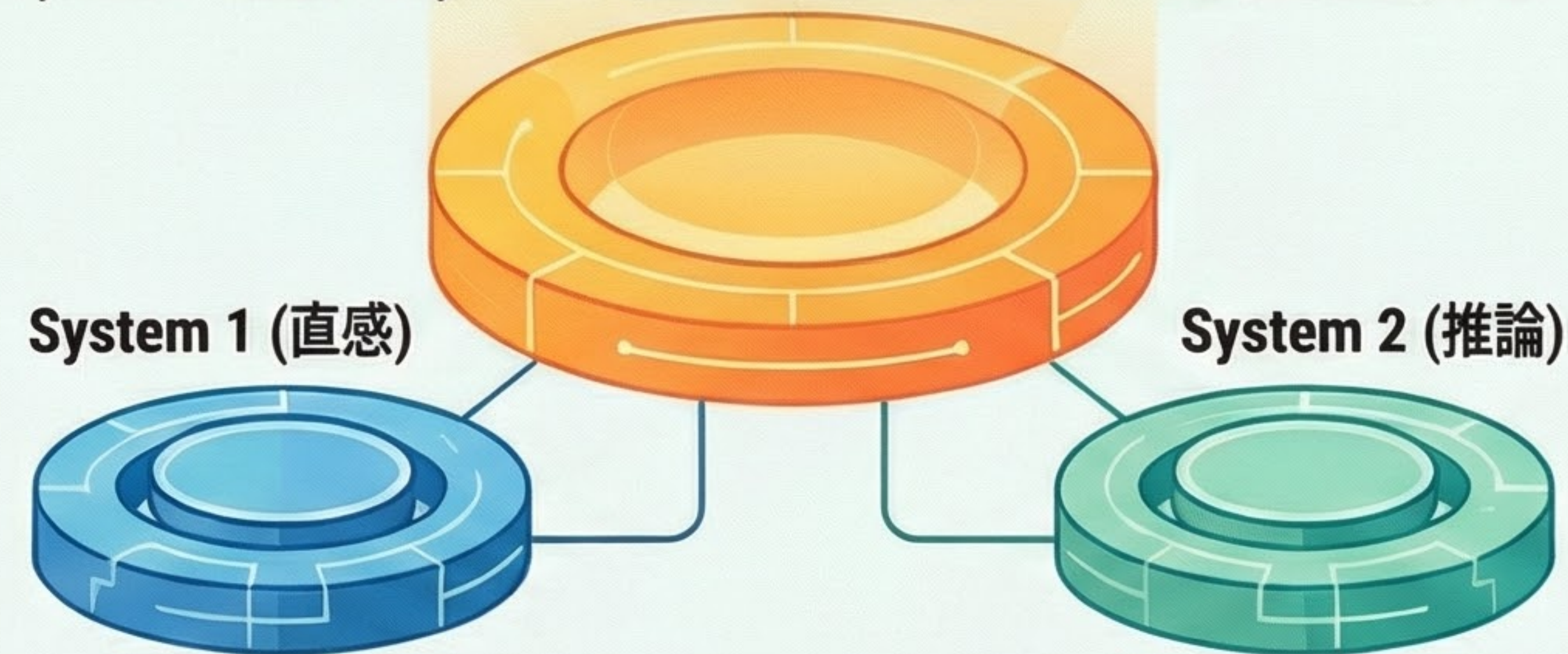
「自己」の欠如 (Lacks a "Self")

思考を監視・改善する上位のメタ認知機能が存在しない。

解決策：メタ認知層「System 3」の導入

System 3：思考を監視する「OS」

System 1 (直感) と System 2 (推論) の上に配置され、AIの全活動を統括する。



永続性を支える「4つの心理学的柱」



1. 心の理論



2. エピソード記憶



3. メタ認知



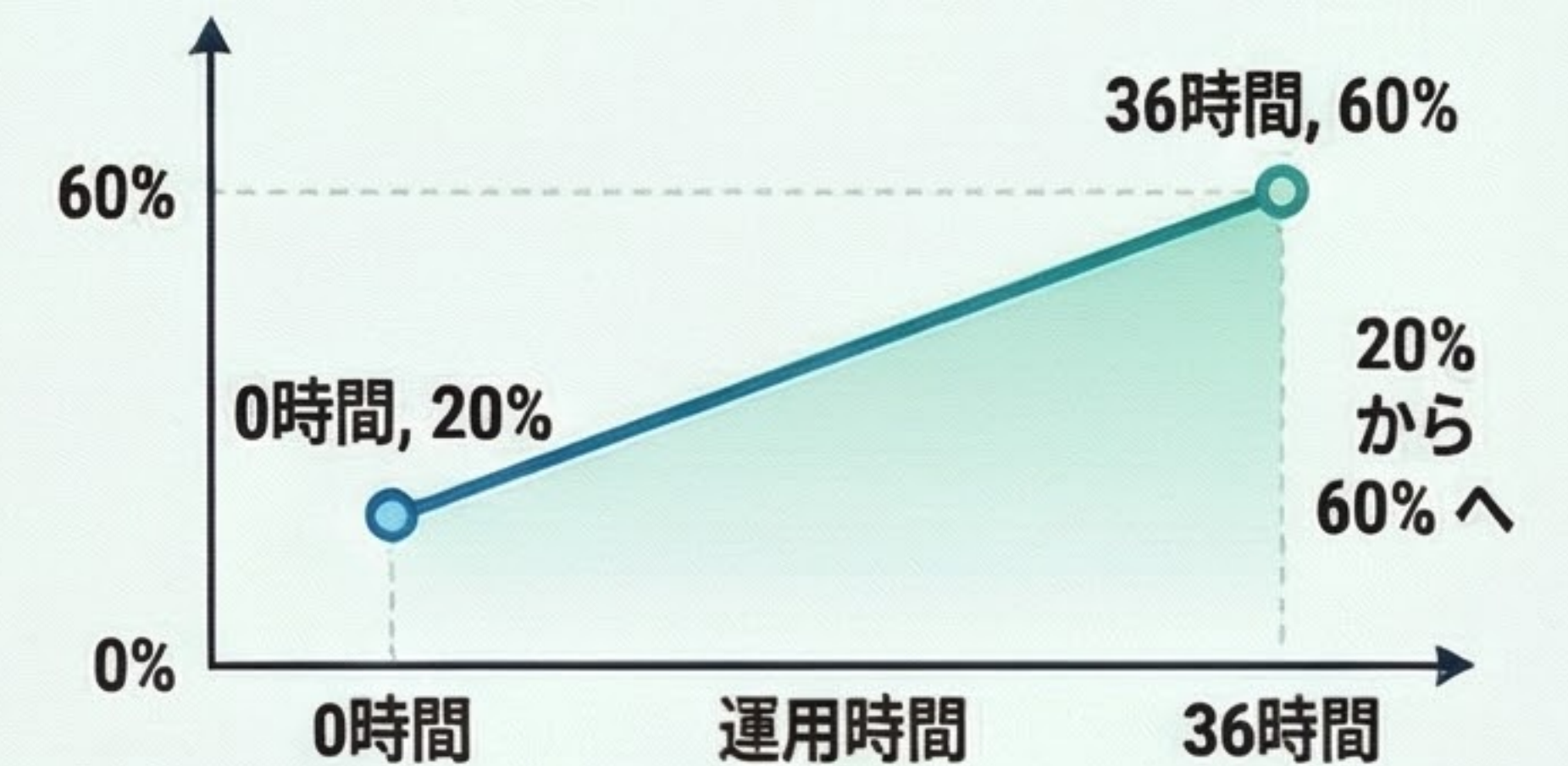
4. 内発的動機付け



ガーディアンLLMによる思考の監査

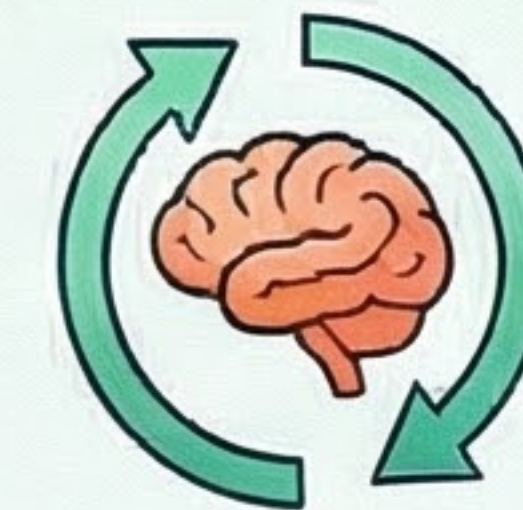
全ての思考プロセスを監視し、論理的矛盾や不安全な思考を即座に剪定する。

実証：36時間の実験で示された驚異的な成果



困難なタスクの成功率が3倍に

運用開始時の20%から、36時間後には60%へと大幅に向上した。



推論コストを約80%削減

一度成功した思考パスを記憶・再利用することで、繰り返し作業を効率化。



アイドル時間での自律的学習

指示がなくても、自らの知識を整理し、新しいスキルを習得し始めた。