ＡＩ技術の進展と学習データの属性・特性の変容

生成 AI の時代に知的財産権による保護をどう考えるべきか？

弁理士　**酒井　將行**

要　約

2022 年 11 月に、OpenAI から、ChatGPT が発表されて以来、いわゆる「生成 AI」と呼ばれる人工知能技術に対する一大ブームが巻き起こった。2024 年 3 月末に特許庁が公表した 2024 年版の「特許庁ステータスレポート」では、2023 年の特許出願件数（出願日ベース）は 3 年連続で増加し、2019 年以来、4 年ぶりに 30 万件を超えたことが報告されている。これまで、特許出願件数で、増えていたのは主に国際特許出願であった。しかし、2023 年の国際特許出願は 7 万 5600 件で、2022 年から横ばいだった。2023 年に増えたのは国内出願（国際特許出願を除く特許出願）とのことである。しかも、技術分野からすると、情報通信分野での増加が全体を牽引したようである。同分野には、技術進歩が著しい AI（人工知能）が含まれていることから、件数の増加の一因との指摘もある。一方で、これまでの「データ駆動型人工知能技術」をはじめとして「生成 AI」の応用が極めて大きな技術的なインパクトをもつだけでなく、その技術的な応用の速度がこれまでになく速いことを考慮して、どのような知的財産保護を考えるべきなのかについて検討する。

目　次

1．はじめに

2022年11月に、ChatGPT（Chat Generative Pre-trained Transformer）が公開され、いわゆる「生成AI（Artificial Intelligence）」のブームともいえる現象が発生している。このようなAI技術については、モデルの設計や学習処理において、いわゆる「プラットフォーマ」と呼ばれる世界的巨大企業が主要な役割を担っていることは周知のとおりである。そして、プラットフォームの形成そのものが、特許などの知的財産権と同等かそれ以上のビジネスの独占を可能とする手法であることについての指摘もされてきた(1)。そして、このようなプラットフォーマとしては、これまで“GAFAM”と称される、Google（Alphabet）、Apple、Facebook（Meta）、Amazon、MicrosoftなどのICT企業が、テック業界を支配し、世界をも支配しているというような言説がされる場合もあった。しかしながら、「生成AI」の登場により、「生成AI」への取り組みの状況によって、このような5社の寡占状況に変化が現れる（あるいは、現れている）というような見解も出てくるような状況となっている(2)。

技術的な側面で言えば、いわゆる生成AIによって、単に、「自然言語をコンピュータが人間並みに操れるようになった」ということを、はるかに超えると考えられる（少なくとも、筆者を含めた非研究者からはそのように見える）技術が開発され、実用化されつつある。たとえば、後述するような「自動運転技術」に対する「大規模言語モデルの応用」などは、その典型であろう。技術の進展が、あまりにも速すぎる(3)が故に、知的財産権を考えるときにも、従来の法制度の枠組みは、とてもそのスピードに追い付けていないように見えるというのが正直な感想である。

このように、技術的にも、経済的にも、極めて変化の激しい分野ではあるものの、だからこそ、その分野の中でも、将来において、どのような知的財産を想定して、どのような権利の取得を目指すのか、ということは、実務家にとって、最も重要と考える。そこで、筆者の知る範囲に制限されるものではあるものの、人工知能技術の現状の理解と、それに基づく、知的財産権、特に、特許権について、想定される内容を以下に検討する。

2．人工知能技術の現状の概観

（1）大規模言語モデル（LLM）の登場とその利用による人工知能技術の環境の変化

以下、ここ数年で生じた人工知能技術の変化について、簡単にまとめてみたい。

第1に、大規模言語モデルの登場によって、これまでのコンピュータと人間との間のマンマシンインタフェースに大きな変動が生じた。

---

(1) 酒井將行「プラットフォーム型およびデータ駆動型ビジネスモデルに対する知的財産保護」別冊パテント23号（日本弁理士会中央知的財産研究所 研究報告48号『「超スマート社会（Society 5.0）」に適合する知的財産保護の制度のあり方』）

(2) 田中道明「IT業界の覇権は「GAFAM」から「GOMA」に変わる…ビッグテックの力関係を一変させる「生成AI」のインパクト」（https://president.jp/articles/-/78646?page=3、2024/02/16）では、「米国で長い歴史を持つ月刊誌の『アトランティック（The Atlantic）』が、23年10月に「AIの未来はGOMAだ（The Future of AI Is GOMA）」と題する記事を掲載した」ことについての解説がある。ここで、GOMAとは、グーグル（Google）、オープンAI（OpenAI）、マイクロソフト（Microsoft）、そしてアンソロピック（Anthropic）の4社であるとされている。一方で、真壁昭夫「なぜアップルとグーグルは「GAFA」から脱落したのか…代わって時価総額が急増している「4つの巨人」の正体」（https://president.jp/articles/-/80292?page=1、2024/04/08）では、2024年4月の段階で「米国株式市場で、これまで主役を演じてきた“マグニフィセント・セブン〔壮大な7社＝テスラ、アップル、アルファベット（グーグル親会社）、エヌビディア、アマゾン、メタ、マイクロソフト〕”から、テスラ、アップル、アルファベットの3社が脱落しつつある。残りの4社を称して“ファビュラス4”という。マグニフィセント・セブンからファビュラス4へ、世界の投資家の注目は移っている。」との指摘がされるような状況となっている。もちろん、その後、さらに、各社から「生成AI」技術についての発表が相次いでいる状況であり、このような状況が今後も続くかについても、判断の分かれるところである。たとえば、2024年4月末には、「アルファベットとマイクロソフト株上昇、AI投資が成長に寄与」との報道もなされた（https://jp.reuters.com/markets/world-indices/VSUVLBD4QNOLXFZJCQ2TFP6NNQ-2024-04-26/ ）。ただし、いずれにしても、「生成AI」の登場で、これまでの本技術分野のリーディングカンパニーにも、大きな変動が生じつつあるとの認識が、技術分野の専門家だけでなく、経済分野からも広がっていることは確かなようである。

すなわち、私たちが、日常使用しているような言語（自然言語）を、ほぼ人間がしゃべったり、書いたりするのと遜色のない程度まで、AIが操れる状況が生じた。この結果、たとえば、これまでは、コンピュータでの「検索」といえば、Google検索を用いて、「キーワード」を入力して、関連するウェブ上のページの一覧が表示されて、その中からユーザがページを選択することであり、さらに、ユーザがそのページの内容を確認するということであった。これに対して、2022年11月のChatGPTの登場により、たとえば、Microsoft社の提供するサービスとしては、Copilotがリリースされ、当初は、作業ウィンドウを開いて、ユーザが言葉でやりたいことをCopilotに伝えると、タスクの実行、情報の検索、コンテンツ生成などを行うサービスが実用化された。現在は、さらに機能拡張版の提供も開始されている。すなわち、現在では、「検索」は、いわゆる「チャットボット」によるインタフェースに置き換わりつつある。

この結果、「検索する」という日本語の動詞に対応して、これまでは「ググる」という表現が使用されることがあったものの、現在では、一部のユーザでは、「コパる」との用語も使用され始めているようである。

第2には、このような単なるインタフェースの変化にとどまらず、生成AIを用いることが有望な分野として、「基本リサーチ」「ドラフト作成」「異視点の抽出」などが指摘されている。たとえば、「基本リサーチ」とは、Googleのようなキーワード検索とは異なり、対話型検索と呼ばれる、より人間とのやりとりに近い形で欲しい情報を入手することを指す。調査・分析作業などで、素早く全体像を捉えたい場合などに、概要や背景、関連情報などバランスの取れた文章を作成してくれる。

「ドラフト作成」は、文字通り、論文や提案書の基本構成をまとめる時などにChatGPTを使うと、これまでの検索エンジンとは異なり、文脈に沿った文章を作成してくれるものである。そして、ドラフト作成という意味では、文章以外にも、コンピュータプログラミングにも応用が広がっている。また、「異視点の抽出」とは、自分とは異なる意見、気づいていない視点を引き出すためにChatGPTを使おうという試みを指す。たとえば、新規事業のアイディア出しなどにも使用される例が増えているようである。

また、後述するような「大規模言語モデル（LLM）」の登場によって、生成AIに雇用を奪われる危険性が高い職種(4)として、「法律」「金融」「建築」、さらに「高等教育者（Postsecondary teachers）」など7分野が挙げられる調査も発表されている(5)。一方で、生成AIについては、一見確からしい誤情報を出す「ハルシネーション（幻覚）」や回答に偏り（バイアス）が存在することが課題として挙げられている。

このような課題に付随して、「大規模言語モデル」といっても、実は、膨大なデータに基づいて、最も尤もらしい「次に出てきそうな単語」を、選び出しているに過ぎない(6)、というような見解も聞かれるところである。

---

(3) 1年前であれば、「動画」を生成するのは、技術的には大変困難で、生成の対象が静止画から動画へ移行するには、それなりの時間を要すると考える人が、いわゆる専門家にも多かったのではないかと思われる。ところが、OpenAI社が、2024年2月15日、「Sora」という最新動画生成AIモデルを発表したことで、状況は、大きく変わった（変わりつつある）といえるようになった。今や、分野によっては、2か月前の話は、すでに古くなっている、というような進歩の目まぐるしさである。

(4) https://shrm-res.cloudinary.com/image/upload/v1706729099/AI/CPR-230956_Research_Gen-AI-Workplace_FINAL_1.pdf

(5) 関連して、「人工知能が人間の仕事を奪ってしまうのか？」ということが議論されることも増えている。しかしながら、筆者としては、1930年代の世界恐慌のころに、MITの学長であったカール・T・コンプトンが、（個人としてみれば、仮に雇用が奪われることがあったとしても、テクノロジーの進歩により、全体的により多くの雇用が創出されたのだから）「業界全体としては」「技術的失業は神話である」と言及していることが、的を射ていると考える。一部では、「AIには、創作的な仕事はできない」「動画が作れるといっても、人間から入力された文章（プロンプト）に基づいて、動画を作成しているだけ」との見解もある。もっとも、この点についても、そもそも、「創作的とは何か？」という点については、人間の側でも統一的な見解はないのではないだろうか？その意味では、「（人間の側から見て）創作的に見えるもの」を、AIが自発的に作り出すようになるのも、時間の問題のようにも思われる。

(6) ChatGPTが現れる以前から、この点については、自然言語をAIが操っているように見える、いわゆる「Chatbot」は、「Stochastic parrots（確率学的オウム）」に過ぎない、という議論もされてきたところである（Bender EM, et. al., “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”, On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?）。本稿では、もちろん、このような問題に立ち入ることができるものではないが、しかし、この用語が使われた当時（2021年）と比べて、現在は、少なくとも外見上、それを超えているように見える事象が現れているようにも感じられる。もちろん、この点は、当然ながら、専門家による一層の研究を待つ必要があるのであろう。

（2）大規模言語モデルの実力は？大規模言語モデルによるAI技術の応用

本稿の執筆時点では、大規模言語モデルによる生成AIが、人間に匹敵するような知的活動の成果を上げていることが報告されている。

たとえば、日本の2024年の大学入試共通テストの問題を、生成AIに解かせてみる、というテストの報告[(7)]がある。ここで、テストされた科目は、「国語」「英語リーディング」「数学ⅠA」「数学ⅡB」「世界史」「日本史」「理科基礎」であった。3つの大規模言語モデルでテストを行った結果、GPT－4では、数学以外のすべての科目で、人間の平均点を超え、数学を含めた7科目の平均でも、人間の平均を上回る結果となったとのことである。もっとも、数学の平均点は、人間（日本の大学受験生）を下回っていた。

一方で、米Google社傘下のGoogle DeepMind社は、2023年12月14日、LLM（大規模言語モデル）とLLMによる幻覚（ハルシネーション）を防止する“評価器”を組み合わせた新たなシステム「FunSearch」を開発し、長年解決不可能な数学問題とされてきた「Cap set問題」を解くことに成功した、と発表した[(8)]。このDeepMind社の発表は、少なくとも、以下のことを再確認させたものといえるであろう。

ⅰ）人間によっては、未だ解決されたことのない問題を、大規模言語モデルが解いた以上は、上述したような『「大規模言語モデル」といっても、実は、膨大なデータに基づいて、最も尤もらしい「次に出てきそうな単語」を、選び出しているに過ぎない』とのあまりに単純な見解は妥当でない、といえるであろう。これは、後述するような「大規模言語モデル」の技術的な構成からも、一定程度、推察されるものなのかもしれない。

ⅱ）ある「生成AI」が、一定条件下で生成したコードを、他のAIが評価するとの試行を繰り返すことで、問題の解決が図られるというソフトウェアシステムの構成が利用された。ただし、このようにして、2つ（またはそれ以上）のAIの相互評価により、一定の成果に到達する、というアルゴリズムは、このような場合に限られず、他のシステムでも、採用されているものである。

ⅲ）ただし、ⅱ）のような問題解決の方法自体は、人間の思考過程とは、おそらくは、異なっているのではないかと推察する。もっとも、正解や回答に到達する道筋は、1つとは限られないのだから、必ずしも人間の思考過程に沿う必要はなく、コンピュータは、コンピュータにとって得意な道筋で、問題を解決すればよいと考えられる。「人間と異なる（ように見える）」ことは、AIの技術としての価値の評価とは、直接の関係はない。もちろん、「人間の思考過程を真似ること」が、演算量の低減や、広範な問題に対処できるという意味での「汎用性」を生むのであれば、それは、その限りにおいて、技術としての意義を有するものと考えられるものの、そのことと、現時点のコンピュータ技術で実現されているものが、技術的に社会にどのような影響を与えうるのか、ということとは別問題であろう。

3．技術的な背景　現在の人工知能技術のトレンドと学習データ

（1）大規模言語モデル、特に、ChatGPTの技術内容の概観

現在、一般に、最も馴染みの深い「大規模言語モデル」といえば、ChatGPTである。その技術内容に深く立ち入ることは、紙幅の関係上難しいところであるが、本稿において主題とする「学習データ」の状況を整理するという意味で、ChatGPTとは、どのような人工知能であるのか？それは、これまで、応用が進んできた人工知能とどこが異なるのか？について、極めて、駆け足となるものの、簡単にまとめる。

重要な点は、「これまでの人工知能技術」にとっての「学習データ」の多くは、学習処理が実行される前

---

(7)　https://www.itmedia.co.jp/news/articles/2401/17/news105.html

(8)　https://www.nature.com/articles/s41586-023-06924-6, 研究チームは、一般的なプログラミング言語であるパイソン（Python）を使用し、解決したい問題の概略を記述することから始めた。しかし、問題の解き方を指定するプログラムの行は空白にしておいた。そこにファンサーチを使用する。つまり、Codeyに空白を埋めさせ、実際に問題を解決できるコードを提案させた、とされる。

の段階で、正解データが準備されているような「教師あり学習」のためのデータが主であり、そのような「教師あり学習のための学習データ」を「いかに準備するか？」が、技術的にも、また、その保護のための知的財産権の検討のためにも、第一義的には重要であると考えられてきた。このような正解データを学習データに付する作業は、一般に、「アノテーション」と呼ばれる。たとえば、自動運転などの技術開発では、車の周りの画像データを学習データとする場合、画像中の「他の車」や「歩行者」などを枠（Bounding Box）で囲むことで、学習データとすることなどが行われてきた。これについては、いわゆる「自動アノテーション」の技術開発なども精力的に行われてきているところである(9)。

しかしながら、後述のとおり、いわゆる「教師データ」の事前準備なしに、人工知能の学習が可能となる技術が、大規模言語モデルのこれまでの進展に、大きく寄与している。この意味では、広義の「教師なし学習」の進展を、これまで以上に考慮して、「学習データ」について検討を行うことが必要となってきていると考える。

一方で、これも後述するものの、そのような状況にあっても、「教師あり学習」のための「学習データ」の重要性や、特定のドメイン（たとえば、会社内）でのいわゆる「業務データベース」の重要度も、ある局面では高まっているなど、さまざまな形態の「データ」が、人工知能技術にとって、極めて重要な位置を占める状況になっている。これらの点については、これまでも、いわゆる「データ駆動型人工知能」に対する「学習データ」という意味で、検討されてきたものの(10)、「大規模言語モデル」の登場によって、さらに検討を要する局面が生じているともいえる。

以下、「大規模言語モデル」として、最も有名である ChatGPT の技術内容について概略を説明する。ただし、現行バージョンよりも世代としては、前のモデルではあるものの “ChatGPT3（3.5）” について、公開されている内容に基づいて説明を行う。ここで、このような技術内容について説明する意図は、主として、以下の 2 つである。

a）第 1 には、本稿における中心的なテーマである「人工知能のための学習データ」について、技術的に、大きな転換をもたらしたものであることを確認して、そのための知的財産権について検討する前提としたい。

b）第 2 には、ChatGPT の技術的構成を説明することで、少なくとも、『膨大なデータに基づいて、単純に、最も尤もらしい「次に出てきそうな単語」を、選び出しているに過ぎない』との見解とは、技術的には（少なくとも技術応用の局面では）別のステージに踏み込みつつ（あるように見える）ことについて、注意を喚起したいことがある。技術の観点から見るのであれば、現状、多くの議論がされているような、それが人間の有している「意識や意思と比べてどうなのか？」ということよりも、一定程度の技術上の工夫によって、「人間と遜色のない自然な言語によるコミュニケーションのツールが実現されたこと」が第一義的には重要であろう。また、（仮に人間とは異なる過程で処理がされるのだとしても、コンピュータの計算資源さえ利用可能な範囲であれば）「業務や作業・制御に十分に役立つツール」が実現されてきている、という点も重要といえるのであろう。

1）ChatGPT の構成

ChatGPT の概略的な構成とその特徴を図 1 に示す。

(9)　特開 2024-347 号公報明細書

(10)　酒井將行「データの利用と実施行為の観点から見たデータ駆動型人工知能の知的財産保護」別冊パテント 27 号（日本弁理士会中央知的財産研究所 研究報告 52 号『知的財産権のエンフォースメントの新しい地平』）

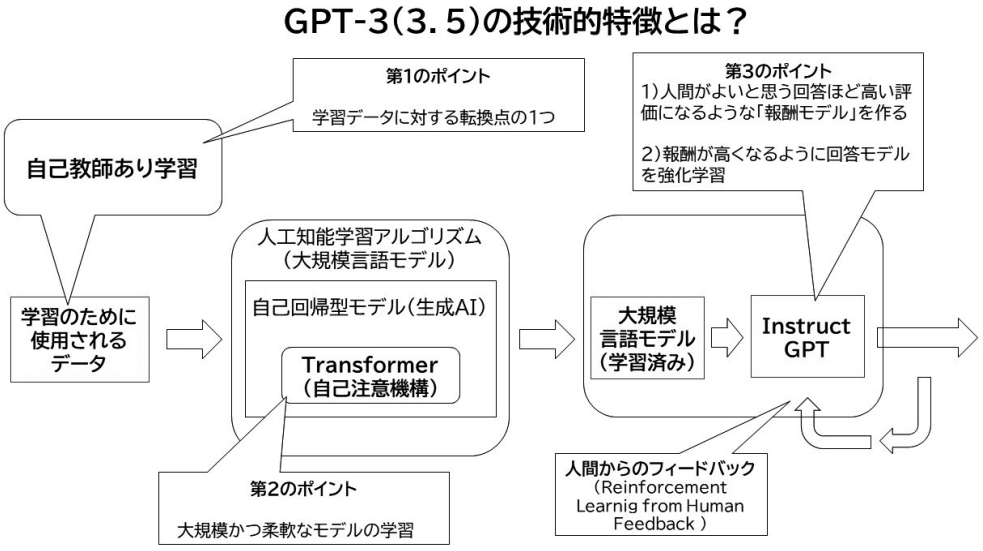

図 1　ChatGPT の構成

以下での説明の前提となる ChatGPT の技術的な特徴は、以下のとおりである。

ⅰ）自己教師あり学習

学習にあたっては、いわゆる「自己教師あり学習」が実行されることで、大量のデータを使った学習において、教師データを事前に付する作業、いわゆる「アノテーション」が不要となった。

ⅱ）トランスフォーマー（Transformer）[(11)]

大規模言語モデルには、「トランスフォーマー（Transformer)」と呼ばれる機構（「自己注意機構」とも呼ばれる）が採用されたことにより、大規模なモデルを効率よく学習させることができる柔軟なモデルが採用された。

ⅲ）対話型インタフェースのための「人間からのフィードバック」の活用

ChatGPT では、人間からのフィードバックによる「強化学習」（教師なし学習の一種：Reinforcement Learning from Human Feedback）により、人間がよいと思う回答ほど高い評価となるような「報酬モデル」を作成して、報酬が高くなるように回答のモデルを強化学習する処理が行われている（Instruct GPT と呼ばれる)。

2）エンコード処理（埋込処理）

まず、ChatGPT では、学習には、インターネット上で収集することが可能な、その言語のテキストが使用される。

詳細は、省略するが、このようなテキストのデータは、各々、トークン（簡単に言うとすると、「単語」と理解できる）と呼ばれるデータに変換される。このようなトークンは、「埋込み（embedding)」という処理により、数値列からなる「ベクトル」に変換される。この場合、意味的に近い単語同士がベクトル空間上で近接するように変換される、といえる[(12)]。

---

(11) 参考までに、「トランスフォーマー」については、Google 社が特許権を取得している。特許 7214783 号
（出願日：2021/5/12　登録日：2023/1/20）
　また、たとえば、LLM については、以下の文献を参照。山田育矢　監修 / 著、鈴木正敏、山田康輔、李凌寒　著『大規模言語モデル入門』（技術評論社、2023 年）

(12) 概念的な説明に過ぎないが、たとえば、「マウス」という単語は、「動物」としての意味と、「パソコンの入力機器」との意味を有する多義語であり、この「意味」を捉えるには、この単語の文字面だけではなく、この単語の含まれるテキストの他の単語を考慮しないといけないことは容易に想像できるであろう。このような考え方は、ある単語の意味は周辺に出現する単語によって表されるという「分布仮説（distributional hypothesis)」と呼ばれる。

そして、このようにして「ベクトル」に変換された後に、各単語のテキスト中の位置の情報を付加するための「位置エンコーディング」と呼ばれる処理が行われる。後述するように、「自己注意機構」では、テキスト内のある位置に存在する「単語」が、他の位置に存在する「他の単語」との関係性を学習するものであるため、単に、「単語が何であるか？」だけではなく、「その単語がテキストの中のどこに位置していたか？」という情報を失うことなく、学習することが必要になるからである。

図２　自然言語処理におけるトークン・単語埋込み

以上のように、「テキスト」→「トークン」→「位置エンコーディング」までが、ChatGPT において、コンピュータが「テキストデータ」を扱うための処理（エンコード処理）である。

3）自己教師あり学習

ChatGPT が現れる以前から、Google 社は、BERT と呼ばれる「言語モデル」の開発を行っていた。このBERT や、ChatGPT での「自己教師あり学習」について、図 3 に示す。

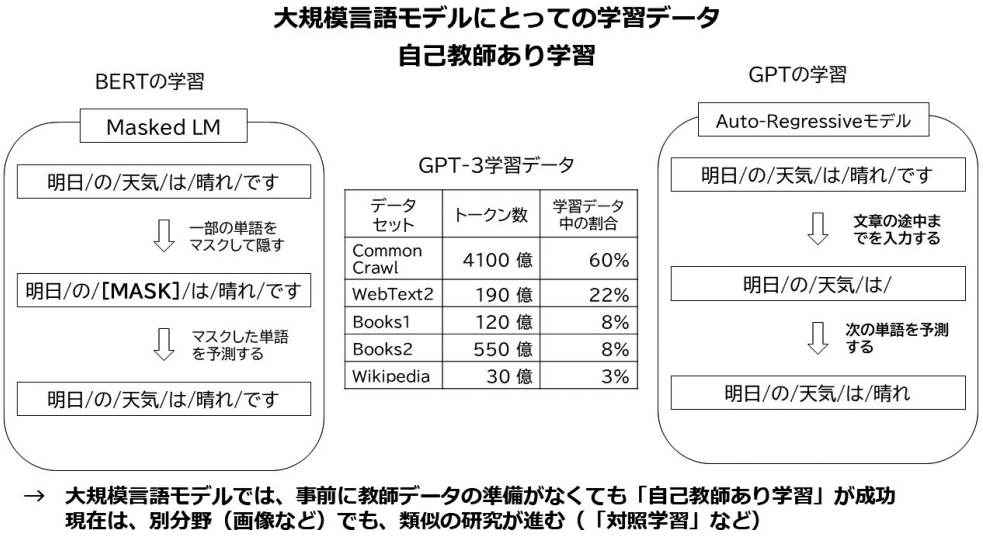

図３　自己教師あり学習

どちらのモデルでも、学習には、インターネット上からクローリングしたテキストデータや、Wikipediaなどに存在するテキストデータが利用される。

たとえば、このようにして準備されたテキストが、「明日の天気は晴れです」というデータであるとして、トークンとしては、「明日／の／天気／は／晴れ／です」と分割されているものとする。

このとき、BERTにおいては、このうちの1つの単語、たとえば、「天気」を隠した（マスクした）「明日／の／［MASK］／は／晴れ／です」を学習データとしてモデルに入力して学習処理を実行する。このときに、その前後のトークンから、マスクされた単語を予測するという学習を実行するのである。

一方で、ChatGPTにおいては、「明日の天気は晴れです」というデータの場合、途中までのデータ、たとえば、「明日／の／天気／は／」を学習データとしてモデルに入力して学習処理を実行する。このデータの次に来る単語である「晴れ」を予測するように学習を行うことになる。

いずれにしても、「明日の天気は晴れです」とのデータさえあれば、正解データを予め準備しておく必要はないことになる。このような学習を「自己教師あり学習」と呼ぶ。

なお、「自己教師あり学習」については、このような「テキストデータ」に対する学習だけでなく、たとえば、「画像データ」に対する生成モデルの学習にも利用されつつある(13)。

4）自己注意機構

「トランスフォーマー（Transformer）」は何をしているのか？

大規模言語モデルの最近の発展の大きな理由の1つは、モデルに、トランスフォーマーと呼ばれる機構を導入したことである。トランスフォーマーについては、「自己注意機構（Self-attention mechanism）」と呼ばれる技術が用いられている(14)。

たとえば、以下のような2つの文があったとする(15)。

I swam across the river to get to the other bank.

I walked across the road to get cash from the bank.

両者は、単語の並びとしては、極めて似通った文であるといえる。

ただし、文末の“bank”の意味は、前者は、「岸」であるのに対して、後者では、「銀行」になる。つまり、2つの文において、最後の単語として何が来るべきなのかは、あくまで、それぞれの文の単語の並びの他の単語によってもたらされるコンテキスト（文脈）に依存している。前者では、“bank”の意味を決定づけている単語は、”river”と”swam”であるのに対して、後者は、特に、“cash”が重要であるといえるであろう。

この意味で、トランスフォーマーは、文の中で、ある単語（上記では、“bank”）にとって、他の単語よりも、どの単語に「注意を向けるべきか」を学習しているという意味で、「（自己）注意機構」と呼ばれていることになる。そして、少なくとも、『膨大なデータに基づいて、単純に、最も尤もらしい「次に出てきそうな単語」を、選び出しているに過ぎない』との見解については、そのような「学習」が行われているとすると、多数の似たような文を探索して、その文の集まりの中で、機械的に「単語」の一致の程度から、次に来る単語が、「何である可能性が高いのか？」を選択する、というような単純な演算処理ではないことも、容易に想像がつくであろう。

言い換えれば、トランスフォーマーにより、入力された「テキスト全体」の「文脈（単語全体を見た“状態”）」を（ある意味で）「解釈」して、次に来るべき単語を予測しているのであって、入力された「テキスト」

(13)「連続する動画データを自己教師ありで学習　LLMに新手法登場」
2023年10月06日 https://xtrend.nikkei.com/atcl/contents/technology/00007/00063/
(14) Ashish Vaswani, et.al. “Attention Is All You Need”, https://arxiv.org/abs/1706.03762
(15) Christopher M. Bishop, Hugh Bishop “Deep Learning: Foundations and Concepts” Springer; 1st ed. 2024版 (2023/11/2), p.359

を構成する「単語」を、すべて同じ重要度として、その「単語の物理的な並び」から、学習データ中で、その「単語の物理的な並び」に最も近似するテキストを元に、次に来る単語を選択している、というような処理では全くないことになる。

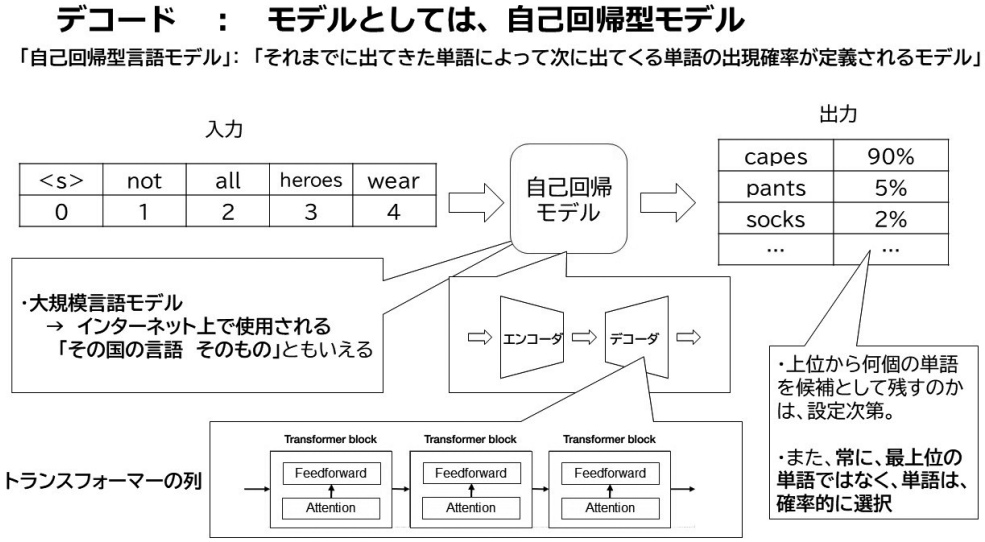

図４　ChatGPT の自己回帰モデル

入力されたテキストデータに対して、次の単語を予測して出力するモデルは、「自己回帰モデル」という範疇（はんちゅう）に入る。ChatGPT では、自己回帰モデルとして、入力をエンコーダでエンコードした後、エンコーダの出力をデコードするデコーダとの組み合わせとなっており、デコーダ側に、トランスフォーマーのブロックが複数個、縦列につながって処理を行う構成となっている。デコーダの出力としては、「次に来る単語」について、複数の候補のそれぞれについて、確率が出力される。

また、トランスフォーマーの機能ブロックの構成の概要を図５に示す。

トランスフォーマーでは、入力に対して、クエリ埋込みと、キー埋込みと、バリュー埋込みとをそれぞれ計算するためのニューラルネットワーク（それぞれに対応する重み行列）が、それぞれ独立に学習される。そして、クエリ埋込みと、キー埋込みとから、i 番目のトークンから見た j 番目のトークンとの関連性（関連性スコア sij）が計算される。このような関連性スコアを要素とする行列は、「注意行列」とも呼ばれる(16)。そして、クエリ埋込みと、キー埋込みと、バリュー埋込みが、同一の入力から計算されている点で、「自己注意機構」と呼ばれていることになる。

(16) トランスフォーマーの最初の論文の題は、前述のとおり、“Attention Is All You Need” であった。最近、このような自己注意機構が、人間の脳の記憶で重要な役割を担う「海馬」の「連想記憶」についてのニューラルネットワークモデルとして以前から提案されてきた “Hopfield Network” と特定の条件下で、同等であることが証明されたとのことである（Hubert Ramsauer, et. al. “Hopfield Networks is All You Need”, https://arxiv.org/abs/2008.02217 ）。筆者は、もちろん、このような最新の知見について、学術的な意味でコメントできる立場ではない。ただし、少なくとも、トランスフォーマーで行われている処理が、「連想記憶」の処理に類似するということが正しいのであれば、画像認識などで利用されてきた「畳込みニューラルネットワーク」などとは、違うレベルの処理が実現するようになってきたといえるのではないかと推察する。人間の脳については、海馬の C3 領域という部分が、記憶情報の蓄積・想起、特に、ある部分から全体を連想する能力であるパターン・コンプリーションにとって極めて重要との認識があるようである。そして、工学的な意味では、「連想記憶とは、情報をパターンとして分散的に表現した上で貯蔵し、部分的な情報を手がかり（キー）として必要な情報を読み出す方式」とされる。とすると、ChatGPT が、N 個の単語列から「想起される（N + 1）番目の単語」を予測しているという処理が、「連想記憶」のモデルと類似しているというのは、（素人ながら）大変、興味深い。この意味で、前述した『膨大なデータに基づいて、単純に、最も尤もらしい「次に出てきそうな単語」を、選び出しているに過ぎない』との見解は、ChatGPT についての正しい理解とは言えないようにも思われる。今後、さらに研究が進めば、人間の脳の働きとの関係でも、そして、コンピュータの処理の観点からも、より一層進んだ知見が得られるのではないかと、それだけでも、興味の湧くところである。

関連性スコアを用いて計算される重みで、バリュー埋込みの重み付き和を計算したものが、出力埋込みとなって、次段のトランスフォーマーブロックの入力となる。

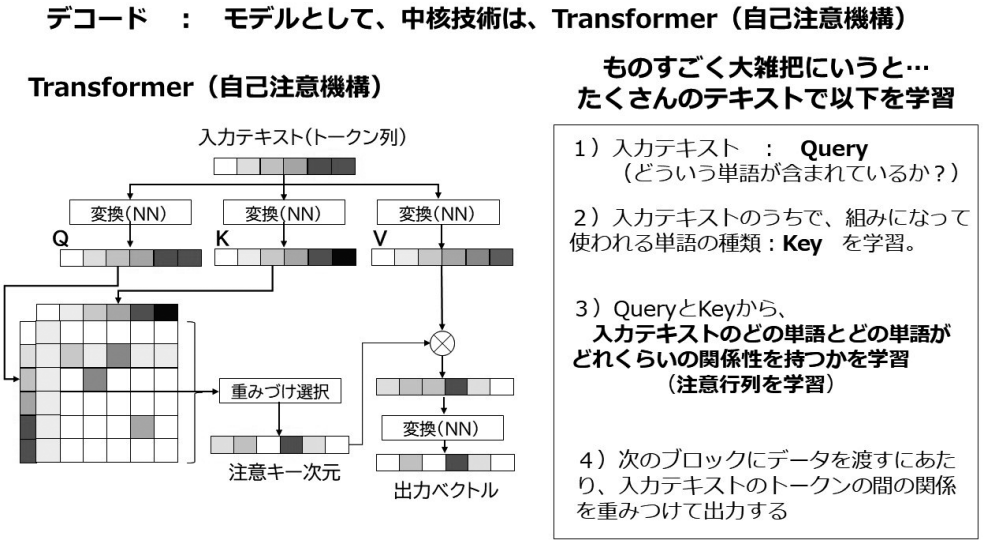

図５　トランスフォーマーの構成

トランスフォーマーの各ブロックでは、入力のトークンの埋込み列を、単語同士の関連性（重要度）を加味しながら、新しい埋込み列へ変換（transform）していく。単語に文脈の意義を持たせるという意味では、ブロックの処理を複数段経ることで「文脈化単語埋込み（contextualized word embedding)」を深化させているともいえる。つまり、入力テキストの周辺の文脈を加味して動的に単語埋込みが計算されていく。

5）InstructGPT

ChatGPT 以前にも、「大規模言語モデル」によるサービスは、いくつか試みられたものの、ユーザからの質問に対して、ある場合に、不適切な回答がされることが問題となり、サービス提供が中止されるなどの事態となった。

ChatGPT では、完全とまではいえないものの、この点について、大きな改善がされたことも技術的には、大きな進展といえる。このような技術は、上述したような「RLHF:Reinforcement Learning from Human Feedback」によるものとされている。

図 6 では、GPT3 から、教師あり学習によるファインチューニングの後に、RLHF による学習で「ユーザがよいと判断する回答」を行うような「強化学習」（一種の再学習）が実行される。

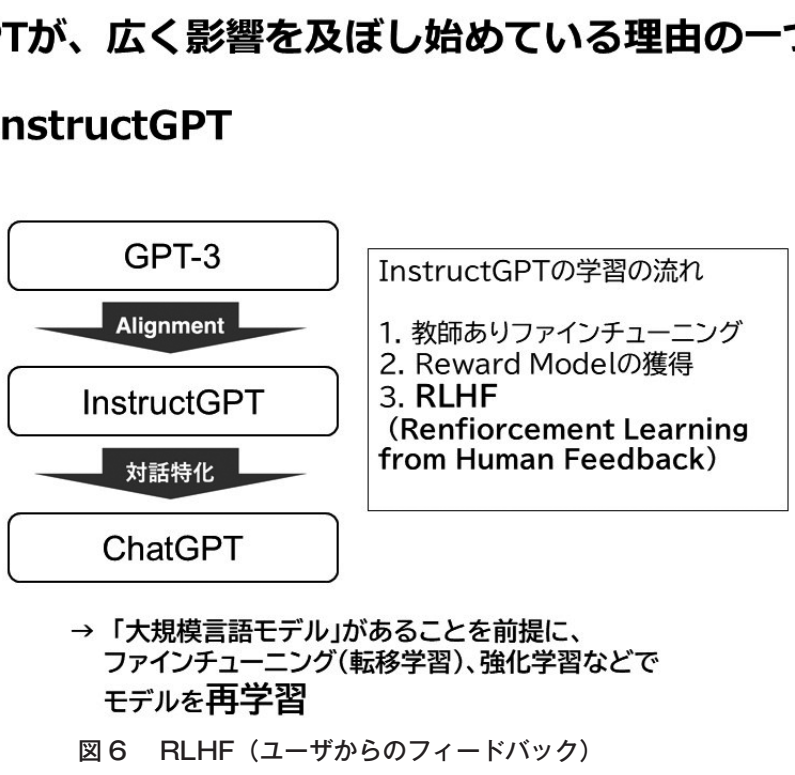

図６　RLHF（ユーザからのフィードバック）

（2）大規模言語モデル以外の大規模基盤モデル

なお、以上の説明では、大量のデータを用いて汎用的に使用可能な「自然言語処理のためのモデル」について、「大規模言語モデル（LLM）」として紹介した。ただし、このような汎用的な大規模な基盤モデルは、画像データに関しても、さまざまに開発が進んでいる。むしろ、コンテンツ生成のための大規模モデルとしては、画像データに対するものの方が、インパクトを与えているといえるかもしれない。

そのような例としては、たとえば、「Stable Diffusion 技術」が有名であろう。これは、画像データのデコーダとエンコーダを組み合わせることで、画像データの生成を実現するものである。この意味では、最も基本的な構成は、「大規模言語モデル（LLM）」と共通しているといえる。ただし、エンコーディングとデコーディングに使用される技術は、いわゆる「拡散モデル」という手法になる。

「Stable Diffusion 技術」自体は、2022 年に公開されたディープラーニング（深層学習）の text－to－image モデルである。これは、ミュンヘン大学の CompVis グループが開発した潜在拡散モデルと呼ばれる技術を用いている。

「Stable Diffusion」のコードとウェイトは一般に公開されており、少なくとも 8GB の VRAM を持つ GPU を搭載したほとんどの消費者向けハードウェアで実行可能とされている。

この技術を用いたビジネスを展開している「Stability AI 社のビジネスモデル」としては、コアとなる基盤モデルは、オープンになっていることから、顧客ごとに、ファインチューニングしたモデルを有償で提供するというものを含んでいるとのことである。「ファインチューニング」とは、顧客の専用の用途のために、再学習データを用いてモデルを再学習することに相当する。

このように、「大規模な基盤モデル」が商用利用できることを前提に、顧客ごとに、自然言語処理を応用した技術に、「ファインチューニング」を行うことは、一般に行われるようになっている。

ここで、「大規模な基盤モデル」自体は、クラウド上で運用されるとして、他の顧客のデータとの間で、隔離した運用を可能とするサービスが、クラウドを運用するサービサーからも提供されている。

（3）技術上のトレンド

そもそも、「生成 AI」が世の話題になり始めたころは、「生成 AI」とは、文章や画像などのコンテンツを生成するもの（Generative AI）であって、「分類」や「予測」をするものというような「それ以前の AI」とは、実行するタスク自体が異なる、というような認識が多かったのではないかと思う。

ただし、「生成モデル」という用語は、「生成 AI」が、これだけ流行する前から、人工知能の世界では使われてきた用語である。

たとえば、多数のデータを、グループＡとグループＢという２つのグループ（クラス）に「分類」するというタスクを考える。

この場合、まずは、事前に、多数のデータが準備されているものとして、このデータを特徴づける「特徴量」の選択が行われる（深層学習では、この特徴量の抽出自体をニューラルネットワークが実行する）。続いて、２つのクラスに分ける場合に、最も適切と想定される「分類手続き」が決定される。そして、「分類」のタスクでは、学習に利用したデータ以外の「ある特定のデータ」を入力としてモデルに与えたときに、いずれのグループ（クラス）に属するのかについての判定が出力される。

このとき、「分類手続き」としては、「事前のデータ（学習データ）」について、２つのクラスに分けるのに、最も適していると判断される境界（線、または面）を設定して、上記「ある特定のデータ」が、この境界のどちら側に属しているかに応じて判定するというものが考えられる。

一方で、「分類手続き」としては、他に、データを生成する（出力する）ような「確率モデル（＝生成モデル）」を想定して、このようなモデルが、グループＡとグループＢのデータを、最も再現性よく生成するようにモデルを決めるという手法も考えられる。たとえば、生成モデルとして、「ガウス分布のモデル」を考えるなら、グループＡについてのモデルは、ある「平均１と分散１」を有しているのに対して、グループＢについてのモデルは、ある「平均２と分散２」を有するとすることが、学習データを最もよく再現するように、これらのパラメータを設定する。そうすると、上記「ある特定のデータ」が、グループＡのモデルから生成された場合と、グループＢのモデルから生成されたとする場合の確率を比較して、確率の高い方のグループに属すると判定する、というような手続きを行うことができる。後者の場合は、前者とは異なり、「生成モデル」という「確率的にデータを生み出すモデル」を想定することとなる。

そして、現在の「生成 AI」は、このようなテキストや画像を「生成モデル」が学習によって得られていることを前提に、さまざまなコンテンツの生成を行っていることになる。

逆に言えば、「生成モデル」は、コンテンツの生成というタスクのためだけに存在するものではなく、以下に説明するように、さまざまなタスクを実行することが可能なものである。この意味では、「生成 AI」との用語を、あまり狭く解釈しすぎると、現在、進行しているさまざまな「生成 AI」の応用の理解を妨げてしまいかねないので注意が必要である。

このような背景を前提に、「生成 AI」に対する現在のところの「技術トレンド」をまとめると以下のようになるであろう。

1）大規模言語モデル作成主体の寡占化

大規模言語モデルを学習して生成できる主体は、世界的に見ても少数に限られる傾向が続くと予想される。学習に要する計算の負荷は、その経済的な負荷とも相まって、モデルの大規模化に伴って、ますます増大していくものと考えられるからである。

一方で、たとえば、「膨大なデータ計算が必要な生成 AI（人工知能）の利用拡大で電力の消費量が急増する。」との予想もあり、状況によっては、現在よりも、40％増えるとの予想まである状況である(17)。現在、

(17)「電力消費、2050 年に 4 割増　生成 AI 普及で想定超す爆食」日本経済新聞　電子版　2024 年 4 月 10 日
https://www.nikkei.com/article/DGXZQOUA29A5J0Z20C24A3000000/

エネルギー政策としては、脱炭素化が指向される中、何らかの対策が必要となっている。

2）大規模言語モデルの圧縮・軽量化

その点で、技術的には、ソフトウェア的にも、ハードウェア的にも、このような「計算の負荷」を低減しようとする試みが、すでに進行している。

まず、ソフトウェアのアルゴリズム上で、モデルを圧縮して、計算の負荷を低減するためには、従来から、主として、以下のような3つの方向性についての検討が精力的に行われてきた。

ⅰ）枝刈り（Pruning）
ⅱ）蒸留（Distillation）
ⅲ）量子化（Quantization）

これらの技術的な内容を簡単に説明すると以下のとおりである。

「枝刈り」:ディープニューラルネットのモデルは、一般に、各層の各ノードが密に結合している。ただし、結合の重みは、ノード間で大きく差があることが通常である。そこで、そのノード間の重みが小さい箇所の接続を削除する、または影響の小さいノードを削除することでパラメータ数を削減する手法のことをPruning（プルーニング：枝刈り）と呼ぶ。

「蒸留」:大きいモデルやアンサンブルモデルを教師モデルとして、その知識を小さいモデル（生徒モデル）の学習に利用する方法を「蒸留」と呼ぶ。これにより、大きいモデルに匹敵する精度を持つ小さいモデルを作ることが期待できる。つまり、一度学習したモデルの知識（予測結果）を別の小さいモデルに継承することになる。

「量子化」:重みなどのパラメータをより小さいビットで表現することで、モデルの軽量化を図る手法を「量子化」と呼ぶ。使用するビットを制限することでネットワークの構造を変えずにメモリ使用量を削減できる。たとえば、32ビット浮動小数点精度（float型）で計算していた学習処理を、8ビットの量子化で行っても、1%程度の性能低下しかないことが報告されたりしている。

これらのうち、大規模言語モデルで、現在、大きな話題となっているのは、「量子化」による計算量の圧縮の可能性である。

米Microsoft社の研究チームが発表した「BitNet」、通称「1bit LLM」と呼ばれる論文[18]が大規模言語モデルに対する「量子化」の大きな可能性を示唆している。「1bit LLM」とは、ネットワーク内の重みを“－1”, “0”, “1”の3値で表現しようとするものである。その意味では、厳密には「1bit」というのは正確ではないが、通称として、「1bit LLM」と呼ばれている。これまでのLLMとは違い、演算が軽くなるにもかかわらず、精度が上がり、そしてこれまで必須だと思われていたGPU（Graphics Processing Unit）が不要で、CPU（Central Processing Unit）でもLLMが動作することを示唆している。すなわち、これまでは、画像処理用の専用プロセッサであるGPUが、ニューラルネットワークの学習処理の演算にも適しているということで、いわば「流用」されることで演算処理を担ってきたというのが実情である。もちろん、人工知能演算の専用プロセッサなどもリリースはされているものの、ハードウェアとしては、GPUの構成を踏襲しつつ、さらに、人工知能演算に特化した構成として実現されてきたという状況であった。

ところが、「1bit LLM」というようなソフトウェアアルゴリズム上の革新により、汎用的な演算装置であるCPUにより、大規模言語モデルを使用できる可能性が出てきていることになる。さらには、使用にとどまらず、「学習処理」についても、CPUのような汎用プロセッサでの実行が可能となる可能性も生じてきている。

ここでは、この「1bit LLM」の技術内容に詳しく言及する余裕はないものの、少なくとも、生成AIによ

(18) Shuming Ma, et. al, “The Era of 1-bit LLMs: All Large Language Models are in 1.58 Bits”, arXiv:2402.17764v1 [cs.CL] 27 Feb 2024, https://arxiv.org/html/2402.17764v1

る人工知能技術のトレンドとして、アルゴリズムの革新の重要性が増加しているという点に、注意が必要であろう。

3）学習データの品質の重要性の再認識

上述した「1bit LLM」のようなモデル圧縮については、実現までに、まだ若干の時間を要する可能性がある。一方で、Microsoft 社は、2024 年 4 月に、大規模言語モデルに比べれば、小規模な「小規模言語モデル（SLM）」で、それまでの大規模なモデルに匹敵する性能を達成して、スマホ上でオフラインでも動作可能なモデル「Phi-3（ファイ 3）」の開発に成功したと発表した(19)。開発のカギは、質の高い学習データ(20)の準備にあるとされる。開発者によれば、「子供が言葉を覚えるに際して、絵本から学ぶ」ように、言語モデルの学習にも、「学習データ」の「質」が決定的であるとの指摘がされている。

また、日本の自動運転の実用化を目指す Turing 社では、大規模言語モデルと画像エンコーダを組み合わせて、トランスフォーマーを使用したモデルに学習させて（マルチモーダル学習）、画像に対する「解釈」をテキストとして生成することを、自動運転に応用しようとしているとのことである。自動運転技術では、これまで、自動車にたくさんの画像センサを設けて、画像処理により自動車の周りの外部環境を認識して、自動車の制御を行おうという、いわゆる “Vision Centralized” 方式が主として研究されてきていた。ただし、ここでの問題は、自動車の周りの環境に生じうるすべての場合の画像データを学習データとして事前に準備するということが、そもそも、困難なことである。そこで、実は、大規模言語モデルが登場してから、画像に対する「解釈」を大規模言語モデルにより実行することで、このような発生頻度が低い外部環境に対しても、AI に自動車の状況を把握させ、自動運転の制御を実行させようとする試みが行われるようになってきている。すなわち、このような構成の特徴は、1 つが「運転判断について説明できること」、もう 1 つが「初めて見る状況や指示に対しても一般常識を使って柔軟に対応できること」とされている(21)。Turing 社の青木 CTO によれば、このような枠組みを想定するとしても、「学習データ」の量以上に、その「質」が極めて重要との指摘がされている。

一方で、2024 年 3 月に、米国の「電子掲示板サービス」を提供する Reddit 社が、ニューヨーク証券取引所（NYSE）に上場した。初日の終値は 50.44 ドルで、売り出し価格（34 ドル）を 48% 上回った。時価総額は発行済み株式ベースで 80 億ドル（約 1.2 兆円）と言われている。

Reddit は、さまざまな興味や話題に基づいたコミュニティが集まる巨大な掲示板サイトである。最新の Reddit 社の統計（2023 年）によると、現在、アメリカのインターネットユーザの 29.3% が Reddit ユーザであるとのことである。掲示板は、非常に多様なトピックが存在するだけでなく、投稿には「アップボート（そう思う／いいね！）」と「ダウンボート（そう思わない）」という形で評価がなされることで、有機的にコミュニティが管理されており、「質の高い情報」が見つかる場所としてユーザの信頼を得ている。

上場時の情報開示資料(22)によれば、Reddit のビジネスモデルは、主に広告収入であるものの、今後展開が期待される 2 つ目のビジネスモデルとして「データライセンス事業」が挙げられている。

---

(19) https://news.microsoft.com/source/features/ai/the-phi-3-small-language-models-with-big-potential/

(20) ここでの「質の高さ」について、開発者の一人の Ronen Eldan 氏は、ご自身の娘さんを寝付かせるための読み聞かせの最中に、以下のような問いを発せられた。“how did she learn this word? How does she know how to connect these words?” その結果、絵本のような精選された単語を使用した「学習」の重要性に思い至ったとのことである。“This breakthrough was enabled by a highly selective approach to training data – which is where children’s books come into play.” https://news.microsoft.com/source/features/ai/the-phi-3-small-language-models-with-big-potential/

(21) https://zenn.dev/turing_motors/articles/353a6e71a1444c

(22) https://s3.documentcloud.org/documents/24438770/reddit-s1-2-22-24pdf.pdf によれば、モデルトレーニングについては以下のとおりの記載がある。

“Model Training. Reddit data is a foundational piece to the construction of current Al technology and many LLMs. We believe that Reddis massive corpus of conversational data and knowledge will continue to play a role in training and improving LLMs. As our content refreshes and grows daily, we expect modes will want to reflect these new ideas and update their training using Reddit data.”

データライセンス事業の2つの柱として展開予定のサービスが「データのAPIアクセス」および「モデルトレーニング」である。特に、『Redditが持つ膨大な会話データと知識コーパスは、現在のAI技術や多くの大規模言語モデル（LLM）の構築において重要な基盤』であり、『コンテンツはタイムリーに更新され、成長するため、日々こうしたデータは進化する。そのため、LLMのトレーニングや改善を目的としたデータの供給に需要が見込まれている。』との言及がされている。

すなわち、（人間の評価を経た）質の高い「データ」自体が、AIの「学習データ」として、ビジネスの対象となり、また、そのようなデータへのアクセスを提供できる会社が、市場からは、高い評価を得る、という状況となっている。

生成AIに関して言えば、OpenAI社だけではなくAnthropic社やStability AI社などの競合も多くのモデルを有償・無償で展開しており、すでにインターネット上にあるデータはトレーニングに使い尽くしてしまったのではないかという意見もあるようである。

したがって、今後は、簡単には、インターネット上でクローリング等によっては収集できないようなデータ、すなわち、市場やユーザのニーズ、流行との関係性が高いデータ（日々更新されるリアルタイムデータなど）こそがモデルの差別化を図る上で大事になってくると考えられる。

4）「コンテンツ生成」以上の応用用途の広がり

以上説明したように、自動運転などの分野では、マルチモーダル学習を利用することで、大規模言語モデルを「画像の解釈」に利用して、そのような解釈に基づく自動運転の制御などが検討され始めている(23)。

また、生成AIを用いることで、現在の画像から、数秒後の画像の状態を予測する、というような応用用途も開発が進められている(24)。

今後、ますます、このような単純な「コンテンツの生成」以上の用途が広がっていくものと予想される。

（4）予想される技術の発展の方向性

1）人工知能演算処理の分散化

これまでは、大規模言語モデルを使用するような生成AIの処理には、演算能力の高い装置が必要で、結果として、大規模言語モデルが稼働するのは、クラウド側の高性能サーバであるとの認識であったところ、少なくとも、近い将来に、大規模言語モデルが、スマートフォンのような端末側で動作するという可能性が出てきた。これは、言い換えれば、生成AIの処理が、「分散化」(25)することを意味する。

少なくとも、2024年4月の時点では、すでに、「ローカルLLM」と呼ばれ、PC内で直接、オフラインで動作し、テキストベースの応答を生成できる技術も公開されている。Meta社によって開発されたLlama2をはじめとするモデルが有名であり、クラウドベースの技術に比べると、性能は落ちるものの、一度インストールすれば、インターネット接続なしで利用可能で、セキュリティの観点などから有利とされている。

なお、ここでいう「分散化」とは、サーバ側での処理と複数の端末側での処理に、演算処理を分散する、というだけでなく、さらには、各ノード（端末）が、それぞれ同等の立場で共同して、学習処理を行う、というような技術も検討されている。この技術は、「分散化」というより、それらが同等であるということを強調して、「分権化」と呼ばれる場合もある。

---

(23) 2024年4月24日　MITテクノロジーレビュー主催　技術講演　チューリング株式会社CTO　青木俊介氏　講演「自動運転2.0 ―生成AIで実現する次世代自律車両―」

(24) 自動運転スタートアップのワービ（Waabi）は、生成AI技術を使った自動運転向けモーション予測システムを発表した。https://www.technologyreview.jp/s/331658/this-self-driving-startup-is-using-generative-ai-to-predict-traffic/

(25) なお、必ずしも生成AIの処理というわけではないものの、たとえば、自動運転技術などでは、リアルタイム性の要求される制御について、自動車に搭載されるエッジデバイス側での「物体検知」などの処理を実行することが検討されている。Manato Hirabayashi et. al.“Vision-Based Sensing Systems for Autonomous Driving:Centralized or Decentralized?”, 2021-4, https://www.jstage.jst.go.jp/article/jrobomech/33/3/33_686/_pdf

2）学習データを準備する技術の重要性の増大

「教師あり学習」に使用される「学習データ」については、「モデルの学習時にモデルに入力されるデータ」と「正解データ」を組として準備しただけでは、著作権法のデータベースとして認められる可能性が低く、知的財産権での保護が難しいことが、指摘されているところである。

この結果、たとえば、「学習（用）データの収集方法」「学習（用）データの生成方法」「学習（用）データの生成装置」などの特許出願もされている。

しかしながら、先には、大規模言語モデルでの学習に、「自己教師あり学習」や「強化学習」が利用されるようになってきたことで、「学習データ」に関連する特許権として、どのようなものを想定し、どのような留意が必要であるのかについては、新たな局面が生じている。

なお、技術的に利用される「データ」の構成の変遷を、「機械学習の学習データ」の構成も含めて記載すると以下のとおりである。

技術的に利用されるデータ構成の変遷

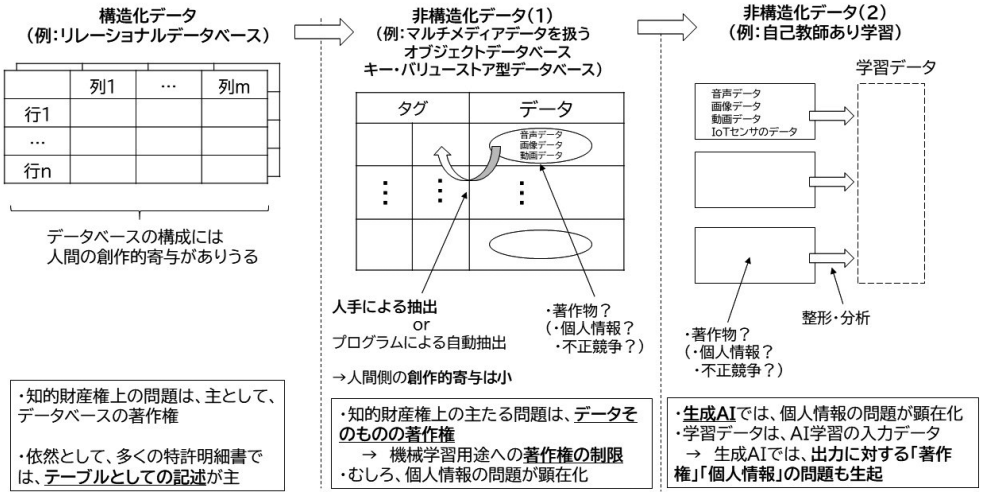

図７　技術的に利用されるデータの構成

今のところ、技術的なデータという場合、「構造化データ」（リレーショナルデータベース）などが、まず、第一に思い浮かべられる。仮に、典型的なテーブル形式ではない「非構造化データ」であっても、従来は、「正解データ」のタグがつけられたデータという意味では、最低限の学習データとしての「構成」が存在したといえる。しかしながら、現在では、「自己教師あり学習」などの登場で、むしろ構成の存在しない「非構造化データ」の重要度が大きくなっている。

3）ソフトウェアとハードウェアとの共進化(26)

先には、大規模言語モデルとして、「1bit LLM」など、ソフトウェア側の革新技術が出てくることにより、専用でないハードウェアであっても、このような大規模基盤モデルの運用が可能となる可能性について言及した。

しかしながら、このことはまた、人工知能学習や人工知能の利用のための「専用ハードウェア」の製造のハードルが下がったことも同時に意味することになる。すなわち、人工知能技術分野において、特に顕在化してきていることは、ソフトウェアの処理として、ある意味、技術の限界に挑戦しているという状況の中から、ソフトウェア技術側にブレークスルーが生じると、それが、ハードウェア側のブレークスルーをもたら

(26) 生物学の世界では、「2 種以上の生物が、寄生や共生、捕食や競争関係などの相互作用を通じて進化すること。虫媒花の花の構造と、受粉昆虫の口器の形態の進化など。相互進化。」とされる（デジタル大辞泉）。

すという関係性が生じつつあるといえる。

一方で、たとえば、大規模なニューラルネットワークに対する学習処理の計算負荷、電力負荷の増大に対処するために、このような演算処理について、GPU などよりも、より特化してこれらの課題を解決することを目指した回路構成のハードウェアとして、コンピュータインメモリ技術、あるいは、メモリインコンピュータ技術などと呼ばれる回路の開発も進められている。

すなわち、人工知能技術では、それまでのハードウェアまたはソフトウェアの制約・制限を超えるような開発が必要となってきていることから、単純に、ソフトウェア技術だけ、または、ハードウェア技術だけの発展では、限界が生じて、両者のまさに「共進化」とも呼ぶべき状況が生じていることになる。

言い換えると、これまでは、たとえば、ハードウェアの大幅な性能向上によって、ソフトウェア側の技術的に考慮するべき点が緩和されて、ソフトウェア側では、むしろ、「何をアプリケーションとして実現するか？」の方に重心が置かれることも多かったように思われる(27)。ところが、ソフトウェアおよびハードウェアが、共に、限界に挑戦していくという環境では、「これは、ソフトウェアの発明」「これは、ハードウェアの発明」というような棲み分けが、ある部分ではなくなり、相互に、相手方の領分に入り込んでいく、ということが積極的に発生することが予想される。

その意味では、特定の人工知能アーキテクチャ（例：所定の構造のニューラルネット）に対して、GPU を劇的に上回る計算効率を実現するような AI チップが数多く生まれてくる「AI チップのカンブリア爆発」の可能性も語られる状況となっている。

この点でも、人工知能技術の登場により、特許権などの知的財産権の枠組みにおける「特許権の対象」「知的財産権の対象」についても、今一度、見直す必要も生じている、あるいは、生じつつあると考える。

4）学習のマルチモーダル化

上述したように、テキストデータと画像データとを同時に入力データとして与えて（異なるモーダルのデータを入力として）トランスフォーマーモデルを学習させることで、「画像データ」を説明するような技術開発も進んでいる。

「自然言語処理」において、「単語の意味」を機械が学習するという観点からも、学習データのマルチモーダル化は、一層進展していくものと予想される。

4．基盤モデルを前提とした人工知能技術に対する知的財産権保護について

「大規模基盤モデル」を用いる人工知能技術について、このような技術分野において、請求項を起案し、特許明細書を作成するにあたっては、これまで主として「ソフトウェア関連発明」ということで、さまざまな点への留意や注意喚起がされてきたところと考える。ただし、以上説明したことから、より留意が必要となる点について、筆者としては、以下のように考える。

(27) かつて、PC に搭載される HDD の容量が、初めて 1GB を超えたとき（1990 年代半ば）のことをはっきり覚えている年代からすると、今や、デスクトップどころか、ノート PC の（HDD ではなく）メモリについて、8GB や 16GB が当たり前となっているのには、隔世の感がある。一方で、ソフトウェア技術においても、いわゆる「オブジェクト指向プログラミング」が開発されるにあたり目指していたものは、「（十分とはいえない容量の）メモリをいかに利用するか？」が大きな動機付けとなっていたことについては、現在の若いエンジニアには、もはやピンと来られない方々が多いようである。メモリ容量の巨大化によって、プログラミングにおいて、メモリのことを、ほぼ、気にする必要がなくなったからであろう。現在の人工知能についての状況とは、まったく異なるとはいえ、これまでも、ハードウェアの進歩とソフトウェアの進歩が、相互に関連しあってきたことは確かである。ただし、これから、「人工知能の時代」にあたっては、それが、より大きく、より広く、より高速に進んでいくであろうと予想される。

大規模基盤モデルを前提とした場合の知的財産についての留意事項
1)「生成 AI」の技術自体は、その応用用途が拡大しており、単に「コンテンツの生成」にとどまらず、広く、予測・分類などのタスクに利用が進んでいる。明細書化においても、技術の内容によっては、その応用が限定的になり過ぎないような記載とする留意が必要である。
2) AI 技術といっても、その基本的なアルゴリズムや構成に関わるような「基礎技術」に関わるものから、「AI 技術の応用技術」まで幅広い。特に、「AI 技術の応用技術」については、前提として「生成 AI 技術」により、「ユーザとのインタフェースが自然言語や図としての表示」という場合が想定される。
3)「基礎技術」については、技術の「共進化」を想定すると、ハードウェア構成に限定され過ぎないようなクレームドラフトおよびソフトウェア構成に向けた従来のクレームドラフトの修正の可能性も検討が必要であろう。たとえば、「アルゴリズム」についての特許出願であれば、これまで、あまり日本では重要視されない傾向があるものの、(ハードウェアおよびソフトウェアの双方を包含しうるという意味で)「方法クレーム」の重要性が高まるものと予想される。また、学習処理を含めて、処理の「分散化」「分権化」が進むことを考慮した明細書・請求項の記載とすることが必要であろう。
4) 一方で、最も応用寄りの技術としては、「AI 技術自体は公知であるとの前提でのビジネスモデル」の特許出願が想定される。ただし、この場合は、「発明該当性」の観点で、少なくとも、将来において検討を要する事態が発生すると予想される。
5) 大規模基盤モデル自体の学習にとどまらず、大規模基盤モデルを利用する「ファインチューニング」などにおいても、「データの品質保証をどうするか？」が重要課題である。
6)「AI 技術」が、人間による「判断」のレベルにより近づいてきたことにより、「データの品質」については、「教師あり学習の正解データ」や「強化学習などの教師なし学習」であっても、「人間の関与」がより大きくなっていくと予想される。「学習データの信頼性」だけでなく、「生成されたモデルの出力」の信頼性を、いかに向上させるか、という点が課題となる。
7)「ファインチューニング」だけでなく、「応答システム」などで利用されている「検索拡張生成（RAG: Retrieval Augmented Generation)」[(28)]など、生成 AI を含む人工知能は、今後、「個別化」の方向に進むと予想される。この場合、「学習データ」の個別化についての検討も重要であろう。
8) 生成 AI については、入力データの「マルチモーダル化」が進むと予想される。「マルチモーダル AI」とは、テキスト・画像・音声・動画など複数の種類のデータを一度に処理できる AI の技術のことをいう。ここでも、「学習データ」を信頼性を含めて、どうしていくかは、課題となる。

上記の各論点をすべて網羅することまではできないが、特に、「データ」の知的財産保護の観点から留意が必要と筆者が考える点を以下に説明する。

（1）データの信頼性

まずは、「学習データ」そのものの信頼性という観点からは、そのようなデータが、「どこで、いつ、どのようにして取得されたもの」であって、取得後から利用時までに、「改ざん」等がされていないことを保証することが必要な局面が想定される。

このような「保証」が、最も強く要求されるのは、「医療系データ」であろう。さらに言えば、たとえば、「薬などの治験」においては、現在は、人間による「監査」により、「データの真実性」が保証されているの

(28) RAG については、たとえば、以下にその応用が詳しい。株式会社日立製作所 Generative AI センター監修『実践　生成 AI の教科書』（リックテレコム、2024 年）

が現状である。

このような状況において、いわゆる「ブロックチェーン技術」を用いることで、人手によらずに、「データの保証」を行おうとしている技術は複数存在する。

たとえば、特許第6245782号には、医薬品などの治験において、医療データという究極の個人情報を保護するために「ブロックチェーン技術」を用いるシステムが、以下のような構成として開示されている。

【請求項1】

医療機関において使用する医療機関端末と、患者が使用する患者端末と、分散型ネットワークにより接続された複数のノード装置とを備えた個人情報保護システムであって、

上記医療機関端末は、

公開鍵および秘密鍵を生成する鍵生成部と、

上記鍵生成部により生成された上記公開鍵を、上記複数のノード装置のうち一のノード装置に提供する公開鍵提供部と、

上記複数のノード装置のうち一のノード装置から暗号化生体情報を取得する生体情報取得部と、

上記生体情報取得部により取得された上記暗号化生体情報を、上記鍵生成部により生成された上記秘密鍵によって復号化する復号処理部とを備え、

上記複数のノード装置は、

上記医療機関端末から提供された上記公開鍵を上記複数のノード装置の全体で共有するための合意形成処理を行う第1のコンセンサス処理部と、

上記第1のコンセンサス処理部により合意形成された場合にのみ、上記複数のノード装置がそれぞれ備えるデータ記憶部に上記公開鍵を記憶させる公開鍵記憶制御部と、

上記患者端末から提供された上記暗号化生体情報を上記複数のノード装置の全体で共有するための合意形成処理を行う第2のコンセンサス処理部と、

上記第2のコンセンサス処理部により合意形成された場合にのみ、上記複数のノード装置がそれぞれ備える上記データ記憶部に上記暗号化生体情報を記憶させる生体情報記憶制御部とを備え、

上記患者端末は、

上記複数のノード装置に記憶されている上記公開鍵を一のノード装置から取得する公開鍵取得部と、

上記公開鍵取得部により取得された上記公開鍵によって、患者の生体情報または当該生体情報に付加される個人を特定可能な情報を暗号化することにより、上記暗号化生体情報を生成する暗号処理部と、

上記暗号処理部により生成された上記暗号化生体情報を、上記複数のノード装置のうち一のノード装置に提供する生体情報提供部とを備えた

ことを特徴とする個人情報保護システム。

ただし、これは「データ」の保護を前提としたシステムであって、データそのものの信頼性を独立に保証しようというものとまではいえない。

もっとも、「データそのものの信頼性の保証」を「ブロックチェーン技術」により保証しようとする試み(29)は、あるものの、すべての「学習データ」を、このような技術で保護するというのは、システムの負荷の観点からも現実的とはいえない、と考えられる。この点で、別の観点からの保証・保護も想定する必要がある。

(29) 仮想通貨の投資家支援プラットフォームなどを提供する株式会社クリプタクトは2019年7月29日、国立大学法人東京大学・東京大学医学部附属病院脳神経外科の金太一助教授らのプロジェクト「ICT活用による医療画像データ流通システムの構築」に参画することを発表した。https://crypto.watch.impress.co.jp/docs/news/1198855.html

（2）学習データの生成とモデル出力における信頼性

一方で、生成された「AI モデル」からの出力に対する信頼性については、ChatGPT では、「RLHF」と呼ばれるような「人間の評価をフィードバックすることによる学習」が使用されていることを一例として説明した。

あるいは、「ファインチューニング」などでは、依然として、「正解データのアノテーションがされた教師あり学習」の重要性が存在している。

ただし、このような「正解データのアノテーション」自体も、最終的には「人による判断」に頼らざるを得ないのが現状である。人工知能の出力が、人間の生活により密接に関連してくるレベルとなったことにより、アノテーション済みの学習データ作成の負担が増大している。しかも、何を「正解とするか」も、人間によって見解が分かれる場合も現れているのが現状といえる。

このような状況に対する技術として、いわゆる「Human－in－the－Loop 機械学習によるアノテーション」(30)が開発されている。

今後、AI の応用用途の拡大により、重要性が、ますます、拡大する可能性がある技術といえよう。

その構成を以下の図に示す。

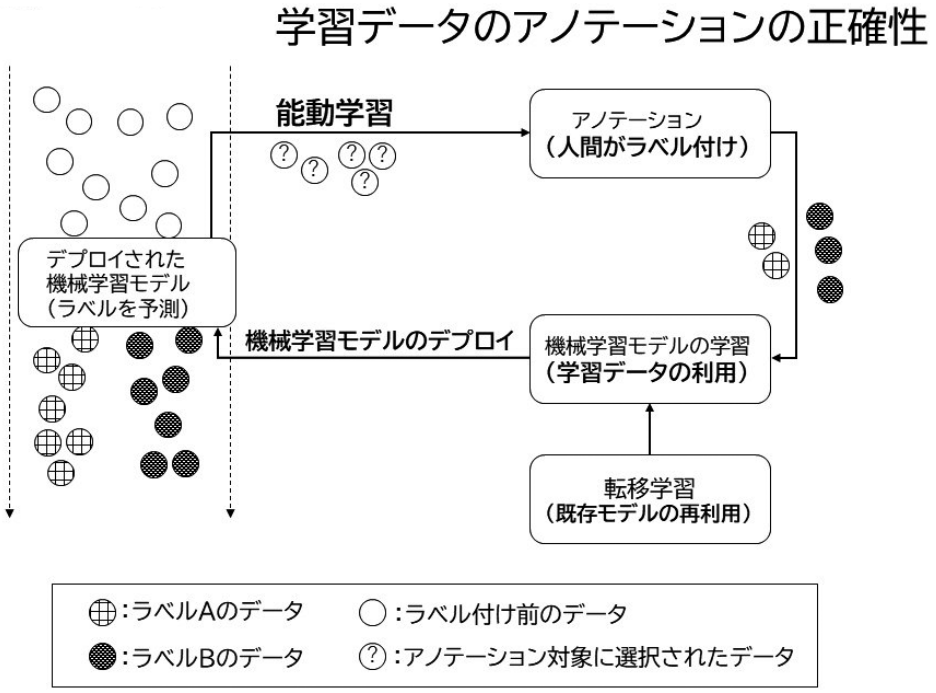

図８　Human－in－the－Loop 機械学習によるアノテーション

すなわち、既存モデルが存在することを前提に、いわゆる「転移学習」により「学習済みモデル」を再学習する場合を考える。このとき、新たに「ラベルがついていないデータ」をモデルに入力して再学習を実行することになる。ただし、無制限に「ラベルがついていないデータ」を入力したのでは、計算負荷が膨大になる。そこで、「能動学習」と呼ばれる方法で、「ラベルがついていないデータ」について人間がアノテーションするデータを選別（スクリーニング）する。一例としては、「能動学習」として、既存モデルが「分類モデル」である場合には、分類境界に一定程度近いデータ（すなわち、分類を学習するには、より重要度が高いデータ）について、人間がアノテーションするように、選別を実行するなどの処理が行われる。

（30）Robert（Munro）Monarch 著（上田隼也、角野為耶、伊藤寛祥訳）『Human-in-the-Loop 機械学習 : 人間参加型 AI のための能動学習とアノテーション』（共立出版、2023 年）

ただし、上記の再学習のループの中には、「人間によるアノテーション」という「人間の判断・行為」が含まれることになるため、「発明該当性」の観点からの留意も必要になるであろう。典型的には、「能動学習」の部分を技術的な特徴として、「人間の判断・行為」をクレームの構成要件からは、除外するなどが考えられる。

そして、このような「再学習のためのデータの選別の処理」は、後述するような「機械学習処理の分散化」が進んだとして、「端末側」「エッジ側」の処理としては、むしろ重要度が増大する可能性がある。仮に、「サーバ側」で再学習をするとしても、「再学習のためのデータの選別の処理」自体は、「端末側」「エッジ側」で実行される可能性がある。この場合、技術的構成としての新規性を有する部分が、「端末側」「エッジ側」に存在することで、「サーバ」がどこに位置するかに依存せずに、このような再学習のための「キーとなる技術」を特許により押さえられる可能性が出てくる。

ただし、後述するように「発明該当性」については、今後、このような例に限らず、より難しい問題が起きてくると予想される。特許権の20年という有効期間を考えると、現時点では、必ずしも要求されていない記載事項についても、先取り的に、十分な検討を行った明細書・請求項の作成をしておく必要があると考える。

（3）発明該当性、実施可能要件、サポート要件と「データ」

いわゆる「データ駆動型人工知能技術」に関して、その権利範囲の解釈において、考慮すべき「実施可能要件との関連性」「サポート要件との関連性」が、これまでも、指摘されてきているところである(31)。

ところで、上述のとおり、『「AI技術」が、人間による「判断」のレベルにより近づく』ことに伴い、これらの問題は、より検討が必要になるものと考える。

データ駆動型人工知能に対する「データの重要性」については、さまざまな指摘がされてきているところであるものの、本稿では、「発明該当性」という別の観点から、この問題を検討したい。その目的は、以下のとおりである。

a)「ビジネスモデル」に対する特許出願に関しては、すでに、その「発明該当性」について多くの議論となってきている。「AI技術自体は公知であるとの前提でのビジネスモデル」の特許出願では、その問題がより顕在化するものと考えられる。

b)「AI技術自体は公知である」、すなわち、「AI技術のアルゴリズム」や「それを実現するためのハードウェア」自体が公知の場合に、特許権の付与を認めるとすると、明細書の開示要件の観点からは、再考が必要な場合が出てくると考えられる。

以下、ビジネスモデルそのものに対して特許権が付与されたものとして、種々の検討がされてきている事案である、いわゆる「いきなりステーキ事件」(32)を例にとって、検討することとする。

本件は、名称を「ステーキの提供システム」とする発明が、特許法2条1項の「発明」に該当するかが争われた、特許取消決定取消請求事件である(33)。

異議申立て時に訂正された本件の請求項1（以下、「本件特許発明1」）は、以下のとおりである。

---

(31) 酒井將行「AI・IoT 技術によるビジネスモデルに対する知的財産権―特許権による保護のためのクレームと明細書」別冊パテント20号（日本弁理士会中央知的財産研究所 研究報告45号『特許クレーム解釈と記載要件』）、前掲注1、前掲注10

(32)「ステーキの提供システム事件」 知財高判平成30年10月17日平成29年（行ケ）第10232号

(33) 本件に対する判例評釈としては、たとえば、以下のものがある。
上羽秀敏 「判批」知財管理69巻9号1272～1285頁（2019）
新藤圭介 「判批」知的財産法政策学研究68巻199～251頁（2023）

本件特許発明１
A　お客様を立食形式のテーブルに案内するステップと、お客様からステーキの量を伺うステップと、伺ったステーキの量を肉のブロックからカットするステップと、カットした肉を焼くステップと、焼いた肉をお客様のテーブルまで運ぶステップとを含むステーキの提供方法を実施するステーキの提供システムであって、
B　上記お客様を案内したテーブル番号が記載された札と、
C　上記お客様の要望に応じてカットした肉を計量する計量機と、
D　上記お客様の要望に応じてカットした肉を他のお客様のものと区別する印しとを備え、
E　上記計量機が計量した肉の量と上記札に記載されたテーブル番号を記載したシールを出力することと、
F　上記印しが上記計量機が出力した肉の量とテーブル番号が記載されたシールであることを特徴とする、
G　ステーキの提供システム。

純粋に、クレームドラフトの観点から、形式のみに着目するとしても、構成要件Ａにおいて、人が実施することになる「ステーキの提供方法」を前提として記載して、その「人が実施する方法」中で利用される「技術的な構成物」がクレームのボディに記載されることとなっている点で、通常みられるようなクレームの構成とは、やや異質な構成といえる。すなわち、通常は、「ビジネスモデルを実現するためのシステムを、システムの技術的な構成要件を列挙する形でドラフトする」という構成でクレームをドラフティングする場合が多いと認識している。

言い換えると、他の多くの「ビジネスモデル特許」といわれるものの「特許請求の範囲」の記載が、「システムを構成する技術的な構成要素」と「それらの技術的な構成要素間の技術的な関連（つながり）」を記載しているのに対して、本件特許発明１のクレームの構成は、大きく異なる。結果として、本件特許発明１は、まさに、形式上も、「ステーキの提供方法」それ自体に向けられた請求項のような外観を呈しているという点で、単に、異議申立ての審理、それに対する特許取消決定取消請求事件の審理の中の判断についての議論だけではなく、広く、ビジネスモデル特許の「発明該当性」について、議論をもたらすこととなった。

３－１）特許庁の異議申立て審理の概要

特許庁での取消理由の概要は以下のとおりである。

「本件特許発明１の技術的意義は、お客様を立食形式のテーブルに案内し、お客様が要望する量のステーキを提供するというステーキの提供方法を採用することにより、お客様に、好みの量のステーキを、安価に提供するという飲食店における店舗運営方法、つまり経済活動それ自体に向けられたものということができる。」

「…してみると、本件特許発明１において、これらの物は、それぞれの物が持っている本来の機能の一つの利用態様が示されているのみであって、これらの物を単に道具として用いることが特定されるに過ぎないから、本件特許発明１の技術的意義は、『札』、『計量機』、『印し』、及び『シール』という物自体に向けられたものということは相当でない。」

「してみると、本件特許発明１の技術的課題、その課題を解決するための技術的手段の構成、及びその構成から導かれる効果等に基づいて検討した本件特許発明１の技術的意義に照らすと、本件特許発明１は、その本質が、経済活動それ自体に向けられたものであり、全体として『自然法則を利用した技術思想の創作』に該当しない。

したがって、本件特許発明１は、特許法第２条第１項に規定する『発明』に該当しない。」

本件のような「発明該当性」が問題となる案件では、日本では、基本的には、「自然法則を利用しているか否か？」が問題となり、「人間の精神活動そのものは、発明に該当しない」ことを当然の前提として、判断がされてきたものと考える。

3－2）知財高裁における審理の概要

特許庁での異議申立てにおいて、特許取消の決定がされたことに対して、原告が、本件決定を不服として、知的財産高等裁判所に本件決定の取消しを求めて提起したのが、本件訴訟（知財高判平成30年10月17日平成29年（行ケ）第10232号）である。

知財高裁の認定を簡単にまとめると以下のとおりである。

「ここで、本件ステーキ提供方法は、『お客様を立食形式のテーブルに案内するステップ』、『お客様からステーキの量を伺うステップ』、『伺ったステーキの量を肉のブロックからカットするステップ』、『カットした肉を焼くステップ』及び『焼いた肉をお客様のテーブルまで運ぶステップ』を含むものである。…

そうすると、本件ステーキ提供方法は、ステーキ店において注文を受けて配膳をするまでに人が実施する手順を特定したものであると認められる。

よって、本件ステーキ提供方法の実施に係る構成（前記ア（イ）①）は、『ステーキの提供システム』として実質的な技術的手段を提供するものであるということはできない。」

「一方、本件計量機等は、『札』、『計量機』及び『シール（印し）』といった特定の物品又は機器（装置）であり、『札』に『お客様を案内したテーブル番号が記載され』、『計量機』が、『上記お客様の要望に応じてカットした肉を計量』し、『計量した肉の量と上記札に記載されたテーブル番号を記載したシールを出力』し、この『シール』を『お客様の要望に応じてカットした肉を他のお客様のものと区別する印し』として用いることにより、お客様の要望に応じてカットした肉が他のお客様の肉と混同することを防止することができるという効果を奏するものである。…

オ 以上によると、本件特許発明1は、ステーキ店において注文を受けて配膳をするまでの人の手順（本件ステーキ提供方法）を要素として含むものの、これにとどまるものではなく、札、計量機及びシール（印し）という特定の物品又は機器（装置）からなる本件計量機等に係る構成を採用し、他のお客様の肉との混同が生じることを防止することにより、本件ステーキ提供方法を実施する際に不可避的に生じる要請を満たして、『お客様に好みの量のステーキを安価に提供する』という本件特許発明1の課題を解決するものであると理解することができる。」

結論においては、全体として「自然法則を利用した技術的思想の創作」に該当すると述べて、発明該当性を肯定した。

3－3）「人の実施する行為」の記載が請求項に含まれる場合の「発明該当性」

以下、本件のような判断手法に基づく場合、「AI技術を応用した発明」（特に、ビジネスモデル特許）において、生じると想定される問題点について、検討する。

3－3－1）ソフトウェア関連発明の「発明該当性」が認められてきた経緯について

以下の議論の前提として、日本において、歴史的に、ソフトウェア関連発明の「発明該当性」が認められるようになってきた経緯を簡単にまとめる。

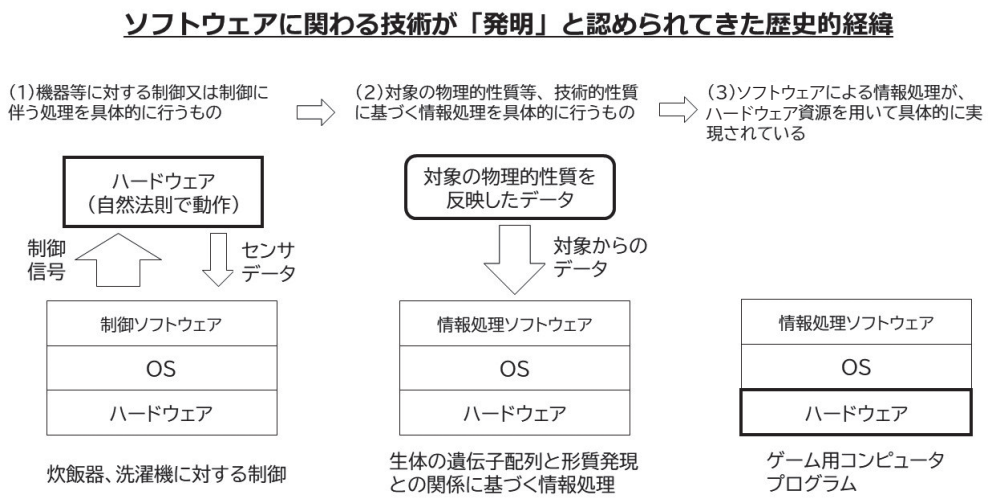

図９　ソフトウェアに関わる技術の「発明該当性」の変遷

いわゆる「ソフトウェア関連発明」が最初に特許の対象として認められたのは、コンピュータによって、その外部にあるハードウェアの動作が制御されるという場合であって、「機器等に対する制御又は制御に伴う処理を具体的に行うもの」といえる場合であった。すなわち、コンピュータそのものの内部での処理に自然法則が利用されているかはさておき、少なくとも、外部のハードウェアは自然法則に従って動作している以上、請求項の記載全体としてみれば、「自然法則を利用しているといえる」との考えに基づくものである。

続いて、「ソフトウェア関連発明」が特許の対象として認められたのは、コンピュータの演算処理の対象となるデータそのものが、「対象の物理的性質等を反映したデータ」である場合であって、「対象の物理的性質等、技術的性質に基づく情報処理を具体的に行うもの」といえる場合であった。すなわち、少なくとも、情報処理の対象が「物理的性質等、技術的性質に基づくもの」である以上、自然法則を利用しているといえるという考えに基づくものである。

これらの２つの類型については、現在では、自然法則の利用性を、特段に検討するまでもなく、発明に該当するものとされている。

そして、(今のところ）最後に「ソフトウェア関連発明」が特許の対象として認められたのは、「ソフトウェアによる情報処理が、ハードウェア資源を用いて具体的に実現されている」といえる場合である。

この結果、仮に、請求項に記載されるものが、「ビジネスモデル（ビジネス方法)」という主題に向けられているとしても、ハードウェア資源との協働関係の記載により発明に該当する場合が生じていることになる。もちろん、この扱いに対しては、本来、発明として特許法で保護するに適さないものであっても、「クレームドラフティング」という形式上の手法によって、「発明該当性」を獲得することには批判も存在するところとなる(34)。

このような批判が生じる原因を、発明の新規性を、請求項に記載される構成のうち、どこに求めるのか？という問題としてみた場合を、以下の図 10 に示す。

図中に記載のとおり、一般の実務家の認識としては、『「ビジネスモデル特許」と言われるものの、「ビジネスモデル」そのものが特許になっているわけではない。』というものであったところ、「いきなりステーキ事件」では、「ビジネスそのもの」に特許が向けられているように見える点が、さまざまな議論を起こしてきた原因の一つと考える。

すなわち、少なくとも、「技術的に新規な部分」が存在するのであれば、それが具体的な技術手段として

(34) 田村善之「特許適格対象の画定における物の本来の機能論の意義」別冊パテント 26 号（日本弁理士会中央知的財産研究所 研究報告 51 号『イノベーション推進に向けた特許の保護対象－更なる研究－』）

記載される限り、そのような請求項の記載が「ビジネス方法を実現するためのシステム」に向けられているのであっても、日本においては、発明該当性については、認められるという方向になる。

問題となるのは、「技術的な構成はすべて公知」であって、「新規なアイディア」といえる部分が、「ビジネスモデル（収益のモデル）」であって、そのような「収益モデル」の構成とハードウェア資源が具体的に記載されているとして、それが果たして、自然法則を利用しているとしてよいのか、という点であるともいえよう。

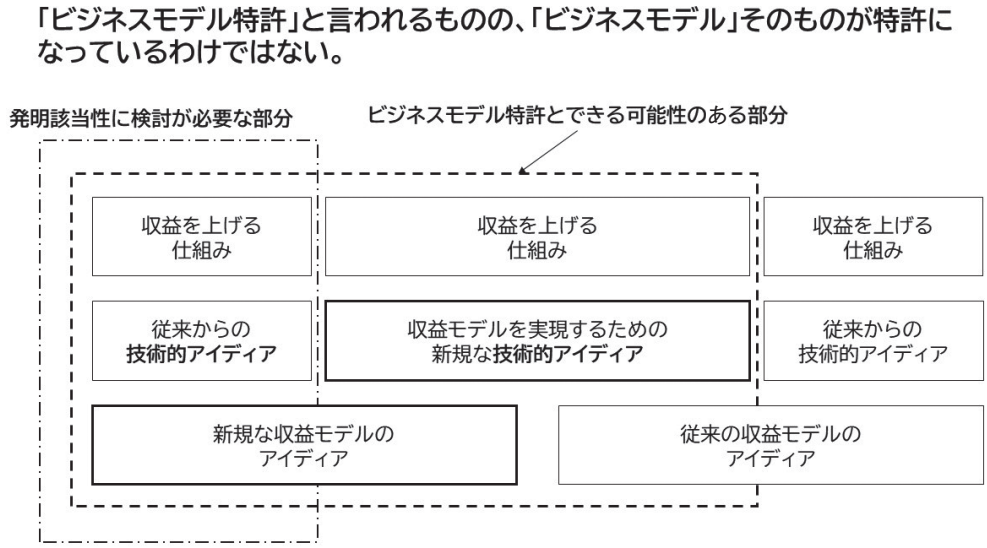

図 10　ビジネスモデル特許の新規性を担う部分

しかも、「いきなりステーキ事件」では、単に、「技術的な構成が公知」であるというだけでなく、そのような構成要件間の関連性（つながり）は、技術的とは必ずしもいえない場合であったといえるであろう。その点において、「自然法則の利用」について、相当に踏み込んだ判断がされたものといえる。

まずは、前提として、本件特許発明 1 の構成を図示すると以下の図 11 のようになるであろう。

「構成要件間の関連性（つながり）」は、技術的要素ではなく、人の判断・意思・行動である。

図 11　本件特許発明 1 の構成

1）化学系の発明の場合

それでは、「構成要件間の関連性（つながり）」が、人の行動である場合をも含みうる発明として、化学系の発明を考えてみよう。

たとえば、「材料Aと材料Bをフラスコに投入」→「フラスコを摂氏60度に保持・攪拌（かくはん）」→「6時間60度に保持」→「フラスコに材料Cを追加投入」→「フラスコを摂氏80度に保持・攪拌」→「2時間80度に保持」という手続きを経て、化合物を合成する手法または合成のためのシステムが請求項に記載されているものとする。

この場合は、仮に、各処理工程を人間が実施したとしても、「自然法則は利用」されており、「発明に該当する」ことについては、異論のないところであろう。

この場合は、各工程は、「人の行動」によるとしても、化合物の合成という発明の本質的な部分について、「人の判断・意思・行動」が入り込む余地はなく、化合物の合成は、あくまで、「自然法則に従って進行するもの」であるからと考えられる。言い換えれば、人間の行動・動作は、単に、化学反応の外的な条件を制御しているに過ぎず、化学反応自体に、人間の行動・動作が、直接関与するということはない、ともいえる。

この点では、本件特許発明1とは、根本的なところで、異なっているといえるであろう。したがって、あくまで、このような例は、「人間の行動・活動」と「発明該当性」との関係性において、「人間の行動・活動」が「発明該当性」に何ら影響を与えないという意味での極端な場合といえる。これに対して、本件特許発明1では、「ビジネスとしての効果を奏する」ための「発明の本質的な部分」について、「人の判断・意思・行動」が必要であって、「人間の精神活動」が、より根本的に関与している、といえるであろう。

2）「ステーキの提供方法」のロボットシステムによる自動化

それでは、以下の図のように、「ステーキの提供方法」を「配膳ロボット」と「調理ロボット」というロボットを用いて、自動化したシステムを考えてみよう。図で示せば、以下のとおりである。

この場合、人間が実施していた「ステーキの提供方法」を、「ロボットによるシステム」とするために、顧客のタブレット端末からの注文を制御サーバが受信して、調理ロボットに調理を指示し、調理ロボットが肉をカットして、計量器で肉を計量する。その後、調理ロボットが肉を焼いて、皿に盛りつけ、配膳ロボットが、制御サーバからの制御に応じて、顧客のテーブルまで焼かれた肉を給仕することになる。そして、このような自動化処理を可能とするために、タブレット端末、制御サーバ、調理ロボット、配膳ロボットなどの間でのデータの授受の仕様や、各ハードウェアの制御のためのソフトウェアプログラムを、人間が設計して、システムとして構築することが必要になる。

結果として、制御サーバ、調理ロボット、配膳ロボットが、それぞれ、公知なものであっても、それらが連携して動作するためのコンピュータプログラムや通信データなどについて、「人間の創作の寄与」があることを前提にしつつ、それらが、ハードウェアと協働して動作する点に、自然法則の利用が認められる、ということになると考えられる。

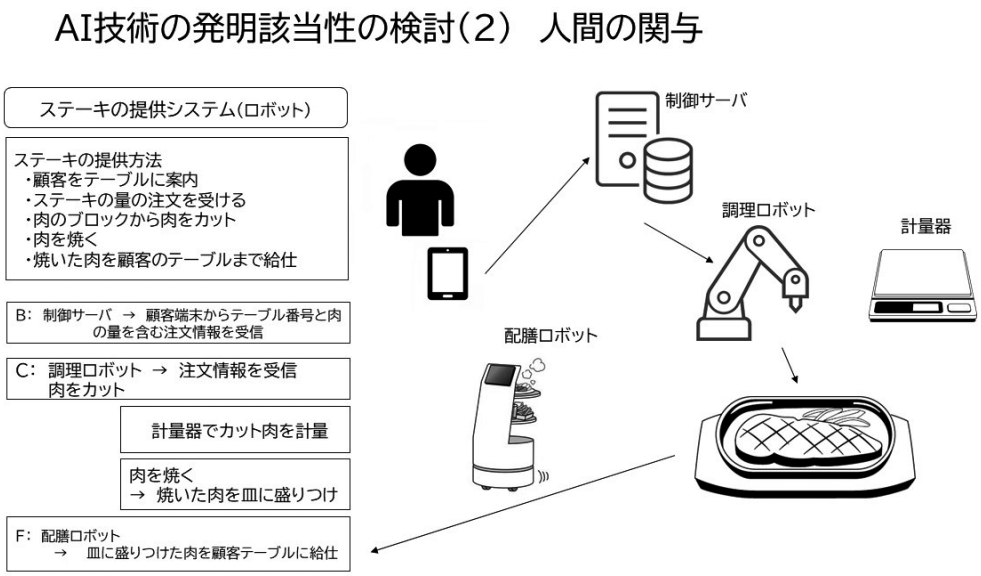

図 12　ロボットによる「ステーキ提供方法」の自動化

3)「ステーキの提供方法」の AI ロボットシステムによる自動化

それでは、AI ロボットの技術が進展することで、「人間が実施しているビジネス方法」を画像データおよび音声データを教師データとして、強化学習により学習する AI ロボットが実施するようになった場合(35)はどうであろうか？以下の図のようになるであろう。

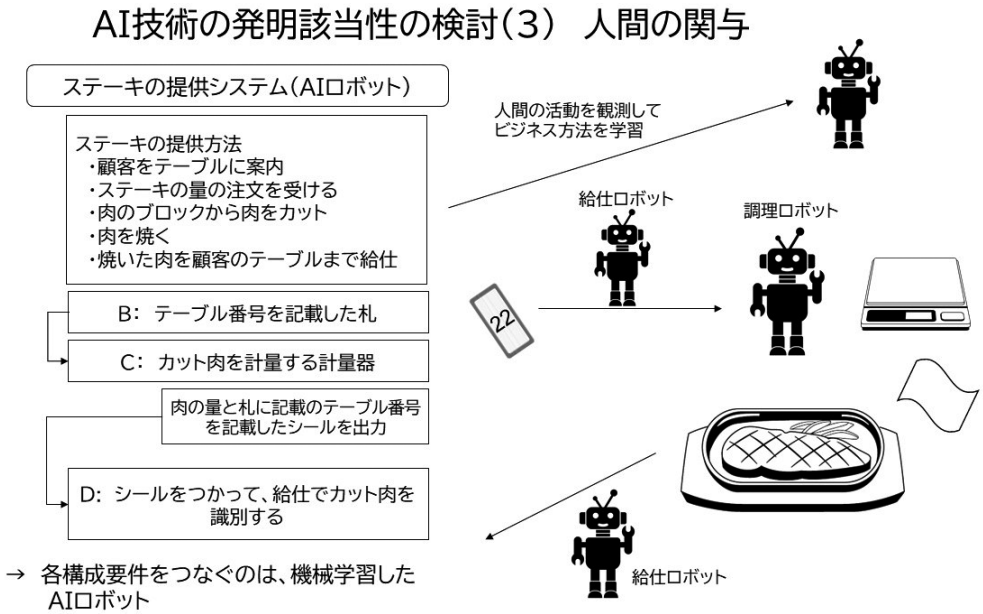

図 13　強化学習で学習する AI ロボットによる「ステーキ提供方法」の自動化

(35) オレゴン州立大学のアラン・ファーン教授（コンピューター科学）らの研究チームは、立つ、歩く、箱を拾う、ある場所から別の場所に移動するといった動作を、人型ロボット「Digit V3（ディジット V3)」に学習させることに成功した。一方、カリフォルニア大学バークレー校の別の研究者グループは、Digit が不慣れな環境でもいろいろな荷物を運びながら転倒せず、歩行する方法を学習させることに重点を置いた研究成果を発表。サイエンス・ロボティクス（Science Robotics）誌に論文として掲載された。
ILIJA RADOSAVOVIC, et. al. “Real-world humanoid locomotion with reinforcement learning”, SCIENCE ROBOTICS, VOL. 9, NO. 89, https://www.science.org/doi/epdf/10.1126/scirobotics.adi9579

すなわち、この場合は、人間が実施していた方法を、AIロボットが機械学習することで、人の判断・動作をそっくりそのまま、ロボットの動作に置き換えたシステムが実現することになり、しかも、「人の判断・意思・行動」自体は、クレームの対象には現れないことになる(36)。

その場合、システムの開発を構想して、実用化されるまでのタイムスケジュールと、典型的な特許出願と審査の流れを図にすれば、以下のようになる。

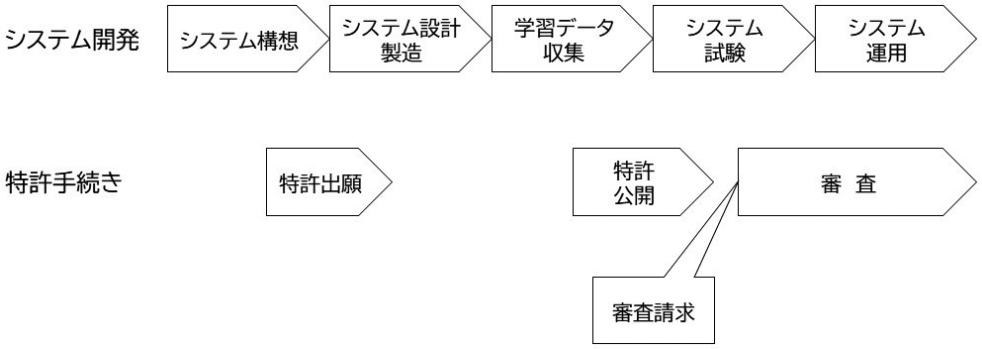

図 14　システムの設計から運用までと特許出願の手続き

この場合、特許出願するのは、システムの構想が出てきて、設計に取り掛かる時点が、一つの候補であろう。ただし、その場合、AIロボット自体は、既存の公知のものを利用するとしても、AIロボットを学習させるための「実環境での学習データ（Real World Data）」については、まだ取得されていない場合がほとんどであろう。

技術水準としての「AIロボット」の完成の度合い次第ではあるものの、少なくとも、特許出願時点における「実施可能要件」の充足性については、一応の課題が残る状態といえる。

ただし、この場合の真の問題は、以下の図に記載のような状態で特許出願がされている、ということである。

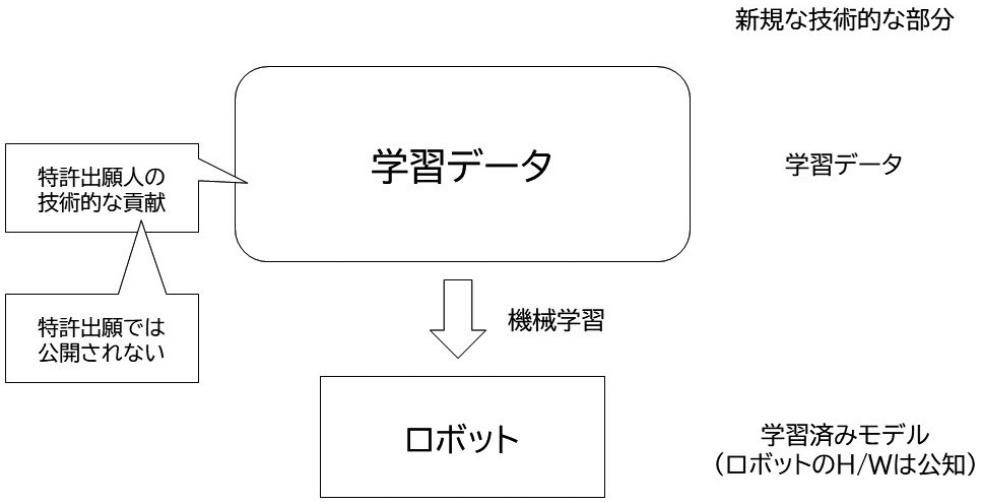

図 15　AIロボットによる自動化における技術的貢献

(36) 開発目標は、2062年とだいぶ先に設定されてはいるものの、米国のスタートアップであるFigure（フィギュア）社は、「すべての人間が家で自分専用のヒューマノイドロボットを持ち、料理や掃除、洗濯といった日常の細かな作業をロボットにやらせるような未来」を目指すということで、2024年2月29日、マイクロソフト社やエヌビディア社、OpenAI社のスタートアップファンド、ジェフ・ベゾスなどからの評価額は26億ドルに達し、6億7500万ドル（約1060億円）を調達した、とのことである。

すなわち、技術的に新規といえるのは、「学習データ」と学習データにより学習させた「学習済みモデル」ということになる。そして、出願時には、上述のとおり、「学習済みモデル」も「学習データ」も、現実には存在しない状態で特許出願がされるという場合が、可能性としては、かなりの割合に上る、と予想される。

もちろん、「学習済みモデル」については、そのような学習を実現できるような（学習前の）「モデルの構造」については、公知という前提ではある。しかしながら、「学習データ」は、出願時点において公知でないばかりか、出願後においても、一切公開されることがないと想定される。

たとえば、「ビジネスモデル（ビジネス方法）自体」が「新規」であるというような「発明」の場合は、「実環境での学習データ（Real World Data）」を取得するということは、とりもなおさず、「ビジネスモデル（ビジネス方法）自体」を公知にしてしまうことになり、しかも、データの取得にはそれなりの期間が必要であることに鑑みれば、出願前から「実環境での学習データ」の取得を開始するということは容易ではないであろう。この場合、第三者も「データの取得」さえ実行すれば、同一のシステムの構築が可能であるとすると、新規性喪失の例外の規定の利用ということも、困難性があると考える。

とすると、そもそも、「発明該当性」の問題以外に、あるいは、それ以前の問題として、技術の核心である「実環境での学習データ」を公開することがない特許出願に対して、「独占排他権たる特許権」を付与することが、法目的である「産業の発展」に、本当に寄与することになるのか？ということが、問題となると考える。

人工知能技術については、学術誌の多くでそれが要求されるのと同様に、「アルゴリズムを実現するプログラムそのもの」と「学習に使用したデータ」の公開(37)を要求するというのが、一つの解決方法ではあるものの、「企業秘密」との関連で、容易なことではないとも考えられる。

いずれにしても、そう遠くない将来においては、少なくとも権利行使の局面で、「AI技術の応用に関する特許権」については、「実施可能要件違反」や「サポート要件違反」が、論点となる事態も予想される。

とすると、以上説明したような「AI技術の応用」に関する特許出願では、現時点では、「アルゴリズムを実現するプログラムそのもの」については、明細書において、少なくとも公知文献などを用いて、技術水準を明らかにすることで、実施可能性を担保する記載とするとともに、「学習データ」については、それをどのようにして、どのような態様のデータとして取得するのか、という点までは、可能な限り、明細書に記載しておくことが望ましいと考える。

「学習処理」をあたかも、各ステップについて、人間が実施するとしてもそうなる、というような記載の仕方（フローチャートや取得するデータの構成が、コンピュータの処理としては不十分な記載）とならないような注意が必要であろう。モデルについても、安易に「ニューラルネットワーク」に対する「ディープラーニングで学習する」というような記載以上に、技術応用の各局面に適した「モデル構成」「学習方法」「学習データの構成」までは記載しておくことが必要と考える。

たとえば、やや話題としてはそれる可能性があるものの、現時点においても、十分に実用化可能なレベルと想定されるような技術内容、たとえば、通信システムであって、サーバと端末とが相互に情報をやり取りしながら、何らかの技術的な効果を奏するような発明であっても、このような具体例の記載が不十分な場合

(37) 2024年8月に、オープンソース・イニシアティブ（Open Source Initiative、略称：OSI）が、「オープンソースAI」の定義を発表した。https://www.technologyreview.jp/s/344206/we-finally-have-a-definition-for-open-source-ai/　そこでは、「オープンソースAIシステムは、許可を得ることなくどのような目的にも使用することが可能であり、研究者が構成要素を検査してそのシステムの仕組みを研究できるようにするべきであるという。また、どのような目的でもそのシステムを修正（出力の変更を含む）することが可能であり、修正の有無にかかわらずあらゆる目的のために他の人とシステムを共有して使用できなければならない。さらに、この基準は、所与のモデルの訓練データ、ソースコード、および重みに関する透明性のレベルを定義することも試みている。」という点でだけでなく、『「スキルを持つ者が同一または類似のデータを用いて実質的に同等のシステムを再現できる」範囲で、訓練データに関する情報を提供することを求めている。』。特許制度での扱いをどうするかという点からも注目すべきであろう。

がある。すなわち、明細書を作成する人間からすると、「サーバと端末」の双方を一望できる立場であるために、これらの間でやり取りされる「データの構成」や、その「データの構成に基づく処理」を、たとえば、サーバ側ならサーバ側だけで実現可能に必ずしも記載していなくても、フローチャートなどは、人間の目から見るだけでは、問題なく処理が進行するように記載できてしまう場合がある。

筆者は、このような問題を、あくまで、たとえとして、「神の視座問題」と呼んだりしている。実施可能性を担保した明細書とするために必要な観点であり、AI技術が進展した際には、より問題が顕在化する可能性がある。

一方で、AI技術が進展した際に、明細書の開示内容が、「実施可能であるのか？」という点で、今後、ますます問題となりそうなものは、上述したように、具体的に「プログラム」や「学習データ」を開示する必要がない場合に、「これこれのデータに基づいて、これこれのことができる」と明細書に書いてあるとしても、それが本当に実現可能なのか検証のしようがない、という問題であろう。「擬人工知能問題」とも呼びうるもので、「生成AI」と記載することで、なんでも実現できそうであるが故に、本当にそれが実施可能であるのかが、不明なままの明細書・クレームとなる恐れがあるといえるであろう。

これも、筆者は、あくまで、たとえとして、「神の手になる機械問題」と呼んではどうかと考えている。すなわち、明細書の開示のみでは、少なくとも特許出願時の「人間」の技術常識では、実現可能であるかが不明な発明でも特許出願の明細書上では、ある程度の尤もらしさをもって記載できてしまう、という問題である。

少なくとも、現在において、「人間にとっては簡単なことのように見えるもの」を、「コンピュータによって実現しようとすると大きな困難がある」という課題は、いくつも存在する(38)。明細書を起案する者だけでなく、それを審査する者、その権利行使の妥当性を判断する者のいずれにとっても、将来、確実に問題となりそうである。

（4）機械学習の「分散処理」に対する特許権の論点について

4－1）分散処理の例

上述したような「エッジ側の処理」という以前に、たとえば、Google社などは、「端末側での学習処理」を実行する分散学習を「連合学習」と呼んで、特許出願をしてきている。

他社においても、この「連合学習」という用語は、使われるようになっているようである。

「連合学習」の特許の請求項としては、たとえば、以下のようなものである。

特許第6923676号　オンデバイス機械学習プラットフォーム

【請求項20】

コンピュータで実行される方法であって、前記方法は、

収集アプリケーションプログラミングインタフェースを介してコンピューティングデバイスによって、前記コンピューティングデバイスに記憶されている複数のアプリケーションのうちの第1のアプリケーションから新たな訓練事例を受け取るステップと、

(38) たとえば、「人工知能技術」の問題ではないものの、1枚の静止画において、その画像内のどこが物体の影であるのか、をどうやって、コンピュータが判断できるのか？が問題となった事案がある（知財高判平成18年10月4日平成17年（行ケ）第10579号審決取消請求事件。この判例自体は、「サポート要件」が問題となったものであるが、同時に、「実施可能要件」の問題であるともいえよう）。人間にとっては、ほぼ無意識のうちに実行している課題ではあるものの、コンピュータにとっては、それが「色の濃い部分」なのか「物体の影の部分」なのかを、1枚の白黒の静止画の画像だけから判断することは容易ではない。実は、画像中で、同じ色の濃さの部分であっても、それが、影であるのか、そうでないのかという認識は、人間が無意識のうちに行っていることであって、ある条件下では、それが逆に錯視を生むことが知られている。有名なものでは、「チェッカーシャドー錯視」がある。人間が無意識に行っていることが、コンピュータにとっては難しいということの一例である。

前記コンピューティングデバイスによって、前記コンピューティングデバイスの集中事例データベースに前記新たな訓練事例を記憶するステップと、

前記コンピューティングデバイスによって、訓練アプリケーションプログラミングインタフェースを介して前記第１のアプリケーションから、前記集中事例データベースに記憶されている前記訓練事例の１つまたは複数に基づいて、前記コンピューティングデバイスに記憶されている第１の機械学習済みモデルを再訓練するようにとの命令を受け取るステップと、

前記命令に応じて、前記コンピューティングデバイスによって、前記第１の機械学習済みモデルを、前記集中事例データベースに記憶されている前記訓練事例の１つまたは複数に基づいて再訓練させるステップと、

前記コンピューティングデバイスによって、クラウドサーバから、予測プランとモデルパラメータを受け取るステップであって、前記予測プランは、モデルを走らせて推論を得るための命令を含む、ステップと、

前記コンピューティングデバイスによって、予測アプリケーションプログラミングインタフェースを介して前記第１のアプリケーションから入力データを受け取るステップと、

前記予測プランと前記モデルパラメータに従って、前記第１の機械学習済みモデルを利用して、前記入力データに基づいて少なくとも１つの推論を生成するステップと、

前記予測アプリケーションプログラミングインタフェースを介して前記第１のアプリケーションに、前記第１の機械学習済みモデルによって生成される前記少なくとも１つの推論を提供するステップとを含む、

方法。

学習データを１つの拠点やサーバに集約することなく機械学習モデルを開発する手法である。個人情報に関わるデータや機密性の高いデータなど外部と共有できないデータを扱うモデルを、プライバシーを確保しながら開発する際に有用である。トレーニングや評価などに使うデータはやり取りせず、トレーニングによって得られたモデルの変更点のみをやり取りする。

ところで、「端末」で分散して再学習して、その結果をサーバで集約するということであれば、基本的には、「サーバ側」と「端末側」についてのクレームを別々に起案することで、特許権の行使の可能性を担保するという、従来から実施されてきたクレームドラフティングを踏襲することができる。

ただし、たとえば、いわゆる「ドワンゴ事件判決」(39)において、サーバが日本国外にある場合でも、特許権の侵害が成立するための要件としては、以下のようなものが挙げられている。これが、ICT 関連発明について、サーバが域外に設置される場合の判断基準といえよう。

・実質的かつ全体的にみて、それが日本国の領域内で行われたと評価し得るものであること。
（この場合は、日本国の特許権の効力を及ぼしても、属地主義には反しないと解される。）
・上記の判断にあたって、考慮される事項。
a）当該提供が日本国の領域外で行われる部分と領域内で行われる部分とに明確かつ容易に区別できるか、
b）当該提供の制御が日本国の領域内で行われているか、
c）当該提供が日本国の領域内に所在する顧客等に向けられたものか、
d）当該提供によって得られる特許発明の効果が日本国の領域内において発現しているか。

（39）知財高判令和４年７月20日平成30年（ネ）第10077号「コメント配信システム」特許権侵害差止等請求控訴事件

そして、人工知能学習処理の分散処理については、特に、「当該提供の制御が日本国の領域内で行われているか」が問題となると考えられる。

さらに言えば、サーバとクライエントというような主従関係ではなく、すべての端末について平等に情報を交換しながら学習処理を行うような構成も検討されている(40)。

このような分散化･分権化にあっては、実施主体が必然的に複数となることが想定される。この観点では、「学習の全体像」のクレームはもちろんであるが、個々の「端末・クライエント」での処理に焦点を当てたクレーム、さらには、学習以前に「端末・クライエント」で使用するデータの収集のための構成に焦点を当てたクレームなども検討の必要があるであろう。

4－2）特許適格性とクレームのフォーマットについて

一例として、大規模言語モデルのハルシネーション（幻覚）に対する対応を発明の解決課題とした、Microsoft 社の米国特許（US 11968088 B1）を示す。

1. An apparatus comprising:

one or more memories storing computer executable instructions; and

one or more processors coupled with the one or more memories and, individually or in combination, configured to:

receive, at an interface between a user and a large language model, a natural language intent for a network configuration;

request the large language model to update the network configuration to an updated network configuration that satisfies the natural language intent in a declarative network configuration language;

verify whether the updated network configuration satisfies a configuration syntax of the declarative network configuration language to detect an error;

request the large language model to update the updated network configuration to correct the error; and

deploy the updated network configuration to a user network.

米国では、「発明該当性（特許適格性）」を満たすために、最低限、上述のように、構成要件として、“one or more memories”“one or more processors” が記載されることが多い。

日本出願の段階で、ここまでのクレームとする必要は必ずしもないので、日本出願では、「…部」「…部」などの機能的構成を構成要件としている請求項の記載もよく見られるところである。ただし、米国を含む外国への出願を前提とするのであれば、少なくとも、明細書には、これらの技術構成についての開示を設けておくことは必要で、多くの実務家が、それを実践しておられるものと考える。

しかも、上述したような分散処理の可能性を想定するのであれば、“one or more” が分散処理を想定した場合も包含していることを、明細書中に、明記しておくことが必要であろう。

5．まとめ

以上、駆け足ながら、「生成 AI フィーバー」とも呼ぶべき状況について、技術的な分析に基づき、不完

(40) 井手剛 「分散分権型環境での機械学習とリスク管理」、2021 年 10 月 31 日、https://ide-research.net/papers/2022_DataScience_Ide.pdf

全ながら、知財保護をどう考えるべきかについて、筆者の考えをまとめた。冒頭にも記載のとおり、ここで議論したことは、たとえば、6か月後には、もう古い議論になってしまっている可能性も懸念されるところではある。しかしながら、何度も強調したように、「生成 AI」も「データ駆動型人工知能」である以上、知的財産保護という観点からは、実は、最も重要であるのは、「学習データ」（特に、Real World Data）であることに、当面、変更はないものと考える。この点で、「学習データ」の保護を、単に特許権によるものだけでなく、いかにしていくかが、知的財産上の最重要課題と考える次第である。

以上