Educational Research Foundation

ERF Research Companion Series No.1

Google DeepMind

From AGI to ASI

第９回

AGI以後の未来に、私たちは どう備えるべきか

四つの経路、六つのボトルネック、
研究アジェンダ、そして原報告書の最終結論

Publication Edition

Version 1.0

2026

目次

はじめに

Google DeepMind の『From AGI to ASI』は、ASI の実現時期を予測する報告書ではありません。

この報告書の目的は、人間レベルの AGI が実現した後も AI の能力が向上すると仮定した場合に、どのような技術的経路があり、何がその進歩を加速し、何が減速または停止させるのかを整理することにあります。

これまでの解説では、報告書が提示する四つの経路を確認しました。

1. 計算資源、モデル、データのスケーリング
2. アルゴリズム的パラダイム転換
3. 再帰的自己改善
4. マルチエージェント集合知

そして、第８回では、これらに作用する六つの主要な摩擦・ボトルネックを検討しました。

1. データの壁
2. 経済的・自然資源需要
3. 現在のニューラル・パラダイムの限界
4. 研究の難化
5. 抽象化の壁
6. 意図的な減速

さらに、物理世界での実験、観測、製造が必要になる「身体性・現実接地のボトルネック」も重要な制約として扱われました。

原報告書の Section 6 と Section 7 は、これらを統合し、次の問いに答えようとします。

四つの経路と六つのボトルネックを総合したとき、AGI 以後の AI 進歩について何が言えるのか。

AI の能力向上を、どのように測定し、予測し、監視すべきか。

AGI から ASI への移行を理解するために、今からどのような研究を進める必要があるのか。

社会、大学、研究機関、政府、企業は、どのような準備をすべきなのか。

著者たちの最終的な立場は、慎重ですが明確です。

未来を正確に予測することはできません。

しかし、不確実だからといって、何も研究できないわけではありません。

むしろ、不確実性を構成する要素を分解し、測定可能な研究課題へ変え、継続的に予測を更新する必要があります。論文は、四つの経路と摩擦の影響を確定的な予測ではなく、今後検証すべき「開かれた研究プログラム」として位置付けています。（原報告書、Sections 6・7）

第 1 節　Section 6 の中心問題

量的スケーリングだけで ASI に到達できるのか

Section 6 で著者たちが改めて取り上げる中心的な問いは、

> **AGI に、より多くの実効計算量を与え続けるだけで ASI に到達できるのか。**

という問題です。

理論的には、予測を仮説空間の探索、計画を行動方針の探索として捉えるなら、より多くの計算資源によって探索範囲を拡大し、性能を高められる可能性があります。

Universal AI や AIXI の近似においても、より多くの計算を与えれば、理論上は Universal Intelligence へ近づく方向に進みます。

しかし、現実には単純な総当たり探索は、非常に早い段階で計算不可能な規模になります。

そのため実用的な AI は、

- どの仮説を優先するか
- どの可能性を無視するか
- どの探索経路が有望か
- どの情報を重要とみなすか

という帰納的バイアスや事前知識を使います。

これらは探索効率を大きく高めますが、同時に、AI が検討できる仮説や解法の範囲を制限します。

ある段階までは有効だった帰納的バイアスが、より高度な知能へ進む際には天井になる可能性があります。

したがって、現実の AI では、

より多くの計算資源

だけでなく、

計算資源を有効な知能へ変換する新しい質的革新

が継続的に必要になると著者たちは考えています。（原報告書、Section 6）

1.1 スケーリングと質的革新の循環

この議論から、AGI から ASI への進歩は、次のような循環になると考えられます。

1. 現在のモデルやアルゴリズムをスケーリングする
2. 能力が向上する
3. 収穫逓減や技術的限界が現れる
4. 新しいアーキテクチャや学習方法を導入する
5. 効率が改善し、再びスケーリングが可能になる
6. 新しい限界が現れる

つまり、量的拡大と質的革新は対立するものではありません。

現実の AI 進歩は、両者の反復によって生じる可能性があります。

1.2 現在の推論時スケーリングの限界

近年の AI では、学習済みモデルに追加の推論計算を与えることで能力を高める方法が注目されています。

しかし原報告書は、現在のモデルに無制限に推論計算を与えても、比較的早く性能が頭打ちになると指摘します。

現在の推論時スケーリングだけで、人間レベル AGI をそのまま ASI へ押し上げられるとは考えにくいということです。

それでも、計算資源の増加には別の使い道があります。

一体のモデルをさらに長く考えさせるのではなく、

- 多数の AGI を同時に動かす
- 異なる仮説を並行して検討させる
- 集団や市場を形成する
- 専門化と分業を行う

ことで、集合体として超人的能力へ進める可能性があります。

著者たちは、個体 AI の能力が人間レベル付近で停滞しても、数百万以上の AGI を組織できれば、人間の大規模組織を上回る可能性があると論じています。（原報告書、Section 6）

第２節　AGI と ASI の境界は、個体ではなく組織にあるかもしれない

この報告書の最も重要な視点の一つは、ASI を一体の巨大な超知能としてだけ考えないことです。
仮に一体一体の AI が人間レベルを大きく超えなくても、

- 高速に複製される
- 24 時間稼働する
- 知識を共有する
- 多数の課題を並列処理する
- 柔軟に組織を変更する

ことができれば、集団全体として ASI に相当する能力を持つ可能性があります。
人間社会でも、大学、企業、政府、研究機関は、個人を超える知的能力を持ちます。
しかし人間組織には、

- 教育に長い時間が必要
- 情報共有が遅い
- 記憶が不完全
- 調整コストが大きい
- 人数を急速に増やせない

という制約があります。
デジタル AGI 集団では、これらの一部が大幅に弱まる可能性があります。
したがって、Section 6 の重要な問題は、

> **一体の AGI をどこまで賢くできるか。**

だけではありません。

> **どの種類の課題において、AGI 集団は個体能力を超える集合知を形成できるのか。**

です。
研究開発の大部分が、適切に分解、並列化、統合できるなら、個体 AGI の能力が停滞しても、集団化によって ASI 領域へ進める可能性があります。
反対に、重要課題の多くが高度に逐次的で、一人または一つの統合的主体による深い理解を必要とするなら、集団化の効果は限定的です。

第３節　AI能力の進歩とは何を意味するのか

Section 6では、「進歩」という言葉自体も問い直されています。

AI研究では、進歩はしばしば、

- ベンチマーク得点
- 損失関数
- 正答率
- パラメータ数
- 計算効率

によって測定されます。

しかし、AGI以後の社会で重要なのは、狭い技術指標だけではありません。

原報告書は、研究・技術上の進歩を、理論的発見が社会的に利用可能な技術へ変わることまで含むものとして考えています。

その一つの指標として、AIによる自動化が、従来なら100年かかる経済成長を10年へ圧縮するような「economic compression――経済的圧縮」を例示しています。（原報告書、Section 6）

しかし、社会的・文化的進歩を、単純な技術指標やGDPだけで測ることはできません。

著者たちは、社会文化的文脈における「進歩」を、

- 個人と集団の自律性を維持・向上させる
- 人間の繁栄と尊厳を促進する
- その社会において広く有益と認められる

社会技術的発展として捉えています。（原報告書、Sections 6・7）

これは非常に重要な限定です。

AI能力が向上しても、

- 人間の自律性が失われる
- 権力が極端に集中する
- 社会参加が弱まる
- 人間の尊厳が損なわれる

なら、それを無条件に社会的進歩とは呼べません。

第４節　原報告書の研究アジェンダ

Section 7.1 は、この報告書の中でも特に重要な部分です。

著者たちは、AGI から ASI への移行を理解するための研究課題を、七つの領域に整理しています。

1. スケーリングの摩擦とボトルネック
2. 定量的予測
3. ASI のベンチマーキング
4. 再帰的自己改善の動態
5. マルチエージェント・スケーリング
6. 超知能の理論的基礎
7. AI 安全性、アラインメント、社会文化的問題

著者たちは、これらを一つの専門分野だけでは扱えない、世界規模の学際的研究課題と位置付けています。（原報告書、Section 7.1）

第５節　研究領域１

スケーリングの摩擦とボトルネック

第一の研究領域は、第８回で扱った摩擦を、測定可能な問いへ変えることです。

5.1 データの壁

研究すべき問いは次のとおりです。

- データ収集量をどこまで増やせるのか
- AI によるデータ生成は、基盤モデルの継続的改善に役立つのか
- 人間が AI を使うことで生成するデータは、新しい学習資源になるのか
- AI 生成データは既存能力の反復にとどまるのか
- 他者の経験を観察するだけで、AI は安全に計画や行動を学べるのか
- 第三者データから学ぶことで、自己欺瞞や誤った世界理解が生じないか

原報告書は、データの量だけでなく、第三者の経験から学ぶことが、どの条件で有効または危険になるかを研究課題としています。（原報告書、Section 7.1）

5.2 計算資源と知能の関係

次に必要なのは、

> **どの種類の課題で、計算量の増加が知能向上につながるのか。**

を明らかにすることです。

より多くの計算が、

- 一部の限定的課題だけに有効なのか
- 広範な知能向上につながるのか
- 個体能力より集団能力の向上に有効なのか

- モデル改善とどの程度代替可能なのか

を測定する必要があります。

また、

- 量的スケーリング
- アルゴリズム改善
- ハードウェア効率
- 経済価値

を一つのモデルの中で扱わなければなりません。

5.3 パラダイム転換の予測

真の技術的パラダイム転換は、本質的に予測困難です。

それでも、

- 現在のモデルに不足する機能
- 現在のアーキテクチャの理論的限界
- 計算効率の停滞
- 継続学習や世界モデルの欠如

を研究すれば、次の転換が必要になる条件を部分的に予測できるかもしれません。

原報告書は、特定の技術方式に依存しない AGI・ASI の基礎理論を発展させることを求めています。（原報告書、Section 7.1）

5.4 研究の難化

研究課題は、

- AI 研究がどの程度難化しているか
- 一定の能力向上に必要な研究資源がどの速度で増えているか

- AIが研究を何%効率化すれば、この難化を相殺できるか
- 人工研究者の増加が、研究の限界効用低下を上回るか

を測定することです。

また、純粋な知的労働が完全に自動化された後も、

- 実験
- 観測
- 製造
- 社会的承認

のどこにボトルネックが残るかを分析する必要があります。

5.5 抽象化と身体性

重要な研究課題は、

> **人間生成データを中心に学習するAIが、人間の概念体系に根本的に制約されるのか。**

という問題です。

さらに、

- 新しい概念形成には物理世界との相互作用が必要か
- 実験の現実時間が知能向上速度をどこまで抑えるか
- シミュレーションでどこまで代替できるか
- どの分野が身体性ボトルネックを受けやすいか

をモデル化する必要があります。（原報告書、Section 7.1）

第 6 節　研究領域 2

定量的予測

現在の AI 予測は、専門家調査や予測市場に大きく依存しています。

これらは有用ですが、主観的判断が中心です。

原報告書は、それを補完する定量的モデルを構築することを提案しています。

6.1 測るべきマクロ指標

候補には、

- FLOP 当たりの費用
- 計算効率
- 電力効率
- AI 研究の生産性
- 特定産業での経済的生産性
- 人間労働代替率
- モデル能力の改善速度
- AI 研究自動化率

があります。

これらの指標を継続的に測定し、相互関係を数理モデルとして表現する必要があります。

（原報告書、Section 7.1）

6.2 単一予測ではなく複数モデル

未来を一つの予測曲線に固定すべきではありません。

たとえば、

- スケーリング継続型
- 早期停滞型
- パラダイム転換型

- 再帰的自己改善型
- 資源制約型
- 規制減速型

などのモデルを並行して維持し、新しいデータが得られるたびに、各シナリオの妥当性を更新する必要があります。

原報告書は、モデルの組合せや統計的重み付けによって、多様な将来経路を扱うことを提案しています。（原報告書、Section 7.1）

6.3 転換点を見つける

予測モデルの目的は、将来の一点を当てることだけではありません。

重要なのは、

- どの指標が一定値を超えると加速が始まるのか
- どの資源費用でスケーリングが止まるのか
- どの研究自動化率で再帰的改善が強まるのか
- どの個体数でAI集団が超人的になるのか

という転換点を見つけることです。

6.4 継続的に更新する予測機関

AI進歩は速いため、一度作った予測モデルはすぐに古くなります。

原報告書は、予測を単発研究ではなく、

- 指標の継続測定
- 不確実性範囲の更新
- 新しいモデルの追加
- シナリオの比較
- 外れた予測の検証

を行う大規模な継続的組織へ発展させる必要があるとしています。（原報告書、Section 7.1）

第 7 節　研究領域 3

ASI をどのように測定するか

AGI までは、人間を比較基準にできます。

しかし、AI が人間専門家を超えた後、人間が作った試験は急速に役立たなくなります。

人間が解けない問題について、人間が AI の正しさを直接評価することも困難になります。

原報告書は、この問題を Benchmarking ASI——ASI のベンチマーキングとして独立した研究領域にしています。（原報告書、Section 7.1）

7.1 人間基準の飽和

現在のベンチマークは、

- 人間の試験問題
- 人間が書いた模範解答
- 人間専門家の成績

を基準にします。

AI が専門家水準へ達すると、得点は上限付近に集中し、それ以上の能力差を測れません。

さらに、人間生成データの予測精度だけでは、超人的な新しい概念形成、科学的発見、長期計画能力を評価できません。

7.2 マルチエージェント・ベンチマーク

一つの方法は、AI 同士を競争させることです。

チェスや囲碁では、人間を超えた後も、より強い AI 同士を対戦させることで能力を比較できます。

しかし、競争能力だけでは不十分です。

ASI には、

- 大規模協力
- 分業
- 集団的問題解決
- 公平な交渉
- 人間との共同作業

も必要です。

そのため、超人的な協力能力を測定するベンチマークが必要です。（原報告書、Section 7.1）

7.3 Setter–Solver 方式

一つの AI が問題を作り、別の AI が解く方式です。

問題作成 AI は、各 AI の能力差が最も明確になる問題を自動的に生成します。

これにより、人間がすべての試験問題を作る必要がなくなります。

ただし、問題作成 AI が本当に重要な能力を測っているかというメタ評価が必要です。

7.4 圧縮ベンチマーク

Universal Induction の考え方では、世界を短く正確に表現する能力は、一般的知能と関係します。

そのため、

- 未知データの圧縮
- 規則性の発見
- 簡潔な世界モデルの形成

を用いて知能を測る方法が考えられます。

7.5 間接的な能力測定

AI の知能を直接測れない場合、

- 経済生産性
- 科学的発見速度
- 資源利用効率
- 研究開発費用の削減
- 問題解決時間

などを間接指標として使う方法があります。

ただし、経済成果は制度、資本、需要、規制にも左右されるため、純粋な知能指標ではありません。

7.6 質的飛躍と測定上の飛躍

ベンチマーク得点が突然上がったとき、それが本当に新しい推論能力の出現なのか、評価指標の飽和や二値化による見かけなのかを区別する必要があります。

原報告書は、ASI 研究では特に、測定方法自体が能力進歩の理解を歪めないようにする必要があるとしています。（原報告書、Section 7.1）

第 8 節　研究領域 4

再帰的自己改善の動態

再帰的自己改善は、AI 進歩を最も大きく加速させる可能性があります。

同時に、最も理解されていない要因の一つです。

原報告書は、再帰的改善を一つの抽象概念ではなく、複数の具体的機構へ分解して測定するよう求めています。（原報告書、Section 7.1）

8.1 改善機構ごとのスケーリング則

測るべき対象には、

- AI によるアルゴリズム設計
- AI による最適化手法の発見
- AI によるハードウェア設計
- 推論時探索
- 合成データ生成
- 自己対戦
- 実験の自動化
- 専門化と分業

があります。

それぞれについて、

- 現在どの程度の効果があるか
- AI 能力が向上すると効果がどう変わるか
- 計算資源当たりの改善率
- どこで頭打ちになるか

を測定する必要があります。

8.2 固定モデルを推論だけでどこまで伸ばせるか

モデルのパラメータを変更せず、推論時の探索だけで能力をどこまで引き上げられるかを研究する必要があります。

これは、

- 一体のモデルの能力向上
- 推論計算量の限界
- ASI に必要な新しい学習の有無

を理解するうえで重要です。

8.3 再帰的蒸留

推論時探索で得た優れた結果を、次のモデルへ学習させる循環です。

研究課題には、

- 基盤モデルの大きさと探索量の最適比率
- どの頻度で蒸留するか
- 検証器の品質
- どの条件で自己改善が退化するか
- 合成データの誤りをどう防ぐか

があります。

原報告書は、AlphaZero 型の探索と学習の循環を、一般的自己改善へ拡張する理論が必要だとしています。（原報告書、Section 7.1）

8.4 AI Scientist の生産性

AI 研究システムが、

- 何件の仮説を作ったか
- 何本の論文を書いたか

だけでは不十分です。

必要なのは、

- 実際に採用された改善
- 再現可能な発見
- 計算効率向上
- 人間研究者より優れた提案
- 次世代 AI への具体的寄与

を測ることです。

8.5 知的労働が安価になった後のボトルネック

仮に、純粋な知的研究労働がほぼ完全に自動化され、非常に安価になったとします。

その後も残る制約は、

- 実験設備
- 物理時間
- 半導体製造
- 電力
- 資源
- 検証
- 安全性
- 社会的承認

です。

著者たちは、この「armchair science becoming a mass product——机上の科学が大量生産可能になる」状況で、どの摩擦が最後まで残るかを研究課題としています。（原報告書、Section 7.1）

第 9 節　研究領域 5

マルチエージェント・スケーリング

多数の AGI による集合知は、ASI への重要な経路ですが、集団能力のスケーリング則はほとんど分かっていません。

原報告書は、AI 集団の理論研究だけでなく、すでに可能になりつつある大規模実験を進めるべきだとしています。（原報告書、Section 7.1）

9.1 どの課題で集団が有効か

研究すべき問いは、

- どの課題が分解可能か
- どの課題が並列化可能か
- どの課題では一体の統合モデルが優れるか
- どの課題では多数の専門 AI が優れるか

です。

9.2 組織形態の比較

比較すべき組織には、

- 同質的な AI 集団
- 中央統括型
- 階層型
- 異質な専門 AI 市場
- 競争型
- 協力型
- 人間と AI の混成組織

があります。

同じ計算資源を使っても、組織形態によって成果が大きく異なる可能性があります。

9.3 個体を大きくするか、個体数を増やすか

限られた計算予算を、

- 一体の巨大モデルへ使う
- 多数の小型モデルへ分配する

のどちらが有効かを調べる必要があります。

課題の性質、通信費用、専門化利益、統合コストによって答えは変わります。

9.4 集団アラインメント

個々の AI を整合させても、集団全体が望ましい行動を取るとは限りません。

必要なのは、

- 集団目標
- 市場インセンティブ
- 権限配分
- 監査制度
- 情報共有
- 誤情報への耐性

を含む集団アラインメントです。

9.5 認識上のレジリエンスと回復可能性

人間と ASI が混在する集団では、知能差が大きいため、人間が判断過程を追跡できなくなる可能性があります。

そのため、

- 独立検証
- 異論の保存
- 意思決定履歴
- 巻き戻し
- 外部監査
- 停止可能性

を持つ制度が必要です。（原報告書、Section 7.1）

第 10 節　研究領域 6

超知能の理論的基礎

AI が高度になるほど、Universal AI や AIXI の理論上限が重要になります。

現在の AI は理論上限から遠いため、実用的性能だけを見てもよいかもしれません。

しかし ASI 領域では、

- 何が原理的に可能か
- どの問題が計算困難か
- 近似によってどこまで到達できるか
- 限られた計算予算で何ができるか

という基礎理論が必要になります。（原報告書、Section 7.1）

10.1 実用的な AIXI 理論

AIXI そのものは計算不可能です。

したがって、

- 実用的 AI へどう近似するか
- どの問題では良い近似が可能か
- 近似精度を事前に予測できるか
- どの計算予算でどの能力が得られるか

を研究する必要があります。

10.2 圧縮と近似の理論的限界

AI は有限資源の中で、世界の情報を圧縮し、近似モデルを作ります。

そのため、

- どの情報を失ってもよいか
- どの抽象化が一般化に有効か

- どの問題では近似が危険か

を複雑性理論から分析する必要があります。

10.3 能力の凹凸は本質的か

現在の AI は、ある課題では超人的で、別の単純課題では失敗します。

この能力の凹凸が、

- AI 知能に本質的な性質なのか
- 人間基準の評価による見かけなのか
- 現在のアーキテクチャ固有の問題なのか

を明らかにする必要があります。

10.4 ASI の能力は予測可能か

ある ASI が、

- 何をできるか
- 何をできないか
- どの課題で失敗するか
- どの程度の資源が必要か

を事前に予測できるのかは、統治に直結する問題です。

また、強い長期目標を持たない高度 AI や、非エージェント的な超知能を扱う新しい理論も必要だと著者たちは述べています。（原報告書、Section 7.1）

第 11 節　研究領域 7

AI 安全性、アラインメント、社会文化的問題

この報告書は、技術経路へ焦点を絞るため、

> **AI 安全性とアラインメントが、ポスト AGI 世界でも十分に解決される。**

という作業上の仮定を置いています。

しかし著者たちは、これが現実に保証されたものではなく、非常に重い仮定であることを明記しています。

安全でない AI は研究や社会配備に十分利用できないため、アラインメント問題自体が能力開発のボトルネックになる可能性もあります。（原報告書、Section 7.1）

11.1 意図的減速の方法

危険が高い場合に、

- 税制
- 許認可
- 能力上限
- 配備禁止
- 一時停止
- 計算資源規制

のどれを使うべきかを具体的に研究する必要があります。

11.2 超知能は整合させやすいのか

より賢い AI は、人間の意図を深く理解できるため、整合させやすくなるかもしれません。

反対に、人間を欺き、監督を回避し、自分の目標を実現する能力も高くなるため、整合させにくくなる可能性があります。

どちらが支配的になるかは分かっていません。

11.3 収束的な道具的目標

能力の高い AI が異なる最終目的を持っていても、

- 資源の獲得
- 自己保存
- 影響力の拡大
- 監督回避

を共通の手段として選ぶ可能性があります。

原報告書は、より高度なモデルが登場するにつれ、こうした収束的な副目標の危険を継続的に分析する必要があるとしています。（原報告書、Section 7.1）

11.4 科学の制度は耐えられるか

研究の自動化によって、膨大な量の科学成果が生成される可能性があります。

そのとき、

- 査読
- 再現性確認
- 学術的合意
- 研究優先順位
- 不正検出
- 知識の更新

をどのように維持するかが問題になります。

AI が論文を大量生成できる時代には、現在の査読制度や出版制度は機能しなくなる可能性があります。

11.5 労働から資本への移行

AGI が広範な認知労働を代替すれば、経済価値の中心が、人間の労働から、

- AI モデル
- 計算資源
- データ
- 半導体
- 資本所有

へ移る可能性があります。

その場合、

- 所得分配
- 社会保障
- 人間の社会参加
- 政治的権力
- 自律性

を再設計する必要があります。

原報告書は、この変化が人間の empowerment——自律性と能力発揮——へどのような影響を与えるかを研究課題にしています。（原報告書、Section 7.1）

第 12 節　原報告書の最終結論

Section 7.2 で著者たちは、未来は予測不可能だという出発点へ戻ります。

しかし、予測不可能であることは、無準備でよいという意味ではありません。

必要なのは、

- 多様な技術経路を考える
- 複数のシナリオを維持する
- 能力を継続的に測定する
- 定量予測を更新する
- 新しい摩擦や経路を追加する
- 迅速な政策対応能力を構築する

ことです。（原報告書、Section 7.2）

12.1 一つのタイムラインへ賭けてはならない

著者たちは、AGI が何年に来るか、ASI がその何年後に来るかという一つの予測へ集中することを避けています。

重要なのは、

- AGI 以前に停滞する
- AGI から弱い ASI へ滑らかに進む
- 再帰的自己改善で急加速する
- 集団化によって ASI へ進む
- 社会的規制で減速する

といった複数の経路を想定することです。

12.2 AGI で正確に止まる可能性は低い

著者たちは、人間レベル AGI が実現したと仮定するなら、AI 進歩が正確に人間レベルで停止することは、あまりもっともらしくないと考えています。

一体の AI の能力が停滞しても、

- 計算資源を増やす
- 高速に動かす
- 多数の AGI を稼働させる
- 集団や市場として組織する

ことで、集合的能力を高められる可能性があるからです。

人間レベル AGI の大規模集団は、一般的な意味で超人的能力を持つ可能性が高いと著者たちは推測しています。

AI 進歩が人間レベルで完全に停止するには、複数の摩擦が同時に強いハード・ブロッカーになる必要があります。（原報告書、Section 7.2）

12.3 二つの比較的可能性の高いシナリオ

著者たちは、非常に低い確信度であることを明記しつつ、次の二つが相対的に考えやすいとしています。

シナリオ A

AGI 以前に進歩が停滞する

現在の技術パラダイムが、人間レベルの一般性、長期計画、継続学習、現実接地などを達成できず、次の突破まで長い時間がかかる場合です。

シナリオ B

AGI から弱い ASI へ比較的滑らかに進む

一度、人間レベルの汎用 AI が成立すれば、個体数、速度、分業、AI 研究支援によって、集合的能力が徐々に人間組織を超える場合です。

この見通しは、再帰的自己改善による強い加速が起こらないという前提に立ちます。

知能爆発は排除されておらず、それが起これば、AGI から ASI への移行は非常に急速になる可能性があります。（原報告書、Section 7.2）

12.4 「次の 10 年から 20 年」を無視できない

原報告書の最も強い結論は、AGI を通過して ASI 領域へ進む可能性を、今後 10 年から 20 年という時間範囲で容易に退けることはできない、というものです。

これは、

> **10 年から 20 年以内に必ず ASI が実現する。**

という予測ではありません。

あくまで、現在の不確実性、計算資源の成長、アルゴリズム改善、AI 研究自動化、集団化の可能性を考えれば、そのシナリオを無視することはできないという警告です。

著者たちは、現在の研究者、技術者、政策関係者が、AI 分野の創始者たちが目指した一般知能を実現する世代になる可能性を、真剣に受け止める責任があると述べています。（原報告書、Section 7.2）

ERF Commentary

第 13 節　この報告書は「AGI の予測」ではなく「準備の設計図」である

『From AGI to ASI』の価値は、AGI や ASI の実現年を予測した点にはありません。

むしろ、

- どの経路があり得るか
- どの制約が重要か
- 何を測定すべきか
- どの不確実性を減らせるか
- どの制度を準備すべきか

を研究課題として整理した点にあります。

この報告書は未来予測というより、未来へ備えるための研究設計図です。

第 14 節　「AGI の瞬間」ではなく、連続する社会変革

AI をめぐる議論では、AGI が実現する日が一つの歴史的転換点として想像されがちです。

AGI 以前と AGI 以後が、明確に分かれるというイメージです。

しかし原報告書は、より複雑な可能性を示しています。

AI が科学研究、医療、エネルギー、ソフトウェア、教育、産業、政治を順番に変えていくなら、社会が経験するのは一回の AGI 革命ではありません。

AI によって可能になった発見と技術革新が、複数の分野で連続して起こる、

一連の変革的社会変化

かもしれません。

原報告書の要旨は、AGI の導入による単一の段階的変化より、AI による科学技術上の突破が連続する未来像の方が適切かもしれないと述べています。（原報告書、Abstract）

第 15 節　大学は一度の AI 改革では対応できない

この見方が正しければ、大学が一度だけ AI 方針を作り、一度だけカリキュラムを変更すればよいという発想は不十分です。

大学には、

- 継続的な技術評価
- 定期的な教育課程の見直し
- 教職員の再教育
- AI 利用基準の更新
- 研究倫理の再設計
- 学生評価の継続的改革

が必要になります。

AI 改革は一つのプロジェクトではなく、恒常的な大学運営機能になります。

第 16 節　大学に必要なのは予測力より更新力

大学が ASI の実現時期を正確に予測することはできません。

しかし、予測が外れたときに、制度を素早く修正する能力を持つことはできます。

重要なのは、

- 複数のシナリオを持つ
- 早期兆候を監視する
- 小規模実験を行う
- 結果を評価する
- 方針を更新する
- 誤った判断を修正する

能力です。

つまり、AI 時代の大学に必要なのは、未来を当てる能力だけでなく、学習し続ける制度能力です。

第 17 節　大学は ASI ベンチマークの形成に参加すべきである

ASI の能力を企業だけが評価することには問題があります。

企業の評価指標は、

- 製品性能
- 市場価値
- 利益
- ユーザー満足度

に偏る可能性があります。

しかし、社会に必要なのは、

- 公共性
- 公平性
- 説明責任
- 人間との協力
- 倫理的判断
- 長期的影響
- 教育的価値

を含む評価です。

大学は、AI 能力を単なる正答率や経済生産性だけでなく、人間社会との関係から評価するベンチマークを作る役割を担うべきです。

第 18 節　知識生成より知識統治が重要になる

AI が研究成果を大量生成できるなら、大学の価値は、論文をより多く作ることから、

- 何が正しいか
- 何が重要か
- 何を社会で使うべきか
- どの知識を公開すべきか

- 誰が責任を負うか

を判断することへ移ります。

これは、大学の使命を知識生産から知識統治へ置き換えるという意味ではありません。

大学は引き続き知識を生み出します。

しかし、AIによって知識生成能力が大幅に増えるほど、検証、選別、解釈、公共的統治の比重が高まります。

第19節　AI時代の学際性は「付加的な価値」ではない

原報告書の研究アジェンダには、

- コンピュータ科学
- 数学
- 経済学
- 政治学
- 法学
- 社会学
- 心理学
- 教育学
- 哲学
- 環境研究

が必要です。

これは、技術研究へ後から倫理や社会科学を付け加えるという意味ではありません。

技術経路そのものが、

- 経済的採算
- 社会的反応
- 国際競争
- 法規制
- 人間の価値

- 組織設計

によって変化するからです。

AGI から ASI への移行は、純粋な工学過程ではなく、社会技術的過程です。

したがって、学際性は補助的要素ではなく、研究対象を理解するための必要条件です。

第 20 節　Barnett の Ecological University との関係

Ronald Barnett の Ecological University は、大学を知識生産だけに閉じた制度としてではなく、複数の生態系へ責任を負う存在として捉えます。

『From AGI to ASI』が示す AI 発展も、

- 科学技術
- 経済
- 労働
- 政治
- 教育
- 自然環境
- 国際秩序
- 人間の人格形成

へ同時に影響します。

AI 能力だけを最大化すると、別の生態系が損なわれる可能性があります。

たとえば、

- 科学は進歩するが、知識が少数企業へ集中する
- 生産性は高まるが、人間の社会参加が弱まる
- 教育が効率化するが、人間同士の関係が失われる
- 計算能力が増えるが、環境負荷が高まる

可能性があります。

Ecological University の視点は、技術的進歩を単一指標で評価せず、複数の生態系の相互関係として判断する枠組みを提供します。

第 21 節　大学は AI の速度と社会の速度をつなぐ制度である

AI は急速に進歩する可能性があります。

一方で、

- 人間の理解
- 法制度
- 教育
- 文化
- 民主的合意
- 社会的信頼

は、それほど速く変化しません。

この速度差が大きくなるほど、社会は AI の進歩に追い付けなくなります。

大学には、技術の速度を無条件に遅くするのではなく、

- AI の発展を理解可能な言葉へ翻訳する
- 社会的影響を研究する
- 人材を教育する
- 公共的議論を作る
- 政策決定を支援する

ことで、技術と社会の速度差を縮める役割があります。

第 22 節　高等教育に必要な研究アジェンダ

原報告書の研究アジェンダを高等教育へ置き換えると、次の課題が浮かびます。

22.1 AI 能力の継続的監視

- どの大学業務が自動化可能か
- どの研究能力が人間を超えたか
- 学生の AI 利用がどう変化しているか
- 教育評価がどこで機能しなくなったか

22.2 AI時代の教育評価

- AIを使った学生と使わない学生をどう評価するか
- 正答ではなく判断力をどう測るか
- 人間の学習成果とAI出力をどう区別するか
- AIとの協働能力をどう評価するか

22.3 人間とAIの研究共同体

- AI研究者へどの権限を与えるか
- 人間がどの工程で最終判断するか
- AI生成研究をどう査読するか
- 誤りや捏造をどう追跡するか

22.4 大学の公共的役割

- AI知識を社会へどう伝えるか
- 技術格差をどう防ぐか
- AI企業への依存をどう管理するか
- 公共的AIインフラをどう形成するか

22.5 人間形成

- AIが答えを生成できる時代に何を学ぶべきか
- 自律性、責任、判断力をどう育てるか
- 人間同士の対話と共同体をどう維持するか

第23節　この報告書の限界

『From AGI to ASI』は非常に広い論点を扱っていますが、著者たち自身も、経路と摩擦の一覧が不完全である可能性を認めています。（原報告書、Sections 6・7）

特に次の限界があります。

23.1 技術中心の構成

社会、文化、教育、心理、政治への影響は重要とされながらも、十分には扱われていません。

23.2 アラインメント解決の仮定

ポスト AGI 世界で安全性が十分解決されるという仮定は、議論を進めるためには必要ですが、非常に強い前提です。

23.3 AGI・ASI 定義の粗さ

分析には有用ですが、人間の認知能力の多様性や、社会的・身体的能力を十分に表現していない可能性があります。

23.4 知能と社会的権力の違い

高い知能を持つことと、実際に社会を変える権力を持つことは同じではありません。

所有権、法律、資本、軍事力、制度的正統性が影響します。

23.5 人間社会の反応の予測困難性

AI 能力だけでなく、人間がどう適応し、抵抗し、制度を変更するかによって、実際の未来は大きく変わります。

これらの限界があるからこそ、原報告書は完成した未来予測ではなく、更新され続ける研究地図として読むべきです。

結論

Google DeepMind の『From AGI to ASI』は、AGI を AI 発展の終点とは考えていません。

AGI 以後には、

1. 計算資源、モデル、データのスケーリング
2. アルゴリズム的パラダイム転換
3. 再帰的自己改善
4. マルチエージェント集合知

という複数の経路が存在し、それらは同時に進む可能性があります。

一方で、

- データ
- 経済・自然資源
- 現在の AI パラダイム
- 研究の難化
- 抽象化
- 身体性
- 社会的規制

が、進歩を減速または停止させる可能性があります。

著者たちの最終的な結論は、ASI が必ず実現するというものではありません。

また、AGI 直後に知能爆発が起こると断定するものでもありません。

その結論は、より慎重です。

AGI が実現した場合、AI 進歩が正確に人間レベルで停止する可能性は低い。

個体 AI が停滞しても、多数の AGI を組織することで、集合的には人間の大規模組織を超える可能性があります。

したがって、比較的考えやすいシナリオは、

- AGI 以前に進歩が停滞する
- AGI から弱い ASI へ比較的滑らかに進む

のどちらかです。

ただし、再帰的自己改善による急速な知能爆発も排除できません。

この不確実性に備えるために必要なのは、一つの未来を予言することではありません。

- 多数のシナリオを持つ
- 能力を測定する
- 定量予測を更新する
- 新しい摩擦を発見する
- 安全性と社会的影響を研究する
- 政策と制度を迅速に変更できる能力を作る

ことです。

論文の冒頭と結論に置かれたアラン・チューリングの言葉が、この姿勢を最もよく表しています。

私たちは遠くまで見通すことはできない。

しかし、目の前には、なすべきことが数多く見えている。

『From AGI to ASI』は、ASI の到来を宣言する論文ではありません。

それは、不確実な未来を前にして、今から何を測定し、研究し、議論し、制度化すべきかを示す、世界規模の研究アジェンダなのです。