

基礎研 レポート

DeepSeek に見る AI の未来 近年の AI 進化の背景とは

金融研究部 准主任研究員 原田 哲志
(03)3512-1860 harada@nli-research.co.jp

1—DeepSeek の衝撃

AI 開発を手がける中国の新興企業 DeepSeek が開発した AI モデルが低コストでありながらベンチマークテストにおいて ChatGPT と同等の高スコアを記録したことが注目を集めた¹。大規模言語モデル (Large Language Model : LLM) など AI モデルの開発は資金力に優れる米国のテック企業が主導してきたが、その優位を揺るがすとの見方が広がった。

DeepSeek の登場は、オープンソースによる大規模言語モデルの改良を加速させると期待されており、世界的に AI 技術の開発競争に拍車がかかるとみられる。これにより、資金力に依存せずとも高性能な AI が構築可能となる可能性が広がり、生成 AI の利用がより幅広い層に浸透する契機となるだろう。

また、AI 開発の効率化により、特定の用途に特化した AI の導入コストも引き下げられると考えられており、製造、金融、教育など多様な産業での活用が進むことが期待されている。一方で、AI の処理を担う半導体需要は、一部で供給過多や価格下落の影響を受ける懸念も出ている。これが半導体市場全体の調整圧力となる可能性もある。ただし、エヌビディアなどの AI 向け GPU に特化した企業は依然として高い注目を集めており、長期的な AI 半導体需要の増加期待を背景に株価も堅調に推移している。DeepSeek の登場は世界の AI モデルの開発において米国の主要テック企業が独占する状況から様々な企業や開発者に裾野が広がる可能性がある。

2—DeepSeek のモデルの特徴

DeepSeek の言語モデルは、最新の技術を積極的に取り入れ、高速で効率的かつ高性能な推論を実現している点が大きな特徴である。その代表的な技術要素として「8-bit 化」、「マルチトークン生成」、「ミクスチャー・オブ・エキスパート (MoE)」の 3 つが挙げられる(図表 1)。

¹ DeepSeek の低コスト化や性能については誇張があるとの指摘もあるが、本稿では近年の AI モデル改良の一例として、その性質や特徴について見ていきたい。

まず 8-bit 化とは、従来の 16-bit や 32-bit の精度よりも桁数とデータ量が小さい 8-bit で重みや演算を行うことで、計算資源を大幅に削減する技術である。これにより、モデルのメモリ消費量が減り、より小さなデバイスでも大規模言語モデルの運用が可能になる。特に推論時の計算コストを抑えることで、リアルタイム応答性の向上にも寄与している。DeepSeek ではこの 8-bit 化を採用し、効率性と速度のバランスを高めている。

マルチトークン生成は、一度のステップで複数のトークン（単語や文字）を同時に生成するアプローチである。従来の言語モデルは 1 トークンずつ逐次生成していたため、長文生成には時間がかかっていたが、この手法により生成速度が大幅に改善された。DeepSeek はこの並列生成能力を活かし、対話応答や文章生成のスループットを向上させている。

ミクスチャー・オブ・エキスパート（Mixture of Experts: MoE）は、複数の専門的なサブモデルを組み合わせ、入力に応じて最適なものだけを動的に選択する構造である。すべてのサブモデルを毎回使うのではなく必要な部分だけを使うため、パラメータ数を抑えながらも表現力の高いモデルが実現可能だ。DeepSeek ではこの MoE 構造を採用し、大規模化による性能向上と計算効率の両立を図っている。こうした改善の組み合わせにより、DeepSeek は低コストで高い性能を実現している。

図表 1 DeepSeek の主な技術的特徴

8-bit化	モデルで用いるデータの大きさを通常の16/32bitから8bitに縮小。モデルサイズとメモリ使用量を大幅に削減し、推論や学習の高速化・省電力化を実現できる。
マルチトークン生成	1ステップで複数のトークンを同時に生成する技術。通常の逐次生成より高速で、推論のスループットが向上し、大規模モデルの応答遅延を軽減できる。
ミクスチャー・オブ・エキスパート（MoE）	複数のエキスパートサブモデルの中から一部のみを入力ごとに動的に選択して使うモデル構造。精度を保ちつつ計算コストを抑えるのが特長。

（資料）各所資料をもとに筆者作成

3—AI モデルの非連続な進化

このように、DeepSeek はモデルの改善により性能を大きく向上させた。ただし、こうしたモデルの改善は過去にも起こってきたものである。AI に関連する代表的なモデルの革新としては「深層学習」や「Transformer」が挙げられる。

深層学習は簡単に言うと、入力層と出力層の間に隠れ層をはさみ、多層構造としたモデルだ。深層学習により、複雑な事象も処理できるようになった。

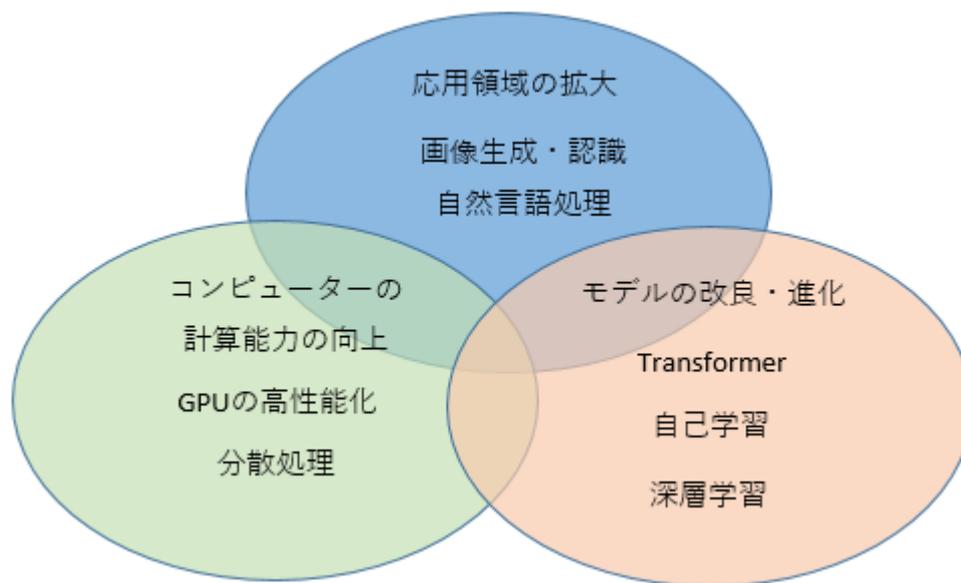
Transformer は入力データの中から重要な部分に焦点を当てる「attention 機構」と呼ばれる仕組みを持っている。Transformer の活用により、精度が向上するとともに大量の情報を並列処理することが可能となった。これにより、それまでの生成 AI よりも大幅に性能が向上した「ChatGPT」が誕生した。

AI の進化は「単純なパラメータ数の増加」だけでなく、「モデルの改善」や「応用領域の拡大」が複合していると言える(図表 2)。コンピュータの計算能力の進化やそれに伴う(AI モデルの)パラメータ数の増加は連続的であり、成長予測は比較的行いやすい。

一方で、モデルの進化は非連続であり、いつ革新的なモデルが登場するのか予測は難しい。ただし、

AI モデルは世界中で様々な改善が研究されており、今後もモデルの革新は起こっていくと考えられる。

図表2 AIの進化



(資料) 筆者作成

4—AIの応用領域の拡大

従来の主要な生成 AI モデルの開発は潤沢な資金を持つ米国の巨大テック企業が先行してきた。しかし、DeepSeek のようなオープンソースモデルの開発が進むことは世界中の開発者が自由に AI モデルを共有し進化させていくことにつながる。

こうした変化により、AI モデルの基盤自体はコモディティー化（汎用品化）し、代わりに AI モデルをどのように応用し、社会に実装していくかが重要となっていく可能性がある。こうした場合、これまで少数の企業に占められていた生成 AI サービスは、様々な事業者や領域での活用が広がっていくと考えられる。

特に、従来の自然言語処理（NLP）分野を超え、動画生成やエンボディド AI（身体性を持つ AI）といった新領域への応用が進んでおり、AI の活用範囲は爆発的に拡大している。

エンボディド AI とは、物理的な実体を持ち、環境と相互作用しながら学習・判断を行う AI を指す。従来の AI が、チャットや画像生成など主にデジタル空間で動作するのに対し、エンボディド AI はロボットやスマートデバイスで実装され、現実世界で物体を操作したり、人間とコミュニケーションを取ることができる。

AI を用いない従来のロボット制御では、ロボットの動きをプログラムで細かく記述する必要があった。しかし、AI を活用したロボット制御では、こうした記述ではなく学習により、人間の曖昧な指示に対しても柔軟な判断と動作が可能となる。

エンボディド AI では、ロボットやスマートデバイスが人間の言語や環境から学び、物理的行動を

最適化する仕組みの実装が進んでいる。これにより、小売や物流、介護といった分野での AI 導入が現実味を帯びてきた。

また、自動車産業をはじめとした製造業でもヒト型ロボットの導入が計画されている。米国の電気自動車大手テスラは自社工場へのヒト型ロボットの導入を計画している(図表 3)²。

図表 3 テスラ社のヒト型ロボット



(資料) テスラ

5—終わりに

今後、AI（人工知能）の性能や応用領域はさらに拡大し、私たちの生活や仕事のあらゆる場面に影響を及ぼすと考えられる。すでに製造、小売、医療、金融、エンターテインメントなど多くの分野で AI は実用化されており、その存在感は日に日に増している。

かつてコンピューターが登場した際、「人間の仕事が奪われるのではないか」という懸念が広がった。しかし実際には、定型的な作業の一部が自動化された一方で、新たな職業や産業が誕生し、人間はより創造的で戦略的な仕事にシフトしていった。これは AI にも当てはまり、例えばコンピューターグラフィックス、ロボット技術、ビッグデータ分析といった分野では、AI の活用によって新しい価値が創出されている。

今後人間に求められるのは、AI を「競争相手」ではなく「協力者」として捉え、積極的に活用していく姿勢である。AI は膨大な情報処理やパターン認識を得意とする一方で、人間が持つ直感や感性、倫理観などには及ばない。AI を活用しつつも、独自の視点や柔軟な発想力を磨くことが、人間としての価値を高める鍵となる。技術の進歩にただ追随するのではなく、人間らしい創意工夫と主体的な姿勢を持つことが、AI 時代における最も重要な生き方であると言える。

² 日経クロステック「人型ロボットが自動車工場や物流倉庫へ、テスラは 25 年から数千台配備」、2024 年 11 月 21 日

本資料記載のデータは各種の情報源から入手・加工したものであり、その正確性と完全性を保証するものではありません。また、本資料は情報提供が目的であり、記載の意見や予測は、いかなる契約の締結や解約を勧誘するものではありません。