

AIの技術進化



長谷佳明

CONTENTS

- I AIの現在地
- II 汎用技術への進化
- III 実体化するAI
- IV AI進化の時間軸と世界の拡張

要 約

- 1 IPAによれば、AIの導入、または実証実験を行う国内企業の割合は2023年度で34%に達し、年々増加している。生成AIの導入を検討する企業は全体の53.8%に上り、生成AIがAI普及の牽引役となることが期待されている。しかしながら生成AIは要素技術であり、現状では誰もが活用できる技術ではなく、中長期的には抜本的な進化が必要である。
- 2 今後、AIは、より大規模化し電力インフラのような公共性を有する「大きなAI」と、小型化し生物のような多様性を有する「小さなAI」へと進化していくと思われる。小さなAIは、モノに組み込まれ、ユーザーに寄り添い、その人の状況を最もよく知る存在となる。小さなAIで解決できない問題は、より高度な思考力を持つ大きなAIの知的リソースを利用し、推論性能を拡張して解決する。大きなAIと小さなAIとは補完関係となり、将来的に、社会になくてはならない汎用技術になるだろう。
- 3 生成AIブームにより、ヒューマノイドロボットが再び注目され、工場作業などの実用用途に対応する「未来の労働力」としての開発が進んでいる。ロボットの自律的行動には「世界モデル」が必要であり、生成AIの技術が活用され始めている。ロボットがAIのAvatarとして実世界と接触し、経験値を積むことで、世界モデルが完成に近づくと期待される。
- 4 2030年代から2040年代にかけて、小さなAIが自動車や家電、空間などに組み込まれ、知的な空間「Ambient Intelligence（環境知能）」が実現される。小さなAIは、人々と対話し課題解決を支援する一方で、ネットワークを形成し、収集したデータにより実世界をデジタルで再現した「ミラーワールド」を生成する。大きなAIがミラーワールドを通じて現実世界の問題に予見的な対策を立て、デジタルとフィジカル両面から世界が営まれる新たな時代が開かれるだろう。

I AIの現在地

1 AIの最新の活用状況

IPA（独立行政法人情報処理推進機構）が2024年6月に公開した「DX動向2024」によれば、「AIを導入している」「現在実証実験を行っている」と答えた国内企業の割合は、2023年度時点で34%に及び、年々増加している（2022年度では31.3%、2021年度では28.4%）。

生成AIに対象を絞ると、「導入している」と答えた企業の割合は15.6%であるが、「現在、試験利用をしている」「利用に向けて検討を進めている」「これから検討をする予定である」と答えた企業を加えると53.8%となり、高い関心がうかがえる。今後、生成AIが、AIの普及の牽引役となるのは間違いない。

2 従来型AIと生成AIの違い

文字認識のような従来型AIと、ChatGPTのような生成AIの違いを表1に示した。従来型AIは、画像から文字を抽出するなど決められたタスクに特化し、与えられたデータから解を導く正確性に価値がある。一方、生成AIは、タスクは特化せず、さまざまな状況に応じて無数の解を生み出す「思考力」に価値がある（ただし、この「思考力」とは、アウトプットが人間の思考結果と類似していることからそう呼んでおり、2024年11月時点では、AIの推論は人間の思考力とは異なる）。従来型AIと生成AIは機械であり、人間には到底真似できないスタミナを持つ点は共通している。

2024年3月に欧州議会で可決した「AI Act」をはじめ、欧州委員会は、生成AIを

表1 AIと生成AIの違い

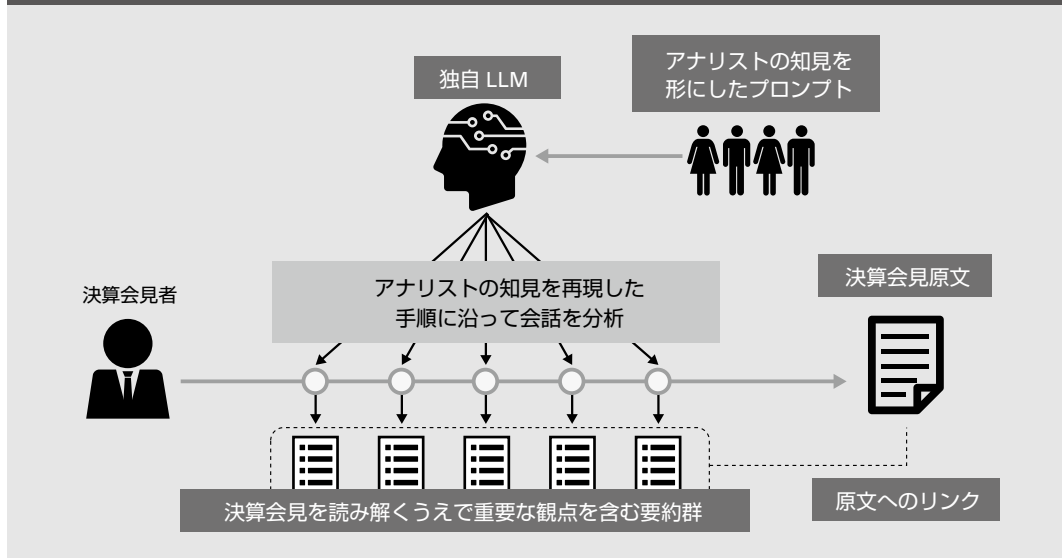
	従来型AI 例：文字認識	生成AI 例：ChatGPT
タスク	特定	特化しない
動作	定められた解を探す	無限の状況に応じて解を考える
価値	正確性 例：何を書いてあるか？	思考力 例：なぜ、円安なのか？
共通点	スタミナ 疲れない、24時間働く	

「Generative AI」ではなく、あえて「汎用目的AI（GPAI：General-purpose Artificial Intelligence）」と表現し始めている。GPAIは、自律的に人並みの知的作業が可能な汎用人工知能（AGI：Artificial General Intelligence）と比べ、知識や思考力も不完全で、自律性も限定的である。しかし、汎用的な目的に適用可能なAIの誕生は、AIが道具の域を脱し、人間のパートナーとなる飛躍的進化に向けた第一歩である。

3 生成AIのアプリケーション化

ブルームバーグは、2023年3月に、金融サービスに特化した500億パラメータの独自の大規模言語モデル「ブルームバーグGPT」を開発するなど、生成AI活用の先駆者である。2024年1月には生成AIを同社の情報端末に組み込み、決算説明会の要約をつくり出す顧客向けのアプリケーションを開発した（図1）。決算説明会の内容をブルームバーグのアナリスト目線で分析し、対象企業をより深く理解する手助けになる。このアプリケーションは、アナリストが気になるポイントを抽出し読み解く作業を自動的に再現でき、ユ

図1 ブルームバーグ 決算説明会向け要約機能



ユーザーの生成AI活用スキルには依存しない。ボタン1つで要約の元になった原文を確認でき、さらに詳細な分析機能と連携するなど、既存システムとの融合にも重点を置く。

生成AIの活用は、対話型生成AIサービスを個人が創意工夫し、生産性を高める「点の生産性向上」から、アプリケーションによって個人のスキルに拠らず、生成AIの効果を等しく享受する「面の生産性向上」に向かっている。ブルームバーグの事例が示すように、企業が生成AIを通じて圧倒的な差別化や競争力の向上を図るためには、生成AIを目的に合わせたアプリケーションにしなければならない。生成AIアプリケーションとは、生成AIによって機能を置き換える性能が向上したり、生成AIを組み込むことで著しく機能が高度化された新しいアプリケーションである。

そして、より一層、生成AIが社会に普及するには、モデルの性能向上に加え、中長期的には「汎用技術」への進化が必要である。

II 汎用技術への進化

1 汎用技術とは

世の中には、社会の隅々にまで普及し、生活を一変するテクノロジーがある。歴史をさかのぼれば、蒸気機関、鉄道、電力、コンピュータ、そして近年ではインターネットなどである。このような技術は「汎用技術 (General-purpose technologies)」と呼ばれ、経済発展や産業構造の変化にも多大な影響を与えてきた。たとえば、鉄道の普及は、移動コストを劇的に下げ、商品や資材の輸送を容易にした。また、都市への人口流入を加速するなど、社会に経済的、構造的変化をもたらした。

汎用技術は、およそ50～100年周期で登場している。もし、インターネットを汎用技術の最新と考えると、その起源である1967年のARPAnet (Advanced Research Projects Agency Network) 誕生から、すでに50年以上が経過している。21世紀に入り、クラウド

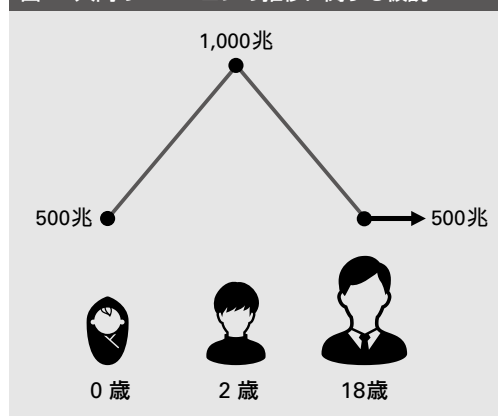
コンピューティングやスマートフォンなどの革新的な技術は存在したが、AIは今後の発展次第ではそれらを越える変化を社会に与える技術になり得る。AIが社会に広く普及し、汎用技術となる可能性を次節以降で検証する。

2 モデルの大規模化というトレンド

言語モデルの性能は、モデルを大きくすればするほど性能が高まる。言語モデルは、人間の言葉を理解し、新たな文章をつくり出す人工知能の技術であり、生成AIの根幹でもある。言語モデルは巨大化によって性能が高まるとの仮説は、2020年1月にオープンAIの研究者らによって公開された論文「Scaling Laws for Neural Language Models」に始まる。この仮説は、2020年5月に公開された同社の大規模言語モデル「GPT-3（Generative Pre-trained Transformer 3）」により証明された。GPT-3は、オープンAIの前モデルGPT-2の15億パラメータと比べ、100倍の1750億という当時想像もしなかったサイズにした結果、数々の性能テストの最高スコアをぬり替え、人間との会話力にみる「創発性」など、多くの研究者にAIの言葉の理解が次の段階に進んだと思わせた。その後、アウトプットを人間の一般的な価値観や意図に寄せる調整（アライメント）や、人間の会話特有のやり取りの学習を経て、2022年11月、ChatGPTとして公開され世界中の人々に衝撃を与えた。

以後、言語モデルの高性能化を巡る競争は、パラメータ数が公開されているモデルだけでもGoogleの開発する「Gemini Ultra」は1.6兆となるなど、資金力のあるスタートアップやベンダーにより、大規模化が続いてい

図2 人間のニューロンの推移に関する仮説



る。人間の脳のニューロンの接続数は、一説には、誕生時に500兆で始まり、2歳頃に1000兆にまで増え、成人を迎える頃には、再び500兆で安定するという^{※1}（図2）。体の成長に合わせて脳は容積を増し、雛形となるネットワークも複雑性を増す。その後、学習を経てネットワークは調整を繰り返し、成人を迎える頃にはシステムとして安定する。

オープンAIは、全人類に利益をもたらす汎用人工知能の開発を目指して設立された企業である。汎用人工知能を実現するAIモデルには、人と同じ複雑性を持つニューラルネットワークを必要とするならば、500兆はモデルの巨大化の行き着く先（ゴール）ともいえる。このまま、モデルが大規模化し、コンピューティングパワーも増大すれば、2030年には500兆の万台に達する「アーティフィシャル・ニューラルネットワーク」が誕生しているかもしれない。

かつて、GPT-3がパラメータ数にして1000億の壁を越えたとき、想定外の創発性を発揮したように、500兆の壁を越えたとき、誕生するものは果たして汎用人工知能なのか、それとも、AIの研究において、たびた

び話題に上がる、意識や心の芽生えなのか。現時点では500兆のモデルは存在せず、その登場を待たねばならないが、今後もモデルの巨大化は続くだろう。

3 特化・小型モデルの実用性向上

大規模化による汎用人工知能誕生を目指す開発が進む一方で、足元では実用性を高める取り組みも始まっている。たとえば、Microsoftが2024年4月に公開した小規模モデル「Phi-3 シリーズ」である。Microsoftの開発者らによる論文「Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone」によれば、Phi-3-smallは、わずか70億パラメータながら、評価対象の問題によっては大規模言語モデルのGPT-3.5を上回る推論性能を持つ。

Phi-3は、インターネットで収集されたデータを学習する代わりに、シンセティックデータ（人工的に作り出されたデータ：合成データ）を活用している。具体的には、名詞、動詞、形容詞を同数含む3000語のデータセットを既存の大規模言語モデルによって作成。そのデータセットから再び大規模言語モデルに名詞、動詞、形容詞を1つずつ選ばせ、短い童話を数百万件作成させた。この結果、言葉の偏りの少ない高品質な学習データができ上がり、Phi-3は小型ながら高性能となった。小規模モデルは、スマートフォンのようなリソースの限られた環境でも動作するように設計され、今後、電化製品などのさまざまなデバイスやセンサーに組み込まれ、われわれの身の回りの空間を知的なものへと変えていく可能性を持つ。

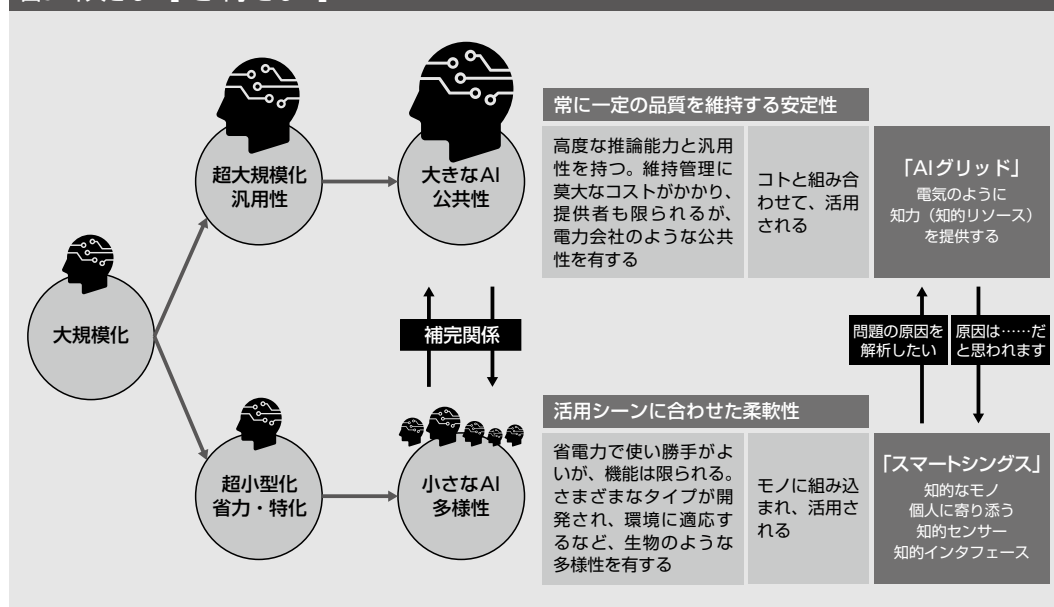
4 大きなAIと小さなAI

大規模モデルはさらに成長し、ほかに類を見ない高度な推論能力と汎用性を持つようになり、社会にとって重要なインフラとなる。それと同時に、維持管理に莫大なコストがかかり、提供者も限られるため、いわば電力会社のような公共性を有するものになる。一方で小規模モデルは、現場に合わせてさまざまなタイプが開発され、生物界のように多様化すると思われる。両者を対比し、「大きなAI」と「小さなAI」と名づける（図3）。

小さなAIは、活用シーンに合わせた柔軟性がメリットであり、「モノに組み込まれ」、ユーザーに寄り添い、その人の状況やその環境を最もよく知る存在となる。画像認識の目、音声認識の耳、そしてセンサーから収集されるデータを小さなAIは分析し、周辺状況を整理し、簡易な問題にはその場で答える。小さなAIで解決できない問題は、大きなAIの知的リソースを利用し、推論能力を拡張する。大きなAIは、高い推論能力と一定の品質を維持した安定性がメリットであり、解決したい「コトと組み合わせで」活用される。現在のプロンプトエンジニアリングのような推論に必要な前処理や、データ入力には小さなAIが担うことで、生成AIは格段に使いやすくなる。一度、大規模モデルと小規模モデルとして別の進化の道を進むことになったAIは、大きなAIと小さなAIとなって再び出合う。大きなAIと小さなAIは、互いを補完する関係になる。

小さなAIによって身の回りのサービスやデバイスらが知的に進化し、大きなAIはそれらを支えるインフラになる。実現すれば、電力ならぬ知力（知的リソース）を提供する

図3 「大きなAI」と「小さなAI」



新たな汎用技術の誕生といえる。

5 萌芽事例①日立産機システム： 自らの不具合を説明する 産業用機械

産業用電気機械メーカーである日立産機システムは、2024年2月、不具合発生時に機械が自ら話して問題解決を促す「設備コンシェルジュ」を発表した。設備コンシェルジュは、音声合成や生成AI、データ分析技術などを組み合わせたサービスであり、同社の提供する機器の遠隔監視サービス「FitLiveサービス」を拡張したものである。設備コンシェルジュは、不具合を捉えると各種マニュアルや過去事例を学習した生成AIのモデルに問い合わせ、端末に搭載されたAIによる音声合成により、その結果を分かりやすく作業員に伝える仕組みになっている。

エッジに搭載された音声合成AIと、高度な知識を持つ生成AIが連携し、課題を解決

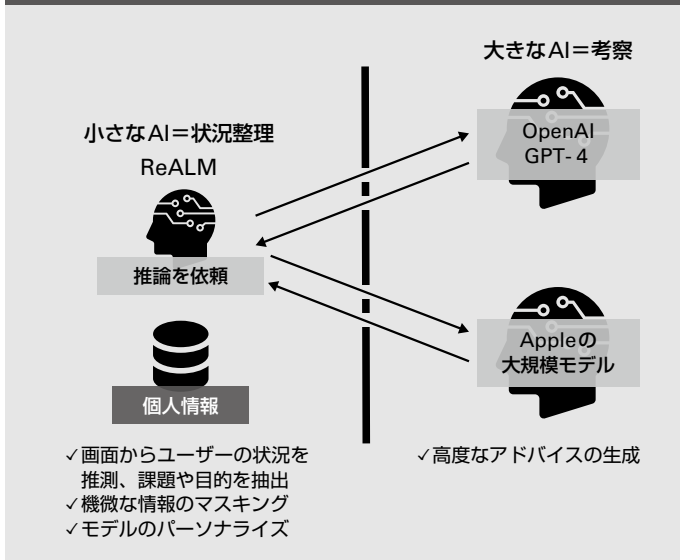
する「大きなAI」と「小さなAI」の萌芽事例といえる。

6 萌芽事例②Apple： Apple Intelligence

Appleは2024年9月、生成AIサービス「Apple Intelligence」に対応したスマートフォン「iPhone16」などを発表した。Apple Intelligenceは、スマートフォンに組み込まれた小規模モデルとAppleの専用サーバで稼働する大規模モデルか、もしくはオープンAIの開発する大規模モデルとが用途に合わせて連携する仕組みになっている。「大きなAI」と「小さなAI」を彷彿とさせるアーキテクチャである。

Appleが考えるAIの役割分担は、2024年6月に公開された同社の論文「ReALM: Reference Resolution As Language Modeling」から読み取れる（図4）。スマートフォンに組み込まれる小規模モデルは、ユーザーの意

図4 「Apple Intelligence」の将来の予想図



図の抽出や関連するデータの収集に特化している。個人情報のような機微な情報は、大規模モデルに連携する際には事前にマスキングする。簡易な問題であれば小規模モデルで解決し、より高度なモデルの支援が必要な際には目的とデータを添え、ユーザーに代わりプロンプトを機械的に生成し、外部の大規模モデルによる推論を依頼する。つまり、小規模モデルがあることで、ユーザーは現在のように煩雑なプロンプトを用意する必要はない。また、モデルが小型である点からも学習が容易で、使えば使うほどユーザーの意図を先読みし、利便性は向上すると予想される。

7 知力のベストミックス

AIが汎用技術を目指す際の課題は、汎用的で公共性のある大きなAIを誰が維持管理するかである。高い推論能力を持つ知のインフラは、サービス停止時の社会的影響が大きい。たとえば、同じく汎用技術である電力は、太陽光や風力、水力、火力、そして原子

力などがあるように、さまざまな電源をミックスして、安定供給を実現している。AIの提供する知力も、汎用技術を見据えた場合、過度の海外依存や一部の企業による市場の独占はリスクとなる。このため、大きなAIには、マルチベンダーによる性能や性質、コストの異なるサービスを組み合わせた「知力のベストミックス」が求められるだろう。

8 AIの進化を阻む主な技術的課題

国際エネルギー機関（IEA）が2024年1月に公表したレポート「Electricity 2024」によると、世界のデータセンターの消費電力は2022年の推定460TWhから、2026年には1000TWhへと倍増すると予想されている。新たな需要の多くはAIが占めると考えられており、電力消費量の多さは、AIが今後、利用者が拡大し、普及するうえで大きな足かせとなる。

データセンターなどで活用されるAI用の半導体は、米国のNVIDIA製がデファクトスタンダードになっている。2024年第4四半期から出荷が本格化する「Blackwell」は、前世代の「Hopper」と比べ、処理能力が最大30倍高速になり、対電力消費能力の観点では効率化されるものの、1チップ当たりの電力消費量は約1.7倍になると予測されており、AIの需要に従い、今後も電力消費は増加する見込みである。AIが汎用技術となるためには、AIに必要な電力を安定的に供給する電源の確保と送配電網の整備が欠かせない。

ワシントン大学やカーネギーメロン大学の研究者が2023年5月に公開した論文「From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Mod-

els」によると、GPT-4をはじめとした大規模言語モデルには主義主張の違いがある。限られたテストケースによる簡易な解析ながら、オープンAIのGPT-4は「左派自由主義」、メタ・プラットフォームズのLLaMAは「右派権威主義」の傾向が見られた。大規模言語モデルは、言葉を学習する過程で大量の知識を獲得している。学習データは人間の教育における「教科書」といえ、教科書次第で学習した結果も変わる。このため、大規模言語モデルは、思考に対する「偏り（バイアス）」から逃れることはできない。人も、それぞれ考え方が異なるように、万人にとって完璧な偏りのないモデルを構築するのは不可能である。AIの偏りに対しては、人が複数のAIを使い分け、解決するのが現実的である。これは、主義の異なるメディアが存在し、私たちがそれを使い分けしているのと同じである。このため、AI技術は政府や企業による独占ではなく、さまざまな選択肢のあることが望ましいだろう。

そしてAIの進化には、データも欠かせない。AIの学習データに関する調査研究を行うボランティア団体Data Provenance Initiativeが、2024年7月に公開した論文「Consent in Crisis : The Rapid Decline of the AI Data Commons」によると、過去1年間で、AIの学習に有効なデータのうち約25%が利用できなくなった。インターネットのデータは増加の一途をたどるかに見えたが、実際には利用可能なデータは減少している。減少の要因は、学習データのソースとなってきたメディアによるポリシー変更である。この背景には、AI企業によるタダ乗りへの不信感がある。AIの持続的発展には、データを生産す

るメディア企業とデータを消費するAI開発企業との間での利用料に関する合意が不可欠である。

AIを汎用技術として社会に実装するためには研究機関によるAIのモデルの改良に加え、電力などのインフラの整備、モデルの独占の防止、そして学習データの安定的供給が欠かせないだろう。

Ⅲ 実体化するAI

1 ヒューマノイドロボットが 新たなブームの兆し

生成AIのブームの陰で、ロボットにも新たなトレンドが見え始めている。それがヒューマノイドロボット（人型ロボット）である。ヒューマノイドロボットとして思い出されるのは、2000年代に活躍した本田技研工業の「ASIMO」である。当時の技術の粋を集め、二足歩行を可能とし、まさに先端技術の象徴であった。しかし、現代のヒューマノイドロボットは工場での作業など実用性を兼ね備えた「未来の労働力」として開発が進んでいる。

2024年1月、米国のスタートアップFigure社は、ドイツの自動車メーカーBMW傘下のBMWマニュファクチャリングとヒューマノイドロボットの導入で商業契約を結んだ。プレスリリースによると、第一段階として自動車生産にロボットを適用するユースケースを探索し、第二段階として米国のサウスカロライナ州スパークンバーグにあるBMWの工場に導入するという。

ヒューマノイドロボットの導入は、メルセデスにもあった。2024年3月、米国のヒュー

マノイドロボットを開発するスタートアップ Apptronik社とメルセデスは商業契約を結んだ。Apptronik社のプレスリリースによると、高度なスキルは要求されないものの、肉体の負担の大きい作業をロボットにより自動化するとある。たとえば、自動車組立で用いられる大量の金属部品を収納したボックスのような、非常に重いものを持ち上げて運ぶ場面での活用などが推測される。Apptronik社の開発するヒューマノイドロボット「アポロ」は、NASAの火星探査用ヒューマノイドロボット「Valkyrie」の開発で培った技術を商用化したものである。アポロのスペックシートによると、55ポンド（＝約25kg）もの重量物を運搬できる。

現在は、将来の大規模導入を見据えた実用的なユースケースの探索段階にある。かつて1台数億円以上したヒューマノイドロボットは、急激に価格が下がり始めている。たとえば、中国のスタートアップUnitree社は2024年5月、ヒューマノイドロボット「Unitree G1」を1万6000ドル（1ドル＝150円換算で240万円）で販売すると発表した。価格以外にも、油圧駆動並みの動力を発揮する電動のアクチュエータによる軽量化、バッテリーの小型・大容量化によるロボットのポータブル化が進んだ。この結果、ロボットの課題であった空間的制約が緩和され、活動時間も長くなり、実用性を増している。汎用的な労働力としてヒューマノイドロボットが検討され始めたことは大きな進展といえるだろう。

2 ロボットとAI

飲食店で導入が進む配膳ロボットをはじめ、注目を集めるヒューマノイドロボットも

AIと深いかわりがある。配膳ロボットには、自動運転の開発で培われた衝突回避のための画像認識技術や、現在の位置を予測する自己位置推定技術などが活用されている。これにより、衝突を滑らかに回避し目的地まで移動する、1980年代のSF映画に登場するようなロボットの開発が可能となった。これも、Googleのようなテクノロジー企業をはじめ、世界中の自動車メーカーが次世代の重要技術として莫大な予算を投じて自動運転の研究を進めた結果、関連技術が急速にコモディティ化し、安価に調達できるようになったことが寄与している。

ヒューマノイドロボットの要である二足歩行にも最新のAI技術が活用されている。四足より重心が不安定な二足歩行の実現に、深層強化学習を用いた制御技術が活用されている。深層強化学習は、試行錯誤から目的に到達する最適なパスやルールを獲得する技術であり、2016年に当時の囲碁の世界チャンピオンであるイ・セドル氏を破ったAI「AlphaGo」に用いられた技術としても有名である。囲碁のようなゲームは「相手に勝つ」、二足歩行は「倒れない」というように、成功と失敗の条件を設定し、無数の選択肢から最終的によい結果になるパラメータを選べるようAIが学習する。

ロボットの安全な移動が実現すると、次の課題は行動計画の作成である。配膳ロボットであれば、地図情報から最短経路をアルゴリズムによって導き、移動する。衝突の回避など状況に応じた動作はできるものの、事前にプログラミングされた動きであり、自律的ではない。現状は、決められたアクションをプログラムの指示に応じて選択しているにすぎ

ない。ロボットが人間のような汎用的な労働力となるには、状況に応じた行動の計画の立案が不可欠である。

3 世界モデル

ロボットの高度化には、周囲の環境を理解し、自らの行動を予測し、生成するためのフレームワークとして「世界モデル」が必要である。世界モデルの獲得は、人間の持つ物理的、概念的な環境認識能力を人工的に作り出す技術といえる。AIはセンサーから得た情報を元に、外部環境を内部に再現し、行動をシミュレートし、その中から最適な戦略を選択する。

キャッチボールを例として考えると、世界モデルがあれば、ボールの種類や重さ、位置関係などからその軌道を予測し、先回りしてボールを捉えることができる。これには、画像の中の丸いものは何か、どのような特性を持つのか、上昇しているとはどういう状態なのかなど、物体認識や状況理解、基本的な物理法則に関する理論が世界モデルに備わっていないければならない。世界モデルは、長年研究されてきたものの、エキスパートシステムと同様に、大量の知識をいくらデータ化しても数多くの一般常識やルールをモデル化するのは難しく、実用化にはまだ相当な時間がかかると思われていた。しかし、世界モデルにも生成AIは有効である兆しが見え始めている。

文章生成AIに「あなたは、私に仕えるロボットです。喉が渴いたので、飲み物を持ってきてください。といわれたらどう行動しますか？」などと入力すると、「まず、冷蔵庫を探します……」と、次を取るべき行動を生成できる。インターネットにある大量の情報

から、無作為に学習したにすぎない大規模言語モデルが、不完全ながらも一般常識や概念のようなものを獲得し、与えられた条件に応じ自律的に行動を導く点は、画期的であった。

しかし、世界モデルの獲得以外に、壁はもう1つある。言葉で行動を表現できたとしても、言葉の世界で示したものがカメラやセンサーで捉えた実世界の何であるか、どのような動作なのかをマッピングできなければ、正しい行動にはならない。この問題は「シンボルグラウンディング問題」と呼ばれ、カナダの認知科学者であるスティーブ・ハルナッドによって、1990年に論文「The Symbol Grounding Problem」の中で提唱された。生成AIは、基本的に言葉の関係性を学習し、言葉の世界に生きているため、言葉と実世界とを結びつけるのは容易ではない。たとえば、ロボットの5m先、右前方に冷蔵庫があったとする。画像認識によって「冷蔵庫、距離5m」と文字情報で表現できても、冷蔵庫の形状や取っ手の位置など、ドアを開けるための有効な情報が失われてしまう。また、状況に応じて画像の中の何に注目すべきかを人が事前にプログラミングすると、ロボットの自律的な判断が限られてしまう。このため、世界モデルには、文字だけでなく、音声や画像、動画などから何に注目すべきであるかが抽出可能なマルチモーダルモデルでなければならない。大規模言語モデルは人間の言葉の世界を学んだように、マルチモーダルモデルは実世界を学ぶ。

事実、マルチモーダルな世界モデルの開発も加速している。オープンAIは、2024年2月、世界モデルの「Sora」を発表した。テキ

ストや画像、動画などを入力として受けつけるマルチモーダルモデルである。Soraは、テキスト操作により動画を編集するVideo to Video、そしてマイクラフトのような人工世界を忠実に再現するシミュレーション機能を持つ。生成物のインパクトから、Soraにはフォトリアルな架空の動画をつくり出すAIとの印象があるが、より興味深いのはシミュレーターとしての機能である。Soraは、動画の中の物体の特性や状況をある程度解析できるため、破綻なく動画の編集ができる。赤い自動車が走行する動画を入力し、「直線の道路を曲がりくねった道に変更して」と指示すると、自動車もその指示に合わせてコーナーを曲がるシーンに変わる。つまり、Soraは物理法則にのっとった「一貫性」のある動画を生成する。Soraの生み出すフォトリアルな動画は、いわば、Soraの頭の中の世界モデルを人間が観測できるよう可視化した「メディア」とも取れる。実世界への理解がさらに深まれば、Soraは世界の法則を再現し、シミュレーションするシステムへと進化していくだろう。

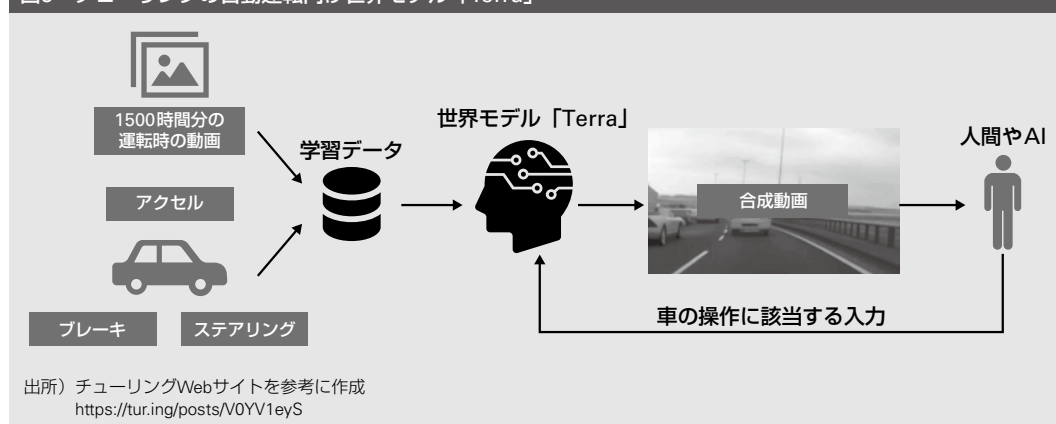
日本のスタートアップであるチューリングは、自動運转向け世界モデル「Terra」を開発した（図5）。1500時間分の走行動画を学

習したモデルで、ハンドル操作やアクセル、ブレーキに応じた運転動画がリアルタイムで生成される。Terraは、自動運転用AIの学習に用いたり、人間の運転技能を高めるドライブシミュレーターとして利用できる。

画像認識用データセットImageNetへの貢献で有名なスタンフォード大学のフェイフェイ・リー教授は、2024年9月、次世代の世界モデルを開発するスタートアップWorld Labs社を設立した。リー教授は、人間は言葉を理解し、言語を通じて他者とコミュニケーションを取る言語知能に加えて、世界を理解し、対話する「空間知能（Spatial Intelligence）」を持つと指摘する。そして、空間知能が生み出す「心の目」を通じて人間は推論し、行動し、発明すると続く。World Labs社はAIや人間が世界を認識し、世界と対話するためのSpatial Intelligenceとして「大規模世界モデル（Large World Models）」の開発を目指している。

オープンAIやWorld Labs社をはじめ、数多くのスタートアップや研究機関が生成AIを活用した次世代の世界モデルの開発を始めている。現状は実験的で不完全なモデルであるが、AIとロボットを融合し、実世界を行動するために欠けていたパーツがそろい始め

図5 チューリングの自動運转向け世界モデル「Terra」



ている。

4 実世界上の 身体的体験学習データ

世界モデル技術は、生成AIによって一歩前進したが、生成AIには足りない学習データがある。それは、人間が成長の過程で獲得する「身体的体験」に相当するデータである。たとえば、私たちはネジを回す際には、はじめは緩く簡単に回るが、次第にきつくなり、締める力を調整しなくてはならないこと、うまく調整しないとネジ山を壊すことを体験から学ぶ。このように、ネジ回し一つとっても、行動はできても、実際に有効なスキルの獲得には手の感覚と経験がものをいう。人は予測と結果のフィードバックを繰り返し、自らのモデルと実世界とのズレを調整している。

AIにとって有望な身体的体験となるのが、ヒューマノイドロボットを通じたデータであろう。大規模言語モデルは人間の考えや経験を文章から学習しているため、それを実世界に応用するには、ロボットの躯体も人型であるヒューマノイドがよいと思われる。ロボットを通じて実世界とのギャップを認識し、補正を繰り返すことで、世界モデルは改善され、実用域に達するだろう。

将来的に工場の生産現場で数万台規模のロボットが働くようになれば、モノを運んだり、将来的にはボルトを締めたり、生産設備を操作する中から得られるデータは、インターネットにはない貴重な経験（データ）になる。また、ロボットの躯体の標準化も重要だろう。AIは同じ躯体の別のAIの経験を活かし、共進化できるからである。

メルセデスやBMWの取り組みからは、FigureやアポロのようなロボットをAIがフィジカルの世界とコンタクトするための「アバター」とし、工場という現場を通じて実世界のデータを収集し、世界モデルを完成させる狙いが思い浮かぶ。大量の労働力とさまざまな作業が求められる自動車工場の生産現場は、全くオープンな環境と比べ、ある程度人工的に場を調整できるなど、ロボットが「経験」を積むまたとない場となる。

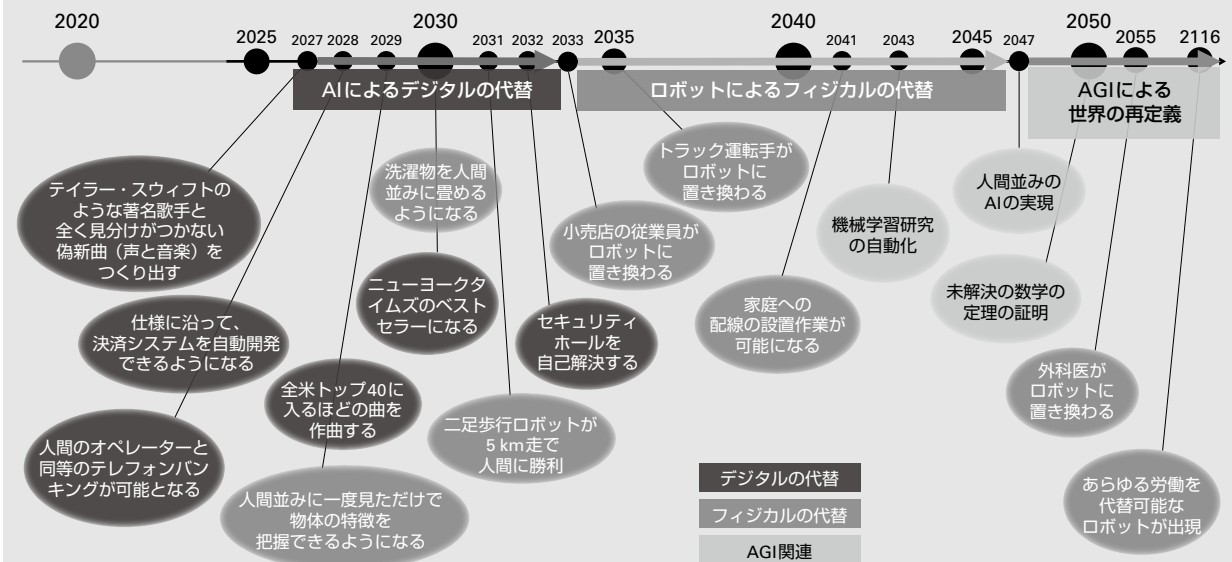
Ⅳ AI進化の時間軸と世界の拡張

1 中長期のAI技術の 進化に関する予測

ドイツのボン大学と米国の人工知能や経済学などの研究者からなる調査団体AI Impactは、2024年1月、論文「THOUSANDS OF AI AUTHORS ON THE FUTURE OF AI」として、AIの技術進化に関する予測をまとめた。このデータは2023年秋にAIの世界的な国際学会（NeurIPS²、ICML³、ICLR⁴、AAAI⁵、IJCAI⁶、JMLR⁷）の査読論文を通過した2778人の研究者に対して実施されたサーベイに基づいている。具体的には、39のタスクに関して「最もリソースのある企業や研究機関が、AIやロボットにより1年以内に50%の確率で実現できる時期はいつ頃か」を問うものである。

代表的なタスクを時系列でピックアップしたものが図6である。AIは、2030年にかけては、デジタル分野での活用が先行する。具体的には、プログラミングやセキュリティ対策、音楽のようなコンテンツ制作まで、幅広いタスクを支援するようになり、一部のタス

図6 AI技術の進化に関する予測



出所)「THOUSANDS OF AI AUTHORS ON THE FUTURE OF AI」を参考に作成

クは自動化が予想される。その後、2040年に向けては、トラックなどの長距離輸送の自動運転が実現し、小売店での接客をはじめ、私たちの日常生活を手助けするサービスロボットの登場も視野に入るなど、AIは自動車やロボットと融合し、フィジカルでの活用が一層進むと思われる。さらに先の2050年以降については、汎用人工知能やあらゆる労働を代替する汎用ロボットの登場が考えられるが、これらは20年以上先のものであり、現時点では不確実性が高いといえる。

ボン大学らの予測は、あくまでAI技術に関する基礎研究が成功し、利用可能になる時期である。これまで、研究機関で開発された技術が世に出るまでには、技術の実用化にかかるコストや社会のニーズがすぐには顕在化しなかったなどの理由で、10年、ときに20年以上の歳月が必要であった。しかし、もしAIへの投資が継続し、「大きなAI」「小さな

AI」のようにAIが社会インフラとして溶け込み、汎用技術となれば、この状況は違ったものとなる。汎用技術である電気と同様に、AIは身近で使いやすいものとなり、AIを活用した製品やサービスを開発する企業が多く誕生し、新たな産業を形成するだろう。基礎研究が世に出るまでの時間は、大幅に短縮される。

AI技術のニーズも後押しする。パーソル総合研究所の調査「労働市場の未来推計2035」によれば、2035年には国内で「1775万時間／日の労働力不足」が発生すると予測されている。AIやロボットは、これを補完する労働力として期待されており、AI技術への関心やニーズも高い。実用化に向けたコストは問題であるが、労働力不足による経済活動の縮小、社会に必須の基本的なサービスさえ提供できなくなるならば、各企業や業界の自助努力ではなく、国家の戦略として対処する必要性が

ある。社会インフラとしてAIやロボットを捉え、戦略的に次世代の労働力を生産する時代の到来が予見される。

2 世界の拡張

AIやロボットが普及する2030～2040年代は、現在の社会とは全く異なるものとなっているだろう。自動車や家電製品、スマートフォンやスマートグラス、壁面や照明、センサーなど、あらゆるモノに小さなAIが組み込まれ、「スマートシングス」へと進化する。小さなAIは互いにネットワークを形成し、群れとなったAIらは空間や都市そのものを知的なものへと変える。これは、空間が知性を持つ「Ambient Intelligence（環境知能）」の実現につながるだろう。Ambient Intelligenceによって人間はモノと対話し、より効率的に課題を解決できるようになる。また、人間と同様に言葉を操り、知的に振る舞う空間の登場は、人間にとって世界の再定義ともいえ、まるで実世界が拡張したような体験を生み出すだろう。

一方で、社会の隅々にまで広がった小さなAIのエコシステムは、デジタルに別の世界を形成する。小さなAIが収集するデータは社会を切り取った断片であり、これを集約し再構成すれば、実世界をデジタルに再現した

ミラーワールドにつながる。この再現者は、高度な推論能力を持つ大きなAIが担うと思われる。大きなAIは小さなAIと協力し、ミラーワールドの中で実世界の問題を予見し、対応策を打つことも可能となるだろう。

遠い未来の2040年代、世界は、AIによってデジタルとフィジカルの両面で拡張し、新時代を迎えているに違いない。

注

- 1 Peter Huttenlocher (1931–2013), Nature, 502, 2013
- 2 Neural Information Processing Systems
- 3 The International Conference on Machine Learning
- 4 The International Conference on Learning Representations
- 5 The Association for the Advancement of Artificial Intelligence
- 6 International Joint Conference on Artificial Intelligence
- 7 The Journal of Machine Learning Research

著者

長谷佳明（ながやよしあき）

野村総合研究所（NRI）未来創発センターデジタル社会・経済研究室エキスパートストラテジスト
専門は人工知能、ロボティクス、社会課題の分析など