AGIからASIへ：Google DeepMindが描く「超知能」への4つの道

知能の定義：AGIからASIへ

AGI（人工汎用知能）
多くの認知タスクにおいて、平均的な人間と同程度の能力を持つシステム

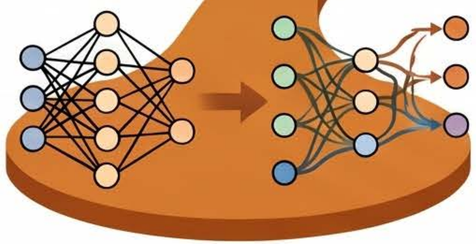
知能は「連続体」である
AGIとASIの間に断絶はなく、知能は連続的に上昇し、複数の進化メカニズムが重なり合う

ASI（人工超知能）
ほぼ全ての領域で、大規模で高度に協調した人間専門家集団を上回る超人的な能力を持つ知能です。

経路1：スケーリング（計算資源・モデル・データ）

既存の基盤モデルの路線を延長し、データ量と計算量を数桁拡大することで性能を押し上げる、現在最も有望視されている経路

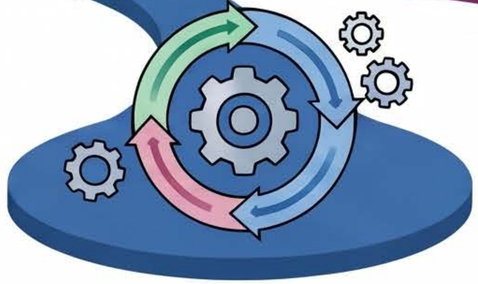
⚠ メカニズムと摩擦・リスク：
計算・データ増大、データ枯渇、莫大なエネルギー消費、収穫逓減



経路2：アルゴリズムのパラダイムシフト

現行の学習原理から逸脱した、新しいアーキテクチャや探索法（世界モデルや相互作用学習など）による進化を目指す

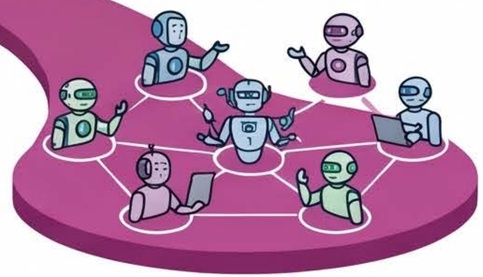
⚠ メカニズム：新しい学習原理
主な摩擦・リスク：内容の未指定、技術統合のコスト



経路3：再帰的自己改善

AI自身がAIの研究開発を高速化・自動化し、知能が爆発的に上昇する正のフィードバックループを形成

⚠ メカニズム：AIによるAI研究加速
主な摩擦・リスク：物理的実験の制約、自己生成データによる劣化



経路4：グループエージェント形成

多数のAGIエージェントが人間組織のように協調し、集団として単一モデルの限界を超えた超知能を実現

⚠ メカニズム：マルチエージェント協調
主な摩擦・リスク：協調コスト、集団的誤作動、予測不能な群挙動

安全性と政策的含意：安全性は「前提条件」

本報告書は「安全対策が十分に解決されている」という作業仮定の上に立っており、実際の展開には別途包括的な安全設計が必要

継続的な能力評価の必要性：
AGI到達判定、事前安全審査、モデル流出防止、エージェント集団の監視を含む反復的な管理

個体から「集団」の監視へ：
グループエージェント経路を想定し、単体モデルだけでなく、エージェント間の相互作用や群挙動の監視体制への移行