

## 生成 AI への期待と日本における課題

三谷 慶一郎

武蔵野大学国際総合研究所客員教授



株式会社 NTT データ経営研究所 主席研究員 エグゼクティブ・コンサルタント  
筑波大学大学院ビジネス科学研究科博士課程修了。博士（経営学）  
企業や行政機関におけるデジタル戦略立案やサービスデザインに関連するコンサルティングを実施。経営情報学会監事、情報社会学会理事、日本システム監査人協会副会長。経済産業省「デジタル産業への変革に向けた研究会」座長、「デジタル時代の人材政策に関する検討会」座長、「DX 銘柄」評価委員等を歴任。監訳書に「デジタルトランスフォーメーション経営戦略」

生成 AI の台頭

「生成 AI（Generative AI）」に大きな注目が集まっている。これは、入力した命令（プロンプト）に対して、適切なテキストや画像、音楽、動画などを自動的に作り出す人工知能のことである。中でも、あたかも人間と会話しているように振る舞う AI である ChatGPT は様々な場面で話題を呼んでいる。ChatGPT は、2022 年 11 月 30 日に OpenAI から無償で公開された後、何と 4 日後には、利用者は世界中で 100 万人を突破したという。これは、インスタグラム（2 年 4 か月）、TikTok（9 か月）と比較しても、驚くべき普及スピードであり、生成 AI への期待の大きさがうかがえる。

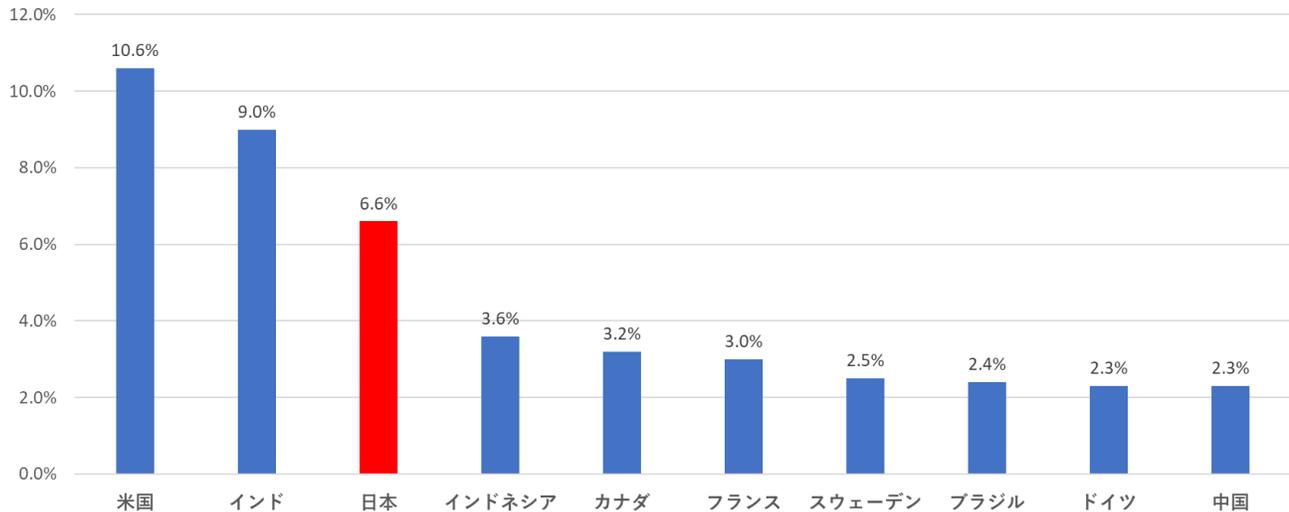
実は日本は、以前から諸外国と比較して AI をあまり積極的に活用していない国だと言われていた。総務省の「令和元年版情報通信白書」\*1 によれば、企業における AI アクティブプレイヤー（「一部の業務を AI に置き換えている」もしくは「一部の業務で AI のパイロット運用を行っている」企業）の割合は、調査対象である 7 か国中最下位（39%）である。情報処理推進機構（IPA）の「DX 白書 2023」\*2 でも、AI の利活用状況において「全社で導入している」と答えた企業は、米国が 19.7% であるのに対して、日本はわずか 2.4% でしかない。このような状況を踏まえると、ここ半年くらいの生成 AI に対する日本国内の熱狂ぶりには少々戸惑いを覚えてしまう。

野村総合研究所の調査\*3 によれば、2023 年 4 月時点での日本における ChatGPT に関する認知度は 61.3% に達しており、実際に利用したことがある人は

全体で 12.1% とのことだ。特に 10 代から 30 代の男性の利用率が 20% 前後と比較的高い。職業別では、大学生・大学院生・専門学生が最も高い（21.6%）という若干考えさせられる結果も出ている。また、2022 年 12 月から 2023 年 4 月までの 5 か月間における OpenAI へのアクセス数を分析すると日本

は、米国、インドに次いで世界第3位（6.6%）だそう。日本語での利用が可能だということ、汎用に活用できる技術であることも大きい、日本人の「新しいモノ好き」という性格が強く現れているともいえる。（図表1参照）

【図表1】 Openai.comの国別トラフィックシェア（2022年11月～2023年4月）



※野村総合研究所「日本のChatGPT利用動向（2023年4月時点）」より作成

一方、フェイク画像や動画が容易に作られてしまうこと、作られたテキストに「もっともらしい嘘（Hallucination）」が含まれる場合があること、知的財産権が脅かされたり、機密情報が流出したりする可能性があること等、生成AIには様々な懸念があるという声も多い。2023年3月には非営利団体 Future of Life Institute が、「AIが人間並みの競争力を持つようになり、人類と社会に対する潜在的なリスクをもたらすので、技術の危険性が適切に評価されるまでGPT-4よりも強力な技術の開発を6ヶ月間休止するよう求める」という公開書簡を出しておりイーロン・マスク氏、スティーブ・ウォズニアク氏ら数千人以上の技術関係者がこれに署名している。また、5月には非営利団体 Center for AI Safety (CAIS) が「AIによる絶滅のリスクを軽減することは、パンデミックや核戦争などの他の社会規模のリスクと並んで世界的な優先事項であるべき」という声明を公表し、OpenAI や Google Deep Mind のCEOを含む350人以上が署名している。

2023年4月に行われた「G7群馬高崎デジタル・技術大臣会合」でも生成AIは大きなトピックになっており、リスクを踏まえつつこの技術を活用しようという議論がなされている。最終的な閣僚宣言では、「生成AI技術の持つ機会と課題を早急に把握し、技術が発展する中で、安全性と信頼性を促進し続ける必要性を認識」し、「AIガバナンス、知的財産権保護、透明性促進、偽情報への対処、責任ある形で生成AIを活用する可能性について、G7における議論を行うための場を設ける」ということが示された。

そもそも「破壊的なイノベーション」というものは、最初は多くの課題を含む劣った技術として現れることが多いと言われている。生成 AI においても、指摘されている課題は多いが、技術の発展と世界中での利用実績の蓄積によって、早晚解決に近づいていくに違いない。また、コロナ禍における Zoom 等と同じように、生成 AI はたくさんの人々が活用し、自らその効用を認識できるような技術であるので、社会全体に大きな影響を及ぼす可能性が高いとも考える。

## 国内における計算資源の増強

それでは生成 AI という新しい技術を日本産業の活性化につなげていくためにはどのような施策が必要になるだろうか。

自民党デジタル社会推進本部がまとめた「AI ホワイトペーパー」<sup>\*4</sup>では、「先行している海外の基盤モデル AI を土台とし、パートナーシップを組む形で応用研究・開発を加速する」とことと、「我が国における基盤モデルに関する基礎的な技術開発能力の構築・強化に取り組む」という二つの方向性が示されている。短期的には前者のように GPT-4 等の確立された技術を活用して独自の用途、目的にチューニングされたサービスを開発していくべきだろう。実際、投資家等とお話をすると、国内においても同様のスタートアップ企業は多く立ち上がり始めているようだ。しかし、より大きな産業の活性化を求めるのであれば、後者の日本独自の生成 AI モデルを作り出していくという方向性は、例えば経済安全保障という側面から見ても避けては通れない。

近年の生成 AI 発展の背景には、Google の研究者が提唱した Transformer というアーキテクチャをベースとした大規模言語モデル (LLM) の存在がある。ここで言う「大規模」とは従来のモデルに対して、「計算量 (コンピュータでの処理量)」「データ量 (入力された情報量)」「パラメータ数 (ディープラーニング技術において扱われる係数の豊富さ)」を大幅に増やして構築されていることを意味する。昨今の研究によってこの3つの要素を巨大化させればさせるほど、より高性能なモデルが生成できることがわかってきている。当然ながらそのためには膨大な計算資源、コンピュータパワーを確保することが不可欠になる。激しい技術開発競争の中で OpenAI が先行できているのも、マイクロソフトによる 10 億ドルあるいは 100 億ドル単位という巨額の投資があったからだろう。2023 年 5 月 30 日に米国の半導体大手であるエヌビディアの時価総額が 1 兆ドルを突破し、META やテスラを越えて世界 6 位に浮上した。この背景にも、生成 AI のための計算資源へのニーズの高まりがあるのは間違いない。

経済産業省はこのような状況を受けて、「半導体・デジタル産業戦略」を改定しつつある。具体的には、生成 AI 産業戦略の方向性として「AI 開発には大規模計算資源が必要だが、足下、国内の開発需要に比して、計算資源は圧倒的に不足している。世界で計算資源の獲得競争が生じており、政府も関与しつつ、可及的速やかに計算資源の整備・拡充を行うべき」ということが述べられている。<sup>\*5</sup>

## デジタルデータの確保

先述の通り、生成 AI の開発は完全に規模の競争になっており、次々と大きなパラメータを持つモデルが作られている。そして、大きなパラメータを持つモデルであればあるほど、学習のために大量のデータが必要になる。パラメータ数に対して学習データが少量だと、これに過剰に適合し過ぎ汎用性がなくなるという「過剰適応（過学習）」という現象が起こってしまうからだ。

2018年に OpenAI が開発した GPT-1 はパラメータ数が 1.17 億個で学習に必要なデータはテキストデータ 4.5GB に過ぎなかったが、2020年の GPT-3 ではパラメータ数は 1,750 億個になり、学習は 45TB の Web データをもとに抽出された 570GB のデータによって行われたそうだ。GPT-4 のパラメータ数は競争と安全性を考慮して非公開になっているが、少なくともこれらよりずっと巨大なことは間違いない。さらに Google の Switch Transfer は 1.6 兆パラメータ、中国の北京智源人工知能研究院が開発した悟道 2.0 は、1.75 兆パラメータに達しているという話もある。（図表 2 参照）もうどのくらいのデータが必要か想像もつかなくなってくる。

【図表 2】生成AIのモデルにおけるパラメータ数

モデル名	開発企業	開発年	パラメータ数
GPT-1	OpenAI	2018年	117,000,000
日本語GPT言語モデル	rinna	2022年	1,300,000,000
GPT-2	OpenAI	2019年	1,500,000,000
HyperCLOVA（日本語特化）	LINE、NAVER	2022年	82,000,000,000
GPT-3	OpenAI	2020年	175,000,000,000
Gopher	DeepMind	2022年	280,000,000,000
GPT-3.5	OpenAI	2022年	355,000,000,000
Switch Transfer	Google Brain	2021年	1,600,000,000,000
悟道2.0	中国北京智源人工知能研究院	2021年	1,750,000,000,000
GPT-4	OpenAI	2023年	非公開

※NTTデータ先端技術「世界で開発が進む大規模言語モデルとは」および公開資料より作成

個人的には、このデジタルデータの確保は、かなり日本にとって大きな課題になるのではないかと考えている。そもそも日本語のデータが少ないということも大きいですが、日本国内にあるデジタルデータは保有している主体ごとに「分断されている」ことが高いハードルになり得る。

日本には、いわゆる GAFAM や BAT といったメガプラットフォーマーは残念ながら存在しない。つまりひとつの民間企業が大規模なデータを保有しているケースはほとんどなく、各企業が自分のデータをそれぞれ持っている状態だということだ。また、国が保有しているデジタルデータも、中国やエストニアのように国家として集中管理しているわけではなく、中央省庁、自治体ごとにバラバラに保有されている。

これらの分断されているデータを社会全体で共有することはかなり難しい。プライバシーに関連する情報、あるいは企業秘密に関する情報は、当然ながら容易に共有するわけにはいかない。ルール上は問題がない場合であっても、各主体が自組織のために収集したデータを目的外に利用されることへの抵抗感は大きいだろう。さらに、日本では民間企業でも行政機関でも、個々のデータそのものに関する標準化が進んでいない。特定の組織内の業務でしか通じない言葉があったり、同じことを意味するデータなのに組織によって違う言葉で表現されたりしていることはよく見受けられる光景だ。この背景には、社会全体として組織を越えた人材移動が少ないことがある。終身雇用を前提とする雇用環境の中では、特定の人材が長期間同一の組織の中に所属し続けるため、例えば同じ業務であっても、企業毎にそのプロセスや扱うデータは自然に独自進化を遂げるからだ。データが分断されておりかつ標準化されていないため、容易に共有化できない環境、これはとても大きな課題である。

## データ連携基盤の構築

先述の「半導体・デジタル産業戦略（改定案）」の中では、データに関しては「公的機関が保存するデータにアクセスできる仕組みを構築する」ということが書かれているがこれだけだと心もとない。民間企業を含め、現在主体ごとにバラバラに分断されているデータを円滑に共有するための何らかの「基盤」を作る必要がある。

全くの私見ではあるが、現時点で最も有効ではないかと考えているのが、先の G7 において日本が提案した、「ウラノスエコシステム（Ouranos Ecosystem）」だ。これは、社会課題の解決に必要となるデータを、企業や業界を横断し連携・活用する「データ連携」に関するイニシアティブを意味するものだ。経済産業省、関係省庁、情報処理推進機構（IPA）、新エネルギー・産業技術総合開発機構（NEDO）等が連携しながら、運用及び管理を行う者が異なる複数の情報処理システムの連携の仕組みに関して、アーキテクチャの設計、研究開発・実証、社会実装・普及に関する検討を進めている。

この取り組みが目指しているのは、脱炭素や人手不足、災害激甚化といった社会課題の解決や産業構造の変革であり、先行して蓄電池・自動車業界を対象としたサプライチェーンデータ連携基盤の構築に取り組む予定になっている。「ウラノスエコシステム」は決して生成 AI を意識したものではないが、「企業等が保有している特定のデータを、データ主権を保ったまま共有化するための仕組み」をつくるのが狙いだというなら、これに拡大適用できる余地があるように思う。

生成 AI の波に乗り遅れず、これを日本の競争力に直結させていくためには、計算資源の抜本的増強と、社会全体としてのデータ連携基盤の構築を急がなければならない。さらに言えば、この 2 つの投資は生成 AI だけに留まらず、デジタル技術活用に関する日本の弱点そのものの克服につながる可能性もある。投資規模は大きいかもしれないが、比較的割の良い賭けかもしれない。

(参考)

- \*1 総務省「令和元年版情報通信白書」2019
- \*2 情報処理推進機構「DX 白書 2023」2023
- \*3 野村総合研究所「日本の ChatGPT 利用動向（2023 年 4 月時点）」2023
- \*4 自民党デジタル社会推進本部 AI の進化と実装に関するプロジェクトチーム「AI ホワイトペーパー – AI 新時代における日本の国家戦略 –」2023
- \*5 経済産業省「半導体・デジタル産業戦略（改定案）」2023