

AIの機能進化予測の難しさ

深層学習の登場以降のAIの機能進化は専門家すら予測ができないほど急速に発達した。一方で、現在の生成AIの機能進化がこれまでと同様のペースで進むのかにはいくつかの課題の存在が指摘されている。しかし、これらの課題を解決するための取り組みも急速に進んでおり、AIの機能進化は未だ予見不可能だ。

■ 専門家ですら過小評価した生成AIの機能進化

OpenAIが大規模言語モデルのGPTを発表したのは2018年のことだった。そこからわずか4年後の2022年には、文書、ソースコード、画像、音声といったコンテンツを人間の指示にしたがって生成できる生成AIのChatGPTが登場した。

生成AIの機能進化はさらに加速している。そしてその進化のスピードは専門家の予想をはるかに上回っている。昨年5月、東京大学の松尾豊教授は「動画生成AIが登場するにはあと数年から十年はかかる」と予想していたが¹⁾、そのわずか9か月後、OpenAIがSoraというリアルな動画生成AIを発表した。

実際、これまでのAIの機能進化の速度を正確に予測した専門家はいない。多くの専門家はAIの機能進化を常に過小評価してきた。

2024年1月に公表された、AIの研究者2,778人に対して「AIが人間の補助なしに人間を上回る能力を示す可能性」を聞いた調査²⁾によると、2027年までに実現する可能性は10%、2047年には50%との結果が紹介されている。ただ、前年に行った同様の調査と比べると、実現時期の予測は13年早くなっている。専門家ですらこのような状況なのだ。

■ AIの進化の制約条件は？

一方で、これまでと同じペースでAIの機能進化が続くだろうという楽観的な見通しに対していくつかの懸念も指摘されている。

■ エネルギーの消費問題

国際エネルギー機関（IEA）が世界の電力市場の動向について分析した2024年報告書³⁾では、世界的な電力消費量の伸びは若干鈍化するものの、データセンター、AI、暗号資産などのデジタル領域でのエネルギー消費は今後ますます増大すると予想している。中でもAI関連の電力消費は2026年までに現在の10倍に達すると予想している。

また同報告書ではChatGPTが一件のリクエストを処理するのに必要な電力量は、Google検索一件に必要な電力量の約10倍にのぼると推計しており、生成AIの利用が急激に拡大した場合、より効率的な電力消費を求められることになる。実際、アメリカでは暗号資産のマイニングの大量の電力消費に30%の課税を実施すべきという提案がなされたこともある⁴⁾。

■ 計算資源の供給制約

また、現在の生成AIが学習データからAIモデルを構築（トレーニング）する際には、想像を絶するほどの膨大な演算が必要だ。これまでコンピューターを動かしていた中央演算処理装置（CPU）は多様で複雑な処理を効率よく行うことを目的に設計されていたため、単純な繰り返しの演算処理を大量・高速に行う作業にはあまり向いていなかった。そこで、繰り返し処理を大量・高速に行える画像処理演算装置（GPU）に注目が集まる。実際にGPUを活用することでAIの学習速度は飛躍的に向上した。GPUの世界的なサプライヤーであるNVIDIAの業績は急拡大し、2月には同社の時価総額は2兆ドルを一時超えた。NVIDIAはAIのトレーニングに最適化されたGPUを供給することで、生成AIのトレーニング領域ではほぼデファクトスタンダードとなっている。

NOTE

- 1) OpenAIがSoraで目指す超知能への道 | AI新聞
<https://community.exawizards.com/aishinbun/news02-20-2024/>
- 2) "Thousands of AI Authors on the Future of AI"
<https://arxiv.org/abs/2401.02843>
- 3) "Electricity 2024" - IEA
<https://iea.blob.core.windows.net/assets/ddd078a8-422b-44a9-a668-52355f24133b/Electricity2024-Analysisandforecastto2026.pdf>
- 4) バイデン大統領、仮想通貨マイニングの電気使用量に30%の課税を提案 | コインテレグラフ ジャパン
<https://jp.cointelegraph.com/news/biden-budget-proposes-30-tax-on-crypto-mining-electricity-usage>
- 5) Will We Run Out of ML Data? Evidence From Projecting Dataset Size Trends - Epoch
<https://epochai.org/blog/will-we-run-out-of-ml-data-evidence-from-projecting-dataset>
- 6) NVIDIAのテクノロジーを搭載した新世代のシステムが、今後のグリーンな道筋を示すNVIDIA
<https://blogs.nvidia.co.jp/2022/11/16/green/>
- 7) Microsoftが1.58ビットの大規模言語モデルをリリース、行列計算を足し算にできて計算コスト激減へ - GIGAZINE
<https://gigazine.net/news/20240229-microsoft-1bit-llm/>

一方でNVIDIAにGPUの供給がほぼ独占されていることで、NVIDIAのGPUの供給能力がAIの機能進化のボトルネックになるのではないかと危惧する声がある。

■学習データの枯渇

生成AIのトレーニングには膨大な学習データが必要だ。学習データにはテキスト、音声、画像、動画といった様々なデータが必要だが、それらの大半はネット上に存在するデータが活用される。

2022年、AI研究グループであるEPOCHIは、高品質の学習データは早晚枯渇する可能性があるという予測を発表した⁵⁾。ここでの高品質の学習データとは、テキストデータで言えばWikipediaの記事、ニュース記事、科学論文などである。これらのデータは早ければ2026年にすべて使いつくされる可能性を指摘したのだ。

実際、OpenAIがこれまで発表してきたAIモデルのパラメータ数（学習データの量とほぼ同義）は、初代GPTは1.1億だったものが、GPT-2では15億、GPT-3では1,750億、GPT-4は1兆とわずか4年で約一万倍の学習データを利用している。

さらに、最近では大手ニュースサイトがAIへの自社コンテンツの利用を禁止する動きも出てきている。良質な学習データが利用できなくなると、AIの機能進化にも悪影響を及ぼす可能性がある。

AIの機能進化予測の難しさ

これまで見てきたようにAIの機能進化が今後ずっと同じか、もしくは加速したペースで進むという楽観論に対する懸念を紹介してきた。しかし、これらの懸念を解消しようとする研究も行われている。

電力消費に関しては、現在の「非効率」なCPU中心のデータセンターの構成をGPUに置き換えることで大幅な計算の効率化と電力消費の削減が可能とされている⁶⁾。またチップ自体の省電力化や、データセンターでのクリーンエネルギーの利用なども進められている。

また計算資源については、一部で話題になっているMicrosoftが発表した「BitNet」と呼ばれるAI学習の新たなモデルは、それまでの生成AIのトレーニングに必要な行列式の乗算処理を加算処理で行えるモデルである⁷⁾。仮にこの技術が本物だった場合、今よりも計算負荷の遥かに少ないAI学習が可能となる。そうすると計算資源の制約を回避できるかもしれない。

また学習データの枯渇への対処として「低品質のデータ」の活用も研究されている。低品質のデータとはSNSへの投稿やブログ記事といった「必ずしも品質が保証されていないデータ」のことだ。これらのデータは高品質なデータよりも遥かに多い。これらの低品質データが利用できれば学習データの枯渇も回避できる可能性がある。さらに、「AI自身に学習データを作らせてそこから学習する」というモデルも研究されている。そうなれば学習データの枯渇という問題自体が解消するかもしれない。

このようにAIの進化にはいつパラダイムシフトが起きるか予測不能だ。そういうわけでAIの機能進化予測を筆者に尋ねるのは今後ともご遠慮ください。

Writer's Profile



柏木 亮二 Ryoji Kashiwagi

金融デジタルビジネスリサーチ部
 エキスパート研究員
 専門はIT事業戦略分析
focus@nri.co.jp