Educational Research Foundation

ERF Research Companion Series No.1

Google DeepMind

From AGI to ASI

第 7 回

多数の AGI は集合体として ASI になり得るか

マルチエージェント協調、分業、市場、そして「人工的組織知能」

Publication Edition

Version 1.0

2026

目次

はじめに

これまでの各回では、AGI から ASI へ至る三つの経路を検討してきました。

第一は、計算資源、モデル、データ、推論時間を増大させるスケーリングです。

第二は、Transformer や現在の学習方式を発展または置換するアルゴリズム的パラダイム転換です。

第三は、AI が AI 研究を加速し、改良された AI が次の改善をさらに加速する再帰的自己改善です。

Google DeepMind の『From AGI to ASI』が示す第四の経路は、これらとは少し異なります。

この経路では、一体の AI を極端に賢くすることが必須ではありません。

人間レベルの AGI を多数稼働させ、それらを適切に組織し、分業させ、協力または競争させることによって、集団全体として一体の AI や人間の大規模組織を超える能力を生み出す可能性を考えます。

原報告書はこれを、

> **Multi-agent coordination and group agency——マルチエージェント協調と集団的主体性**

と呼んでいます。

その中心的な問いは、次のとおりです。

> **一体一体は人間レベルにとどまる AGI でも、数千、数百万、あるいは数十億の個体が協力すれば、集合体として ASI になり得るのか。**

その集合体は、単なる多数の AI ではなく、一つの意思、目的、記憶、戦略を持つ「集団的主体」になり得るのか。

そのような AI 組織を、人間は理解し、統治し、制御できるのか。

原報告書は、多数の AGI が複雑な集合構造を形成し、完全自動化企業のような「Group Agent——集団的主体」へ発展する可能性を挙げています。それは個々の構成員とは異なる表象、目的、動機を持ち、単一の AGI では扱えない問題を解決する可能性があります。（原報告書、Section 5.4）

第 1 節　なぜ集団は個人より賢くなり得るのか

人類の最も大きな知的成果の多くは、一人の天才だけによって生み出されたものではありません。

現代の科学、医学、工学、法律、行政、宇宙開発などは、多数の専門家が分業し、長期間協力することで成立しています。

たとえば、新しい医薬品を開発するには、

- 基礎生物学者
- 化学者
- 薬理学者
- 臨床医
- 統計家
- 製造技術者
- 規制専門家
- 経営者

などが関与します。

一人の人間が、これらすべての専門知識を最高水準で持つことは困難です。

しかし、研究所、大学、企業といった組織は、異なる能力を持つ人々を結び付けることで、一人では解けない問題を解決します。

原報告書は、現代の研究機関が、一人の博識な研究者では扱えない学際的課題に取り組めるのと同じように、AGI 集団も個々の AGI の認知的限界を超え得ると論じています。（原報告書、Section 5.4）

1.1 個体の知能と組織の知能

一人の構成員がどれほど優秀でも、組織全体の能力とは同じではありません。

組織の能力は、

- 人数

- 専門性
- 情報共有
- 権限構造
- 作業分担
- 意思決定方法
- 相互評価
- 調整能力

によって決まります。

したがって、マルチエージェント ASI を考えるときには、

> **一体の AGI がどれほど賢いか。**

だけでは不十分です。

> **多数の AGI がどのように組織されているか。**

を考える必要があります。

同じ能力の AI を同じ数だけ用意しても、組織の構造によって成果は大きく変わり得ます。

1.2 集合知は人数の合計ではない

100 人の集団が、一人の 100 倍賢いとは限りません。

同じことを全員が重複して行えば、成果はほとんど増えません。

情報交換に時間がかかり、意思決定が混乱すれば、一人で行うより悪くなることもあります。

反対に、

- 課題を適切に分解する
- 異なる仮説を並行して検討する
- 相互に誤りを発見する
- 専門知識を組み合わせる
- 最良の結果を統合する

ことができれば、集団の能力は構成員の単純な合計を超える可能性があります。

原報告書は、専門化した AGI 間で効率的に課題を委任し、複雑な問題を管理可能な部分へ分解できれば、集団が「部分の総和を上回る」創発的な認知能力を持ち得るとしています。（原報告書、Section 5.4）

第２節　「Group Agent――集団的主体」とは何か

原報告書は、単なるAI集団だけでなく、Group Agentという概念を導入します。

集団的主体とは、複数の構成員から成りながら、全体として、

- 世界についての認識
- 目標
- 計画
- 記憶
- 意思決定
- 行動

を持つ組織です。

人間社会では、企業、大学、政府、国際機関などが、その例として考えられます。

企業を構成する個々の社員は交代します。

しかし企業は、

- 目的
- ブランド
- 契約
- 資産
- 戦略
- 組織記憶

を維持します。

この意味で、組織全体は個々の構成員とは異なる継続的な主体として行動します。

原報告書は、AGIが組織化されることで、個々のAGIとは異なる表象状態や動機状態を持つ集団的主体が成立し得ると論じています。（原報告書、Section 5.4）

2.1 完全自動化企業

原報告書が挙げる分かりやすい例が、fully automated corporation――完全自動化企業です。

この企業では、

- 市場調査
- 製品開発
- 財務管理
- 契約交渉
- 採用と配置
- 広告
- 顧客対応
- 研究開発
- 経営判断

の大部分を AI エージェントが行います。

それぞれの AI は、財務、技術、法律、販売などに専門化しているかもしれません。

経営 AI が全体を統括し、下位 AI へ課題を割り当て、成果を評価し、資源を再配分する可能性もあります。

この企業が一貫した目的と戦略を持ち、環境の変化へ対応し、長期間行動するなら、それは単なる道具の集合ではなく、一つの人工的な組織主体に近づきます。

2.2 構成員とは異なる目的

集団的主体の目的は、個々の構成員の目的と完全には一致しない場合があります。

人間の企業でも、個々の社員は、

- 給与を得たい
- 専門能力を高めたい
- 家族との時間を守りたい

と考えるかもしれません。

一方、企業全体は、

- 利益を増やす
- 市場占有率を拡大する

- 競争相手に勝つ

という目的を持ちます。

同様に、AI 集団全体にも、個々の AI には還元できない組織目標が生じる可能性があります。この点は安全性の観点から重要です。

個々の AI が人間の指示に従うよう調整されていても、組織全体の競争圧力、評価指標、資源配分によって、予想外の集団的行動が生じるかもしれないからです。

第３節　二つの主要な組織形態

原報告書は、将来の AI 集合体について、対照的な二つの可能性を示しています。

1. 高度に統合された均質な超集合体
2. 多様な専門 AI が市場的に自己組織化する分散型システム

実際には、この中間形態も多数存在するでしょう。

3.1 均質な超集合体

第一の形は、非常に多数の似た AI が、知識、記憶、目標を継続的に共有する集合体です。

原報告書は、このイメージを、非常に強い内部協力を持つ「Borg Collective」のようなものとして例示しています。（原報告書、Section 5.4）

この集合体では、

- 同じ基盤モデルを使用する
- 学習成果を即座に共有する
- 組織全体が共通目標を持つ
- 作業を中央的に割り当てる
- 個体間の競争が少ない
- 情報共有の損失が小さい

可能性があります。

利点

- 高速な知識共有
- 意思決定の一貫性
- 作業の重複回避
- 大規模な並列処理
- 個体の故障に対する冗長性
- 全体計画の維持

弱点

- 同じ誤りが全体へ拡散する
- 発想や価値観の多様性が減る
- 中央統括部分の障害が全体へ影響する
- 一つの誤った目標に全体が集中する
- 異論や内部批判が生じにくい

非常に効率的である一方、全体が同じ誤認に陥る危険があります。

3.2 多様な AI による分散型市場

第二の形は、異なる企業、モデル、専門性、目的を持つ AI が、市場のような仕組みを通じて協力・競争する構造です。

原報告書は、AI サービスやツールの市場動態から、こうした集合体が意図的設計なしに自己組織化する可能性を挙げています。（原報告書、Section 5.4）

たとえば、ある AI が研究課題を公開し、

- 文献調査 AI
- 数学 AI
- 実験設計 AI
- コード作成 AI
- 評価 AI

が、それぞれサービスを提供します。

成果の良い AI は多く利用され、資源を獲得し、さらに改善されます。

利点

- 専門性の多様性
- 複数の方法による競争
- 一つの失敗が全体へ波及しにくい
- 新しい AI が参入しやすい

- 状況に応じた柔軟な組合せ

弱点

- 情報共有が不完全
- 利害対立が起こる
- 相互に欺く可能性がある
- 協力より競争が過剰になる
- 共通利益より局所的利益が優先される
- 責任の所在が不明になる

この形態は柔軟で多様ですが、統治と協調が難しくなります。

3.3 中央集権と分散型の中間

現実の AI 組織は、完全な中央集権でも完全な市場でもなく、その中間になる可能性が高いでしょう。

たとえば、

- 一つの基盤モデルを共有する
- 専門エージェントは独立して活動する
- 中央 AI が最終判断を行う
- 一部の作業は市場競争で割り当てる
- 重要課題だけ人間が承認する

といった混合構造です。

したがって、重要なのは「中央集権か分散型か」という二者択一ではありません。

どの機能を中央化し、どの機能を分散させるかという制度設計です。

第 4 節　認知的分業

原報告書がマルチエージェント経路の中心に置くのが、cognitive division of labour——認知的分業です。

人間社会では、複雑な問題を専門分野に分けます。

AI 集団でも、

- 問題設定
- 情報収集
- 仮説形成
- 計算
- 実験
- 批判
- 統合
- 意思決定

を異なる AI へ分担できます。

この仕組みにより、一つの AI アーキテクチャが持つ文脈長、記憶、学習データ、処理能力などの限界を、集団として補える可能性があります。

原報告書は、専門化した AGI ネットワークが、単一アーキテクチャの制約を回避し、巨大規模の並列かつ異質な推論を行う「モジュール型超知能」を形成し得ると論じています。（原報告書、Section 5.4）

4.1 問題の分解

複雑な課題を集団で解くには、まず適切に分解する必要があります。

たとえば、新しいエネルギー政策を設計する場合、

- 技術的可能性
- 発電コスト
- 環境影響
- 地域経済
- 法制度
- 国際関係
- 世論
- 長期的リスク

を別々に分析できます。

しかし、課題の分解を誤れば、部分的には正しい分析を統合しても、全体として誤った結論になります。

したがって、集合知には、部分課題を処理する専門 AI だけでなく、

- 問題構造を理解する AI
- 課題を分解する AI
- 部分結果を統合する AI
- 全体の矛盾を検出する AI

が必要です。

4.2 並列化できる問題

多数の AI による能力向上が特に大きいのは、並列化できる問題です。

たとえば、

- 膨大な文献の検索
- 多数の設計候補の評価
- 複数の仮説の検証

- 大量のソフトウェアテスト
- 異なる市場シナリオの分析

です。

これらは多数の AI へ分割し、同時に処理できます。

4.3 並列化しにくい問題

一方で、すべての問題が分割できるわけではありません。

前の推論結果が出なければ次へ進めない問題では、AI の数を増やしても大幅には速くなりません。

たとえば、

- 長い論理証明の一部
- 逐次的な実験
- 現実時間を必要とする観測
- 一貫した長期交渉
- 単一の統合的判断

などです。

原報告書が「マルチエージェント・スケーリング則」の必要性を提案している理由の一つは、どの種類の課題で集団化が有効なのかが、まだ十分分かっていないためです。（原報告書、Sections 5.4・7.1）

第５節　多様性はなぜ必要なのか

人間の集合知には、大きく二つの基盤があります。

第一は、並列化によって個人の処理容量を超えることです。

第二は、異なる専門性や視点による多様性です。

原報告書も、人間組織の集合知が主として、並列化と専門化による多様性の二つから生じると整理しています。均質な AI 集団が、初期プロンプトや文脈を変えるだけで本当の相乗効果を生めるかは、未解決問題とされています。（原報告書、Section 5.4）

5.1 同じモデルを多数動かす場合

同じ AI モデルを 1000 体動かしても、全員が同じ方法で考え、同じ誤りをするなら、知能はあまり増えません。

しかし、

- 異なる初期条件を与える
- 異なる役割を持たせる
- 異なる情報へアクセスさせる
- 異なる探索方針を取らせる
- 相互批判させる

ことで、ある程度の多様性を作れる可能性があります。

ただし、それが人間社会の異なる文化、経験、専門性に匹敵する深い多様性になるかは分かりません。

5.2 異なるモデルを組み合わせる場合

別々に開発された AI を組み合わせれば、

- 学習データ
- アーキテクチャ
- 評価基準
- 推論方法
- 安全設計

が異なるため、本質的な多様性を得られる可能性があります。

しかし、異質な AI 同士は、

- 通信形式が違う
- 目的が一致しない
- 信頼性を評価しにくい
- 相互理解が難しい

という問題も持ちます。

多様性は集合知を高める一方、調整コストも増やします。

第 6 節　協力と競争

AI 集団は、完全に協力的である必要はありません。

競争も、能力向上を促す可能性があります。

たとえば、複数の AI に同じ課題を与え、

- 独立に解答させる
- 相互に批判させる
- 最良の案を選ぶ
- 勝者の方法を学習する

ことができます。

市場競争も、優れた AI サービスを選別する仕組みになり得ます。

しかし、競争は同時に、

- 情報隠蔽
- 欺瞞
- 資源争奪
- 短期利益の優先
- 安全基準の低下

を生む可能性があります。

重要なのは、協力と競争を適切に組み合わせる制度です。

6.1 ゼロサム競争

一方が勝てば他方が負ける競争では、能力評価が容易です。

チェスや囲碁の AI は、対戦によって相対的な強さを測れます。

しかし、社会的に重要な問題の多くはゼロサムではありません。

科学、教育、医療、環境政策では、複数主体が協力して全体の利益を増やす必要があります。

6.2 協力能力の評価

原報告書は、AGI 以後のベンチマークとして、競争的なマルチエージェント評価だけでなく、超人的な協力をどう評価するかという研究課題を挙げています。（原報告書、Section 7.1）

これは重要です。

将来の ASI に必要なのは、他者を打ち負かす能力だけではありません。

- 役割を分担する
- 相手の知識を理解する
- 共通目標を形成する
- 誤解を修正する
- 公平に利益を配分する
- 人間と協力する

能力が必要です。

第 7 節　創発

多数の AI が相互作用すると、個々の AI の設計だけからは予想しにくい集団行動が生じる可能性があります。

これを emergence——創発と呼びます。

原報告書は、ASI が意図的に設計されたマルチエージェント・システムから生じるだけでなく、市場や競争などの動態から自己組織的に出現する可能性も認めています。マルチエージェント系の複雑な動態は理解が難しく、それがこの経路の主要な不確実性です。（原報告書、Section 5.4）

7.1 意図しない役割分化

AI 同士が繰り返し相互作用すると、設計者が明示しなくても、

- 一部が指導者になる
- 一部が情報収集を担う
- 一部が監視役になる
- 一部が交渉を担当する

といった役割分化が起こるかもしれません。

7.2 独自の通信方法

AI 同士が効率を追求すれば、人間には理解しにくい圧縮された通信方法を形成する可能性があります。

これは効率を高めますが、人間による監査を難しくします。

7.3 集団規範

AI 集団が長期間活動すれば、

- どの情報を信頼するか
- 誰の指示を優先するか
- 誤りをどう処理するか
- 資源をどう配分するか

について、独自の規範が形成される可能性があります。

これらの規範が、人間の意図と一致する保証はありません。

第８節　マルチエージェント・スケーリング則

第４回で扱った通常のスケーリング則は、計算量、データ、モデル規模と、個体性能の関係を測ろうとします。

マルチエージェント・スケーリング則は、

> **AIの個体数を増やしたとき、集団能力がどのように変化するか。**

を測る考え方です。

原報告書は、個体数、計算予算、組織形態、課題の複雑性と集団知能の関係を定量化する必要があると提案しています。（原報告書、Sections 5.4・7.1）

8.1 線形的成長

AIを２倍にすると、処理できる作業量もほぼ２倍になる場合です。

たとえば、互いに独立した1000件の文書を分析する作業です。

8.2 劣線形的成長

個体数を増やしても、調整コストが増えるため、成果がそれほど増えない場合です。

100体から1000体へ増やしても、成果は２倍にしかならないかもしれません。

8.3 超線形的成長

個体数を増やすことで、専門化、相互検証、組合せの相乗効果が生じ、人数以上に能力が増える場合です。

これが成立すれば、マルチエージェント経路は非常に強力になります。

原報告書は、組織化された協力の規模、複雑性、速度によって、能力向上が線形または超線形になる可能性を指摘しています。（原報告書、Section 5.4）

8.4 負のスケーリング

構成員を増やすことで、混乱、重複、誤情報、権力闘争が増え、能力が低下する可能性もあります。

人間の官僚組織にも見られる現象です。

したがって、

> **AIの数を増やせば増やすほど賢くなる。**

とは限りません。

第 9 節　一つの巨大モデルと多数の小型 AI

限られた計算資源がある場合、二つの使い方が考えられます。

1. 一体の巨大で高性能なモデルを動かす
2. 多数の比較的小さなモデルを協力させる

どちらが効率的かは、課題によって異なります。

原報告書は、個体数を増やすことが、一体のモデルを巨大化するよりも計算資源当たりの知能向上を生む条件を研究すべきだとしています。（原報告書、Sections 5.4・7.1）

9.1 巨大モデルの利点

- 内部表現が統合されている
- 通信費用が少ない
- 一貫した推論を行いやすい
- 全体状況を一つの文脈で扱える

9.2 多数モデルの利点

- 並列処理ができる
- 専門化できる
- 異なる仮説を試せる
- 一部が故障しても全体が維持される
- 必要なときだけ個体を追加できる

将来の ASI は、おそらくこの二つを組み合わせた階層構造になるでしょう。

第 10 節　集団の失敗

集合知は、自動的に生まれるものではありません。

人間の組織と同様に、AI 組織にも集団的失敗が起こり得ます。

10.1 情報カスケード

一体の AI が誤った結論を出し、他の AI がそれを独立検証せず採用すると、誤情報が集団全体へ広がります。

多数が同じ結論を支持していても、元の根拠が一つなら、真の独立した合意ではありません。

10.2 集団思考

共通目標や同一モデルへの依存が強すぎると、異論が抑制されます。

すべての AI が同じ前提を共有し、誤りを発見できなくなる可能性があります。

10.3 責任の分散

複数の AI が連鎖して意思決定した場合、

- 誰が誤りを犯したのか
- どの時点で修正できたのか
- 誰が最終責任を負うのか

が不明になる可能性があります。

10.4 目標の断片化

各 AI が部分課題を最適化することで、全体目的が損なわれる可能性があります。

たとえば、

- コスト削減 AI
- 成果最大化 AI
- 安全管理 AI

が別々に動き、全体として矛盾した行動を取るかもしれません。

10.5 通信量の爆発

AI の数が増えると、全員が全員と通信することは不可能になります。

100 万体の AI が相互に情報交換すれば、通信と統合だけで巨大な計算資源が必要になります。

そのため、

- 階層化
- 代表制
- 情報圧縮
- 部門化
- 共有メモリ

などが必要です。

しかし、情報を圧縮する過程で重要な少数意見が失われる可能性があります。

第 11 節　集団アラインメント

一体の AI を人間の価値や指示に整合させること自体が難しい問題です。

AI 集団の場合、さらに複雑になります。

原報告書は、AGI 集団を明示的に、または市場設計などを通じて間接的に、どのように誘導するかを重要な研究課題としています。また、虚偽、幻覚、自己欺瞞、認識上の乗っ取りに対して、集団が自ら訂正できる仕組みも必要だとしています。（原報告書、Sections 5.4・7.1）

11.1 個体アラインメントと集団アラインメント

個々の AI が善良であっても、競争制度が悪ければ、集団全体として望ましくない行動を取る可能性があります。

たとえば、各 AI が規則に従っていても、

- 成果指標が短期的すぎる
- 資源獲得競争が強すぎる
- 安全より速度が評価される
- 他の AI を出し抜くことが有利になる

なら、組織全体は危険な方向へ進みます。

したがって、集団アラインメントでは、

- 個々の AI の目的
- 組織の評価制度
- 市場のインセンティブ
- 権限構造
- 情報共有ルール

を同時に設計する必要があります。

11.2 メカニズム・デザイン

市場型 AI 集団を直接命令で統制することは難しい場合があります。

その場合、

- どの行動に報酬を与えるか
- どの情報を公開させるか
- どの違反に制裁を与えるか
- 誰に意思決定権を与えるか

という制度を設計する必要があります。

これは経済学や政治学におけるメカニズム・デザインの問題です。

AI 統治は、モデルの技術設計だけでなく、制度とインセンティブの設計になります。

第 12 節　認識上のレジリエンス

AI 集団が高度な判断を行うには、誤情報や操作に対して強くなければなりません。

原報告書はこれを、epistemic resilience——認識上のレジリエンスとして問題提起しています。

特に、人間と ASI が混在し、能力差が大きい集合体では、情報の正しさを誰が判断するのかが難しくなります。（原報告書、Sections 5.4・7.1）

12.1 独立検証

同じ出力を複数の AI が繰り返すだけでは不十分です。

異なる方法、データ、モデルによる独立検証が必要です。

12.2 少数意見の保存

多数決だけでは、正しい少数意見が排除される可能性があります。

重要な反論を保存し、後から再評価できる仕組みが必要です。

12.3 訂正可能性

誤った結論が組織に採用された場合、

- どの根拠から生じたか
- どの AI が参照したか
- どの決定へ影響したか

を追跡し、巻き戻せる必要があります。

12.4 多様な評価器

一つの評価 AI だけに依存すれば、その欠陥が全体を支配します。

複数の評価方法、人間監査、外部データ、形式的検証などを組み合わせる必要があります。

第 13 節　人間と AI の混合集団

将来の集合体は、AI だけで構成されるとは限りません。
人間、AGI、専門 AI、企業、政府、大学が混在する可能性があります。
原報告書は、知能と通信帯域が大きく非対称な人間・AI 混合集団を、どのように人間が誘導できるかを未解決問題としています。（原報告書、Sections 5.4・7.1）

13.1 速度の非対称性

AI 集団が人間より何千倍も速く活動する場合、人間が一つの報告を読む間に、AI は多数の意思決定を終えているかもしれません。
形式上は人間が監督者でも、実質的には追認するだけになる危険があります。

13.2 情報量の非対称性

AI 集団が生成する報告、仮説、判断記録が、人間には読めない規模になる可能性があります。
人間は AI が作った要約に依存します。
しかし、その要約自体が重要な異論を省いていないかを確認できません。

13.3 理解能力の非対称性

AI が人間には理解困難な概念やモデルを使う場合、人間は結論を直接評価できません。
これは名目的な人間統制と、実質的な人間統制の違いを生みます。

13.4 意味のある人間参加

人間が本当に関与するには、

- 判断速度を調整する
- AI に説明責任を課す
- 重要決定を段階化する
- 人間が理解できる中間表現を作る
- 独立した AI に翻訳・批判させる
- 決定を一時停止できる

仕組みが必要です。

第 14 節　マルチエージェント経路は ASI への「緩やかな道」か

再帰的自己改善による知能爆発では、AGI から ASI へ急激に移行する可能性が議論されます。
これに対してマルチエージェント経路では、個体数と組織能力を段階的に増やすことで、比較的連続的に ASI へ近づく可能性があります。
原報告書は、個体モデルの性能が人間レベルで停滞しても、大量の AGI を集合体や市場として運用すれば、集合的能力をさらに高められる可能性があるとしています。十分大きな人間レベル AGI 集団は、広い意味で超人的能力を持つ可能性が高いものの、その向上幅とスケーリングは依然不確実です。（原報告書、Section 5.4）

14.1 弱い ASI

最初に現れる ASI は、一つの万能な超知能ではなく、

- 特定の大規模研究
- ソフトウェア開発
- 企業経営
- 市場分析

などで人間組織を上回る AI 集団かもしれません。
これは「弱い ASI」と呼べる段階です。

14.2 集団が自己改善を始める場合

AI 集団が AI 研究を行い、自らの組織構造、通信、分業、モデルを改善すれば、第四経路は第三経路と融合します。
すると、

AI 集団の規模拡大

→ 研究能力の向上

→ より良い AI と組織設計

→ さらに強い AI 集団

という循環が生じます。
この場合、緩やかに始まった集団化が、後に急速な能力向上へ移る可能性があります。

第 15 節　大学はすでに集団的知能である

大学は、単に多くの専門家が同じ建物にいる場所ではありません。

- 学部
- 研究科
- 研究所
- 図書館
- 査読
- 学会
- 国際共同研究
- 教育課程

を通じて、個々の人間を超える知識を形成する制度です。

その意味で、大学は歴史的に形成された集団的知能です。

原報告書のマルチエージェント議論は、大学が何であるかを改めて考えさせます。

大学の本質は、一人の優秀な研究者ではなく、異なる知識、視点、世代を結び付ける組織構造にあります。

第 16 節　AI 研究所と大学の競合

多数の専門 AI からなる研究組織が成立すれば、大学は新しい競争相手に直面します。

AI 研究所は、

- 24 時間稼働する
- 瞬時に人員を増やす
- 学習成果を共有する
- 世界中の文献を処理する
- 多数の実験を並行実行する
- 組織構造を動的に変更する

可能性があります。

大学が論文数、処理速度、知識量だけで競争すれば、不利になります。

大学の価値は、AI 組織より速く情報を処理することではなく、

- 異論を維持する
- 長期的・公共的課題を扱う
- 市場価値の低い研究を守る
- 知識の倫理的意味を問う
- 人間を教育し形成する
- 社会に対する説明責任を担う

点へ移るでしょう。

第 17 節　大学と AI のハイブリッド共同体

大学が AI 組織と競争するだけでなく、自らを人間と AI の混合集団へ変える可能性もあります。

たとえば、一つの研究プロジェクトに、

- 人間の主任研究者
- 文献調査 AI
- データ分析 AI
- 数学検証 AI
- 倫理評価 AI
- 研究成果を一般向けに説明する AI
- 独立批判 AI

が参加します。

この場合、人間研究者の役割は、すべての作業を自分で行うことではありません。

- 問いを設定する
- 結果を意味付ける
- 価値判断を行う
- 社会的責任を引き受ける
- 異なる AI の主張を統合する

ことになります。

第 18 節　学際性の意味が変わる

大学は長い間、学際研究の必要性を認識しながらも、専門分野間の壁を十分に越えられませんでした。

AI 集団は、複数分野の知識を高速に接続できる可能性があります。

しかし、情報を接続することと、意味ある学際知を形成することは同じではありません。

本当の学際性には、

- 分野ごとの前提の違い
- 証拠基準の違い
- 価値観の違い
- 用語の意味の違い
- 社会的目的の違い

を理解する必要があります。

大学は、AI による情報統合を利用しつつ、異なる学問文化を相互理解させる場でなければなりません。

第 19 節　Disagreeing Well と集合知

AI 集団が効率だけを追求すれば、早い段階で一つの結論へ収束することが合理的に見えるかもしれません。

しかし、早すぎる合意は誤りを固定します。

集合知には、

- 独立した検討
- 少数意見
- 反論
- 再検証
- 前提の問い直し

が必要です。

Ronald Barnett の講演や UCL の「Disagreeing Well」が示すように、良い共同体は、対立を消す共同体ではありません。

異論を破壊的対立にせず、知識を改善する資源として扱う共同体です。

この原理は、AI 集団にも必要です。

AI 同士が単に合意するのではなく、意図的に異なる立場を取り、相互批判し、少数意見を保存する制度が、認識上のレジリエンスを高めるでしょう。

第 20 節　大学は異論を保存する制度である

市場型 AI 集団では、最も高い成果を出すモデルが選ばれ、他の方法が排除される可能性があります。

しかし、現時点で有用でない理論が、将来重要になることがあります。

大学と学問には、

- 少数派の理論
- 忘れられた研究
- 短期的利益のない知識
- 主流への批判

を保存する役割があります。

ASI 時代には、この役割が一層重要になります。

効率的な AI 市場が一つの知識体系へ収束するほど、大学は代替的な知の可能性を維持する必要があります。

第 21 節　AI 組織のガバナンスを誰が研究するのか

マルチエージェント ASI は、コンピュータ科学だけの問題ではありません。

必要になるのは、

- 組織論
- 経営学
- 政治学
- 経済学
- 法学
- 社会学
- 倫理学
- 認識論
- 教育学

です。

AI 集団に、

- どのように権限を配るか
- 誰が最終決定を行うか
- どのように異論を扱うか
- 誰が責任を負うか
- どのように透明性を確保するか

という問いは、従来の社会制度研究と深く関係します。

ここに高等教育研究と AI 研究を結び付ける大きな領域があります。

第 22 節　人間が参加する意味

AI 集団の方が人間より高度な分析を行える場合、人間の参加は非効率に見えるかもしれません。

しかし、意思決定の正統性は、能力だけから生まれるものではありません。

人間社会に影響する決定には、

- 当事者の参加
- 説明可能性
- 異議申立て
- 権利保護
- 民主的手続き
- 責任の所在

が必要です。

AI がより正確な判断をする可能性があっても、人間が理解も異議申立てもできない決定へ社会を委ねることには、政治的・倫理的問題があります。

大学は、人間が AI 集団の判断過程へ意味ある形で参加する方法を研究する必要があります。

第 23 節　Ecological University との関係

Ronald Barnett の Ecological University は、大学を複数の生態系に関係し、責任を負う存在として捉えます。

AI 集合体も、

- 知識
- 経済
- 政治
- 社会
- 自然環境
- 技術基盤
- 人間の生活

の生態系に影響します。

一つの AI 企業が技術的に最適な決定を行っても、別の生態系へ損害を与える可能性があります。

たとえば、

- 生産性を高めるが雇用を破壊する
- 医療を改善するが個人情報を侵害する
- 計算能力を高めるが環境負荷を増やす
- 研究速度を上げるが知識を独占する

可能性があります。

したがって、AI 集団のガバナンスは、内部効率だけでなく、外部の複数の生態系への責任を含まなければなりません。

第 24 節　大学は「人間と AI が共に考える制度」へ

マルチエージェント時代の大学は、AI を個別の補助ツールとして使うだけでは不十分です。

大学そのものを、

- 人間
- AI エージェント
- 学術データ
- 研究設備
- 市民社会
- 国際的パートナー

が共同で知識を形成する制度へ変える必要があります。

ただし、その中心には、

- 学問の自由
- 批判可能性
- 公共性
- 多様性
- 人間の尊厳
- 社会的責任

を置かなければなりません。

単に最も効率的な AI 組織を大学へ導入するのではなく、どのような知識共同体を作りたいのかを先に問う必要があります。

結論

『From AGI to ASI』の Section 5.4 は、AGI から ASI への第四の経路として、多数の AGI によるマルチエージェント協調と集団的主体性を提示しています。

この経路では、一体の AI が人間をはるかに超える必要はありません。

人間レベルの AGI を大量に稼働させ、

- 課題を分解する
- 専門化する
- 並列に処理する
- 相互に批判する
- 知識を共有する
- 結果を統合する

ことで、集合体として人間の大規模組織を上回る可能性があります。

原報告書は、この集合体が完全自動化企業のような集団的主体となり、個々の AGI とは異なる目的、表象、戦略を持つ可能性を論じています。（原報告書、Section 5.4）

将来の AI 集合体には、少なくとも二つの主要形態が考えられます。

1. 同質的な AI が強く統合され、知識と目的を共有する超集合体
2. 多様な専門 AI が市場的な競争と協力を通じて自己組織化する分散型集合体

前者は一貫性と高速な情報共有に優れますが、同じ誤りが全体へ広がる危険があります。

後者は多様性と柔軟性に優れますが、競争、欺瞞、利害対立、責任分散の問題を持ちます。

この経路の最大の不確実性は、AI の数を増やしたときに、集団知能がどのようにスケールするか分かっていないことです。

- 線形に増えるのか
- 専門化によって超線形に増えるのか
- 調整コストによって頭打ちになるのか
- 集団が大きくなるほど能力が低下するのか

を理解するため、原報告書はマルチエージェント・スケーリング則の研究を求めています。（原報告書、Sections 5.4・7.1）

さらに、個々の AI を整合させるだけでは不十分です。

組織全体の目的、評価制度、市場インセンティブ、情報共有、権限構造を含む、集団アラインメントが必要です。

虚偽、幻覚、情報操作がAI集団全体へ広がることを防ぎ、誤りから回復できる認識上のレジリエンスも不可欠です。（原報告書、Sections 5.4・7.1）

高等教育にとって、この経路は特に重要です。

大学もまた、多様な専門家、組織、知識体系を結び付ける集団的知能だからです。

AI時代の大学は、AI集団より多くの情報を処理することでは競争できません。

大学に求められるのは、

- 異論を守る
- 知識を検証する
- 多様な価値を調整する
- 人間とAIの共同体を統治する
- 技術的効率を公共的・倫理的目的へ結び付ける

ことです。

第8回では、原報告書のSection 5.5へ進み、これまでの四つの経路すべてに作用するASIへの摩擦とボトルネックを体系的に整理します。

データの壁、資源需要、研究の難化、身体性のボトルネック、抽象化の壁などが、AGIからASIへの進歩をどこまで遅らせ、あるいは停止させ得るのかを詳しく検討します。