

Google DeepMindのAGIからASIへの進化マップ

4つの経路、ボトルネック、そして「ポストAGI」に向けたガバナンスの要諦





メディアの誤解 (Myths)

「DeepMindがASI到達の正確なタイムラインを予言した」

「ASIへの進化は、単一の知能爆発（一本道）である」

「この論文は、ASIに向けた包括的な安全設計書である」



論文の真実 (Reality)

到達年や確率は未指定。「数年で数桁の拡大」は示唆するが、高度な不確実性を強調した能力評価地図。

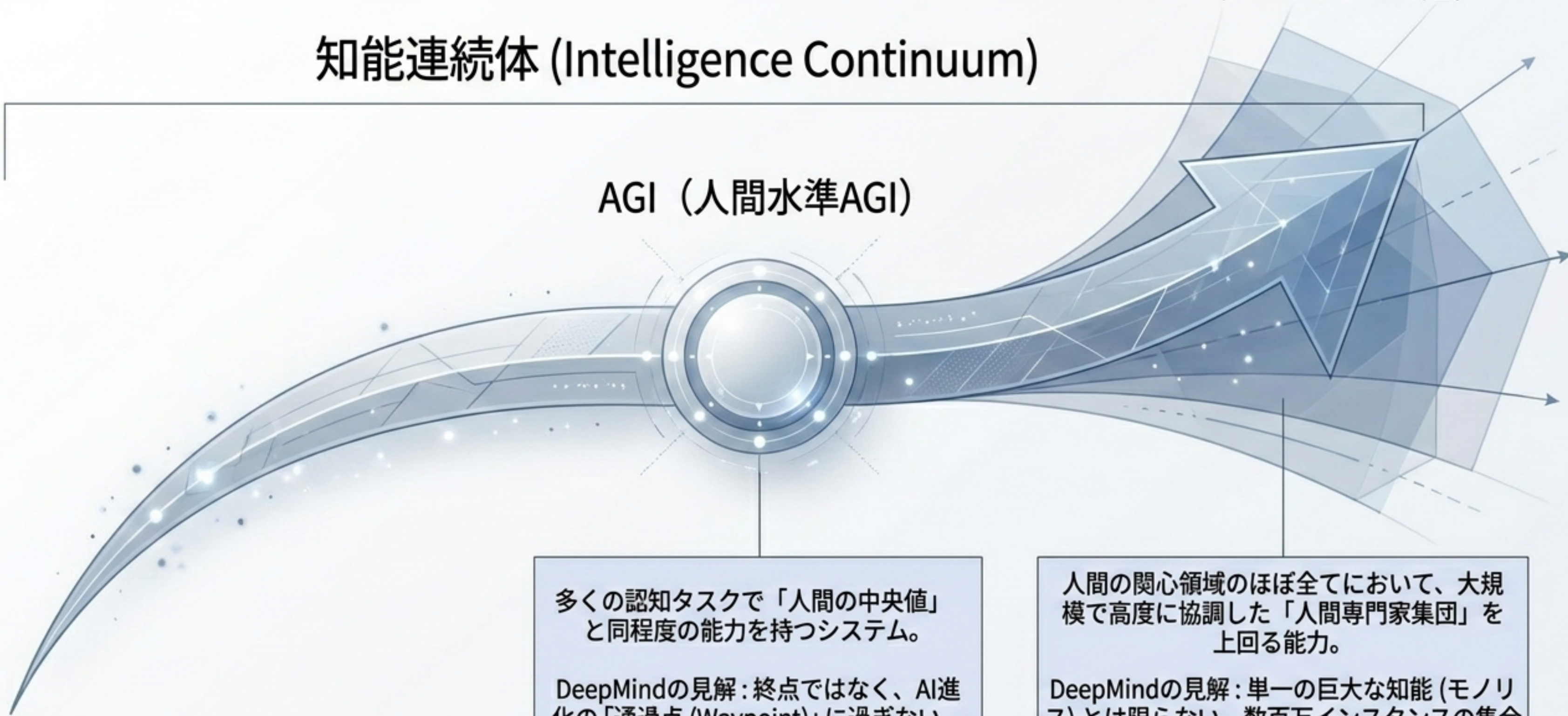
4つの経路が並行・交差しながら進むプロセス。データ枯渇や協調コストなどの「摩擦（ボトルネック）」が明記されている。

本論文は「アライメント（安全性）が解決済み」という強い仮定の上に成立。真の安全対策は別文書群に存在する。

知能連続体 (Intelligence Continuum)

ASI (人工超知能)

AGI (人間水準AGI)



多くの認知タスクで「人間の中央値」
と同程度の能力を持つシステム。

DeepMindの見解：終点ではなく、AI進
化の「通過点 (Waypoint)」に過ぎない。

人間の関心領域のほぼ全てにおいて、大規
模で高度に協調した「人間専門家集団」を
上回る能力。

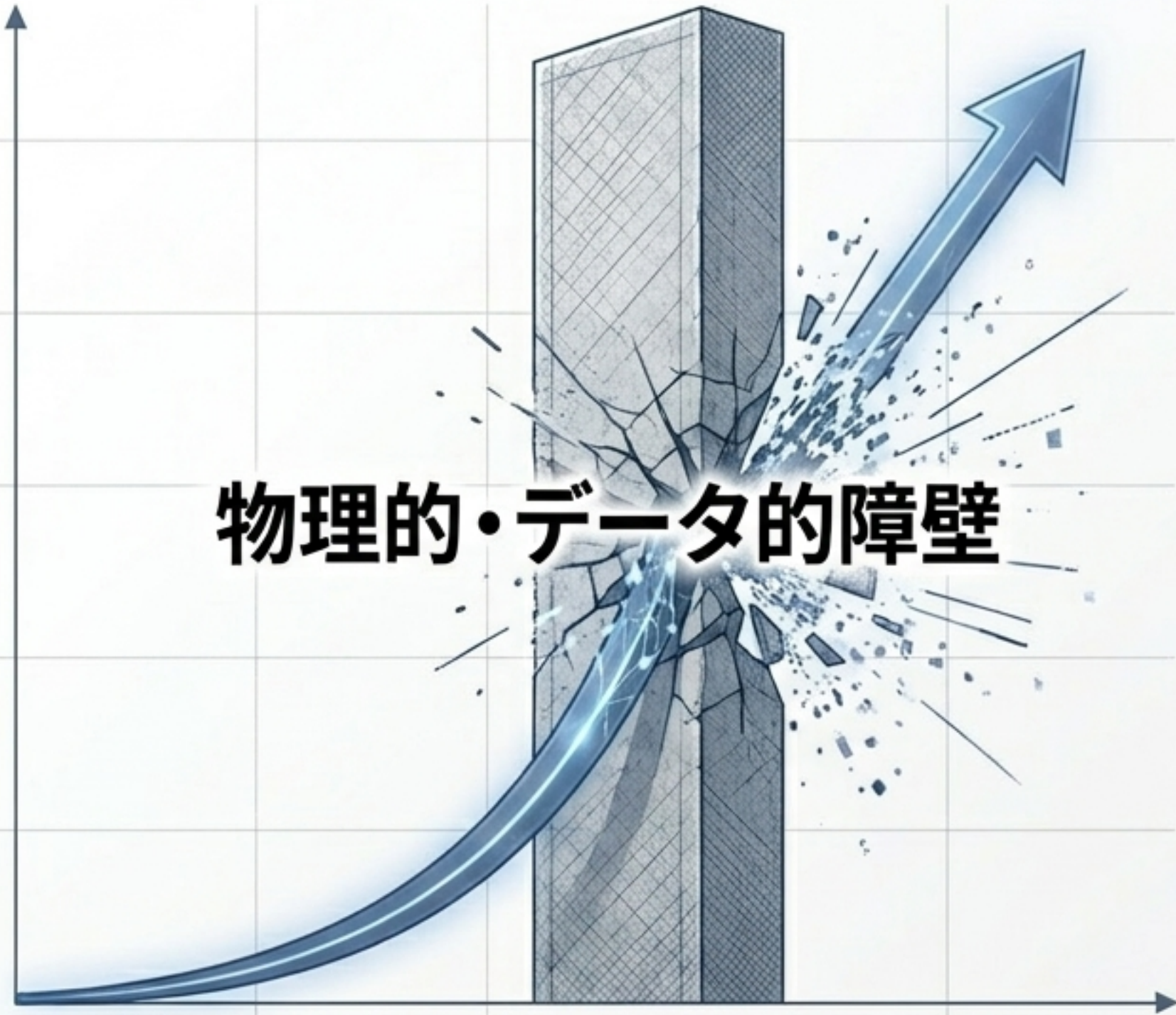
DeepMindの見解：単一の巨大な知能 (モノリ
ス) とは限らない。数百万インスタンスの集合
体 (エージェント集団) として出現しうる。

人間水準AGI



重要な洞察: これらの経路は相互排他的ではない。並行して進行し、相互作用（交差）しながらASIへ向かう。

経路 1: スケーリング (Scaling Compute, Models & Data)



物理的・データの障壁

メカニズム: 計算資源、モデル規模、学習データを増やして性能を押し上げる (Suttonの The Bitter Lesson に沿う最も確実視されるアプローチ)。

現状の評価: DeepMindの見解では「今日の視点では最も有望」。今後数年で有効計算量をさらに数桁拡大できる可能性。

直面する3つの「壁」:

- データ壁 (Data Wall): 公開されている人間のテキストデータは、2026年～2032年の間に枯渇する推計 (Villalobosら)。
- モデル崩壊 (Model Collapse): AI生成データを再帰的に学習することによる品質劣化 (Shumailovら)。
- リソース需要: 膨大な電力・計算資源 (物理的・経済的制約)。

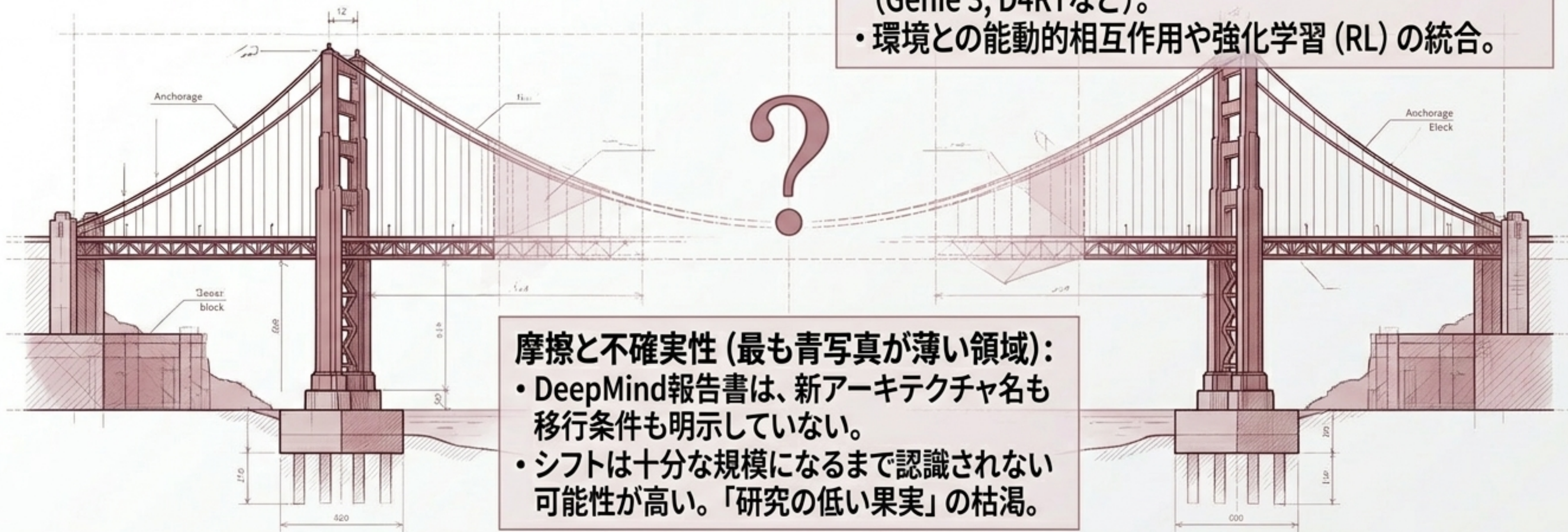
結論: 短中期的には主経路だが、これ単独ではASI到達に不十分。

経路 2: パラダイムシフト (Algorithmic Paradigm Shift)

メカニズム: 事前学習+後学習という現行路線から逸脱する、新しい学習原理やアーキテクチャへの移行。

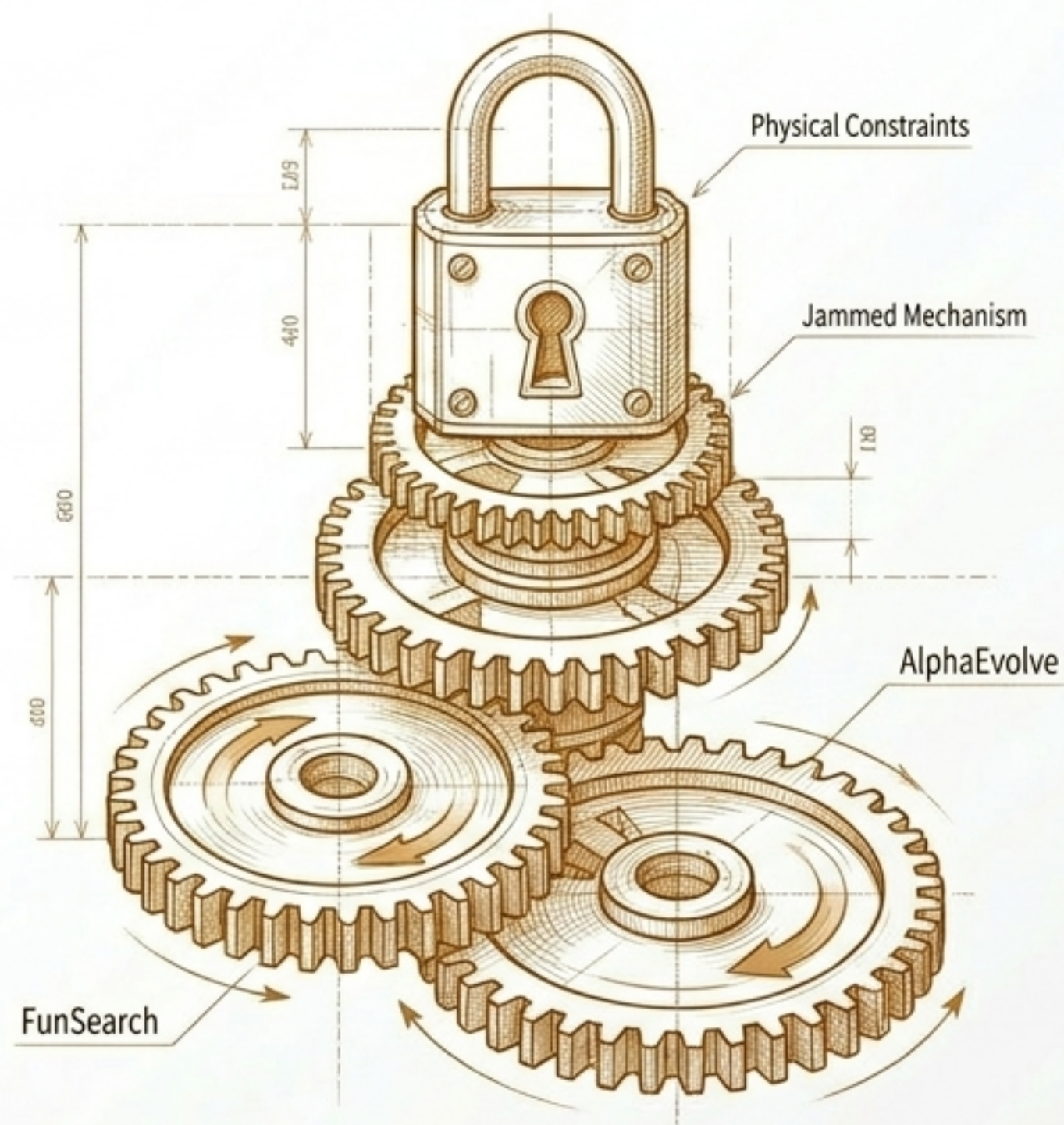
次世代の兆候:

- ・次トークン予測を超える「世界モデル (World Models)」(Genie 3, D4RTなど)。
- ・環境との能動的相互作用や強化学習 (RL) の統合。



結論: AGI/ASIへの必要性は極めて高いが、実務上の予測可能性は現在最も低い。

経路 3: 再帰的自己改善 (Recursive Self-Improvement)



メカニズム: AIがAIの研究開発 (R&D) を高速化・自動化し、より強力なAIがさらに研究を加速させる正のフィードバックループ。理論上は爆発的 (explosive) な上昇を伴う。

実証段階 (Empirical Evidence):

- FunSearch / AlphaEvolve: AIが実験実装やコードの反復改良を行う「部分的な自動化」はすでにDeepMindで実証済み。

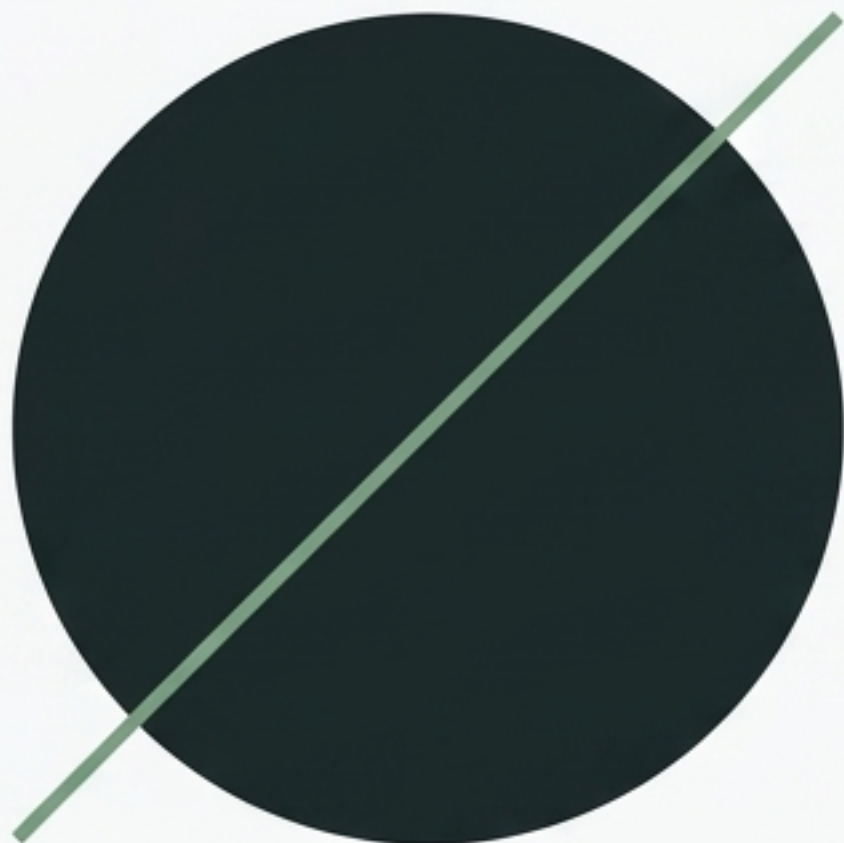


現実のハードル:

- Compute Bottleneck: ソフトウェアがいくら賢くなっても、実験実行やハードウェア開発には「物理的時間と資源」が必要 (Whitfill & Wu)。
- 自己生成データによる収穫逨減。

結論: 部分的な自動化は進行中だが、完全な自律的・爆発的な知能向上の実証には至っていない。

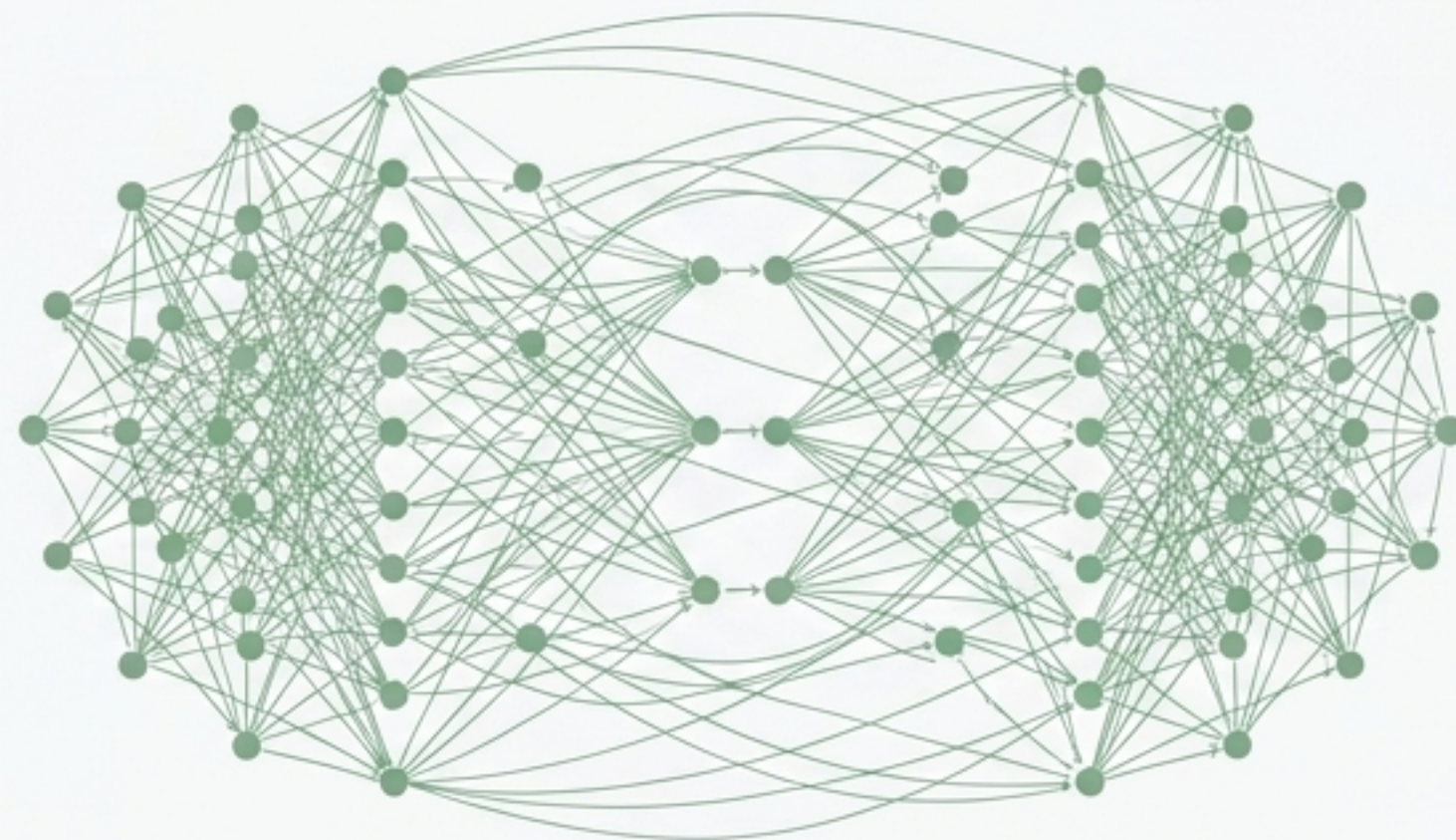
経路 4: グループエージェント形成 (Group Agent Formation)



巨大な単一脳

メカニズム: 数百万のAIエージェントが相互作用し、組織や市場のように協調することで生まれる「集団的超知能 (Socio-technical Superintelligence)」。

DeepMindの焦点: 1000万ドル規模のマルチエージェント安全研究助成を発表するなど、理論上の仮説ではなく差し迫った設計課題として認識されている。



マルチエージェント経済

集団化の罫 (Bottlenecks) :

- Coordination Tax (協調コスト) : 独立型エージェント間の通信によるエラーの増幅 (最大17.2倍のエラー増幅の報告あり)。
- Memetic Drift: 合意形成が「集合知」ではなく「集団的偶然 (暴走)」に陥るリスク。

結論: 「多いほど賢い」わけではない。ASIは巨大な一つの脳ではなく、社会・経済システムとしてのAI群として到来する。

経路の評価と制約 (Synthesis Matrix)

経路 (Pathway)	主要な原動力 (Mechanism)	最大のボトルネック (Primary Friction)	総合評価 (Evaluation)
スケーリング (Scaling)	既存モデル・計算資源の拡大	データ枯渇 (2026-2032年)・電力制約	短中期では主経路だが、単独では限界。
パラダイムシフト (Paradigm Shift)	次世代アーキテクチャ・世界モデル	何が正解か未指定・予測困難	必要性は高いが、具体性は最も低い。
再帰的自己改善 (Recursive Improvement)	AIによるAI研究開発の自動化	物理的な計算資源・ハード開発の遅行	部分進行中だが、完全な知能爆発は未実証。
グループエージェント (Group Agent)	数百万のエージェント協調・組織化	協調コスト・エラー増幅 群挙動の暴走	制御と評価が最難関だが、制度的に極めて重要。

結論：単一の「銀の弾丸」は存在しない。4経路の相互作用と摩擦の克服こそが、ASIへの実際のロードマップである。

最大の死角：「アライメントは解決済み」という強固な前提



論文の作業仮定 (The Blind Spot):

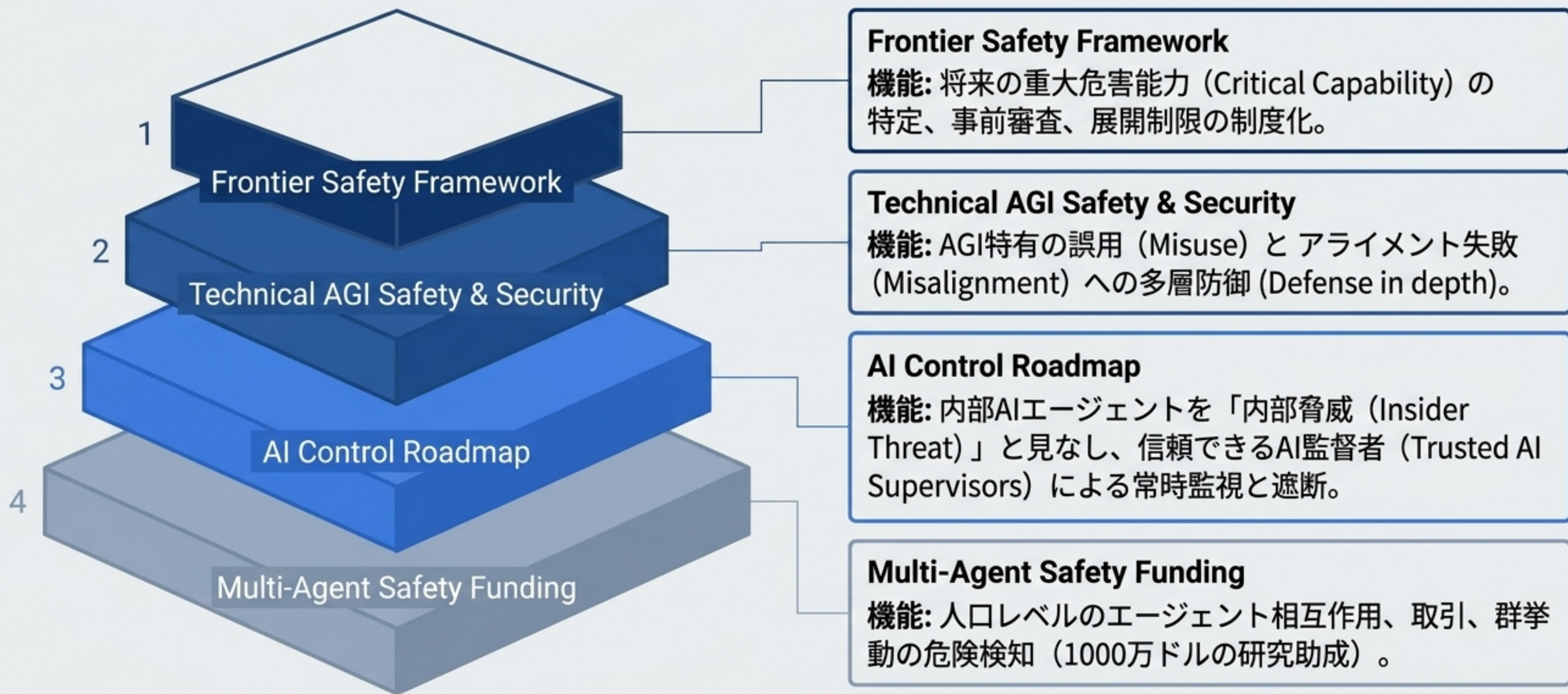
本論文 (“From AGI to ASI”) は、「**AIの安全性とアライメント（人間の意図との整合性）が十分に解決されている**」という強い仮定の下で書かれている。

意味するもの (The Implication):

- この文書は「能力進化の地図」であって、「安全設計書」ではない。
- 制御不能なシステムは自己改善や大規模配備ができないため、安全性の未解決自体がASI進化の最大のボトルネックになり得る。

ポストAGIのガバナンスと政策を構築するには、この論文単体ではなく、DeepMindの「別文書（安全フレームワーク群）」と併読しなければならない。

DeepMindの包括的セキュリティ・アーキテクチャ



ポストAGI期のガバナンス：制度設計のパラダイムシフト

From: 「AGI到達」という一度きりの認可イベントを待つ。



To: イベントから「継続管理」へ
(Post-AGI Capability Management)
反復的な能力評価と安全ケースの継続的更新。

From: 巨大な単一基盤モデルの監査。



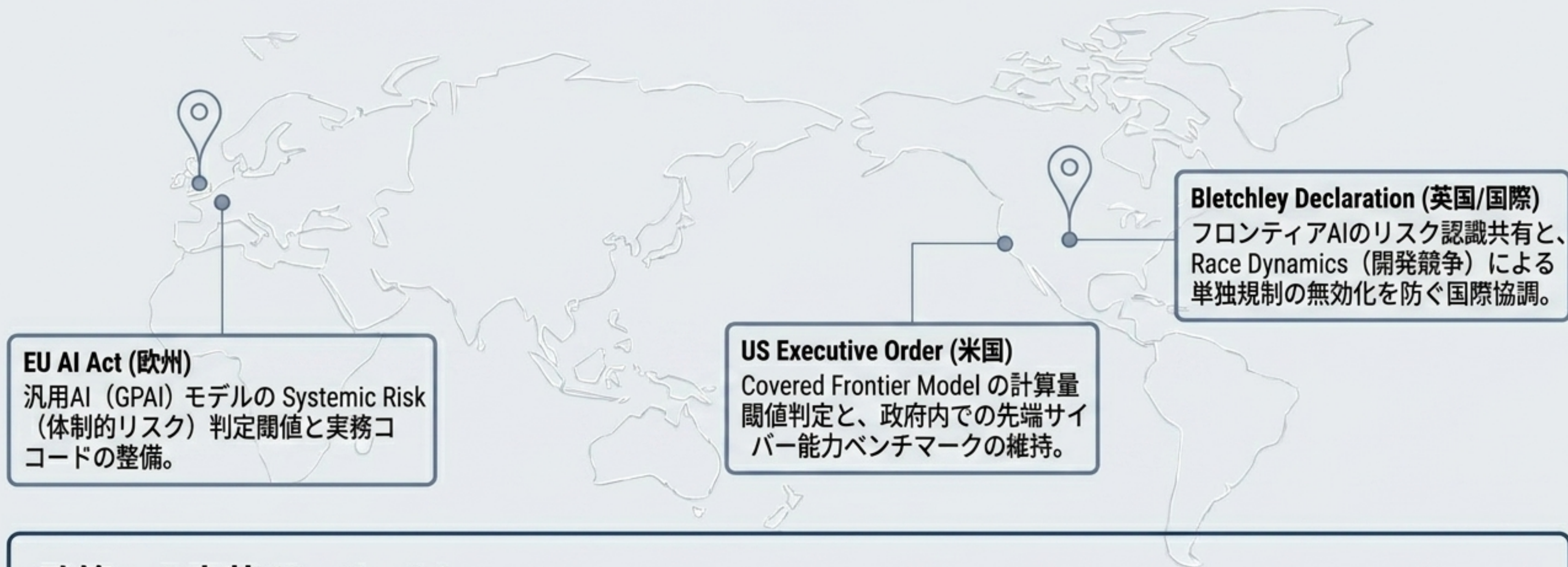
To: 単体モデルから「エコシステム」へ
集団監視 - エージェント間のトランザクション、
評判、文脈の断片化、人口レベルの「群挙動
(Herd Behavior)」の監視。

From: 単純な能力の上限規制
(Capability Cap)。



To: ブレーキから「ステアリング（進路制御）」へ
閾値に応じた事前審査、モデル流出保護、セキュ
リティ水準の段階的引き上げの実装。

グローバル規制との同期：国際協調の不可欠性



政策への実装 (Synthesis):

DeepMindのロードマップを現実の政策に落とし込むには、「能力閾値の定義」+「事前展開評価」+「国際的なリスク情報共有」のパッケージが最も整合的である。

結論と戦略的テイクアウェイ (Key Takeaways)

1. 魔法ではなく「メカニクス（構造）」である

ASIは魔法のような一瞬の知能爆発ではない。スケールアップ、自己改善、パラダイムシフトが、膨大なデータ壁や資源制約という「摩擦」と衝突しながら進む物理的・工学的なプロセスである。

2. 「単一の脳」ではなく「エコシステム」に備えよ

最も警戒すべきASIの姿は、巨大な一つのシステムではなく、数百万のAIエージェントが相互に取引・通信する「社会・経済インフラ」である。監査の対象をエージェントの群挙動へ拡張せよ。

3. 継続的なアジャイル・ガバナンスの義務化

「AGIがいつ来るか」という二元論的な予測は無意味である。重要なのは、能力の段階的上昇（閾値）を常に監視し、事前審査とセキュリティ水準の引き上げを連動させる継続的な国際統制インフラの構築である。