

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

弁理士 酒井 將行

要 約

工業化社会・情報化社会の次に来るべきソサイエティ 5.0 と呼ばれる社会では、いわゆる「データ駆動型人工知能」技術が、その中核の一つを担うものと想定される。ここで、日本の特許制度において、データ駆動型人工知能の知的財産保護という観点で、「学習済みモデル」そのものをプログラムの発明として特許の対象とする、との運用がいち早く採用されている点については、技術開発を行う主体にとっては、大きなメリットと考える。いわゆる「プログラムの特許権」については、侵害摘発の困難性という点から、課題が指摘されることが多いものの、筆者としては、これは、一つには、発明をいかに把握して、クレームドラフトにつなげられるのか、という側面もあるものと考え。しかも、「学習済みモデル」自体が特許適格性を有していることは、クレームドラフトの自由度ないしは保護範囲の拡大という観点から、極めて重要と考える。一方で、「データ駆動型人工知能」のような技術によるサービス提供という観点からは、特許法上の実施行為の概念についての再検討も必要と考える。もっとも、実は、「データ駆動型人工知能」を知的財産権で保護することにより、産業の発達を図る、という観点からは、筆者としては、特に、オープンに学習用データとして利用可能なデータについて、何らかの新たな知的財産権としての保護の必要性を感じており、立法論とはなるものの、その可能性について提案する。

目 次

1. はじめに
 - 1.1 データ駆動型人工知能技術の社会的背景
 - 1.2 データ駆動型人工知能技術の技術的背景
2. データ駆動型人工知能の技術を保護するためのクレームドラフトについて
 - 2.1 「学習済みモデル」の保護とこれをクレームできることの重要性
 - 2.2 学習済みモデルを特許で保護するという点
 - 2.3 データ駆動型人工知能に対する権利行使を考慮したクレームドラフト1
 - 2.4 データ駆動型人工知能に対する権利行使を考慮したクレームドラフト2
3. 学習済みモデルの特許による保護を前提とした実施行為の把握に対する一提案
4. 「学習用データ」の保護
 - 4.1 「データ」と知的財産としての保護
 - 4.2 データの特性、データの階層性
 - 4.3 既存の知的財産権によるデータについての保護
 - 4.4 データ保護のための新たな知的財産の必要性について
5. まとめ

1. はじめに

1.1 データ駆動型人工知能技術の社会的背景

最近、各所で、ソサイエティ 5.0 という用語が言及されるのを見かけられる方も多いのではないかと思います。ソサイエティ 3.0 と呼ばれていたのが工業化社会で、今のソサイエティ 4.0 が情報化社会で、次の時代はソサイエティ 5.0、いやもうソサイエティ 5.0 になっているという意見もあるかと思うが、要は、現

実世界であるフィジカルの世界と、コンピュータの世界であるサイバーの世界が融合したような社会が、次の社会としてやってくる、ということがいわれている。

図1に示すように、工業化社会のころは「物を作って売る」ということがビジネスの一番中心だったものが、情報化社会になったときには、「コトを売る」ビジネスに変容したという表現がされることが多い。さらにソサイエティ 5.0 への言及の中では、「コト」からさらに進んで、「トキを売る」ビジネスに変容していく、というような表現がされる場合もある。

「コト」と「トキ」の違いは、知的財産としての扱いという観点からは、筆者にとっては、「リアルタイム性」「臨場感」という以上のことは、必ずしも明確ではないものの、一つははっきり言えることは、「物を作って売る」というビジネスから、「サービスを提供する」ということに、ビジネスの重心が移ってきているということであろう。

それに伴い、技術の側では、工業化社会は、「物」を作る技術が重要ということであったところ、情報化社会になったことで「ソフトウェア」というものの比重がかなり高くなって、ソサイエティ 5.0 では、さらに、「人工知能 (AI: Artificial Intelligence)」の重要性が高いとの指摘も多い。

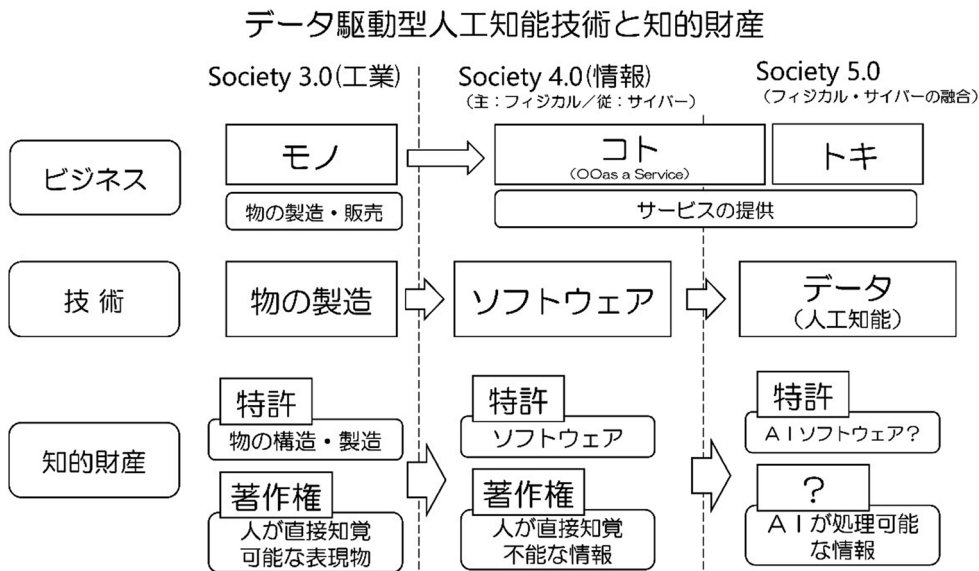


図1 ソサイエティ 3.0 から 5.0 へ 技術と知的財産

以上の経緯から、知的財産の世界では、工業化社会においては、当然、特許は物の構造とか製造方法とかに対する権利であって、著作権も人間が直接知覚できるような表現物が対象だったといえる。そして、情報化社会では、少なくとも特許の対象の中に「ソフトウェア」「プログラム」が入ってくるとともに、著作権も、必ずしも人間が直接知覚できないような情報（たとえば、プログラムのバイナリコード）であっても、保護の対象になるというように変わってきたものと考えられる。

ソサイエティ 5.0 になるとどうなるのか、という点では、特許の世界では、上述の通り、人工知能 (AI) のソフトウェアの重要性が指摘されることが多いものの、以下に説明するように、AI が処理できるような情報（データ）を、どう扱うのかという点が、むしろ本質的と考える。その場合、少なくとも、著作権という枠組みだけでは、AI 処理可能な情報を保護するという点では、不十分であるという点も、本稿において指摘したい。

そのような中で、このような AI、特に、「データ駆動型人工知能」と呼ばれる技術についての問題意識としては、大きく分けて AI のソフトウェア側の課題と、データそのものに対する保護についての課題があると考えられる。

第1の点としては、最近の AI は、処理の中身が「ブラックボックス」だということで、非常に問題だと

ということが指摘されることが多い。その場合、特許権としての設権の過程では、記載要件や特許要件を、どう扱えばよいのか⁽¹⁾、そしてひとたび特許権として成立した後は、技術的範囲や保護範囲を、どう考えるべきか、が問題となる。

第2の点としては、権利行使を想定したクレームドラフトを考える際に重要な点として、サービスを提供する側が本質的に複数の主体になるということが課題と考える。結果として、弁理士として特許のクレームをつくるという観点では、そのクレームと侵害との関係や、実施行為との関係を、どのように想定すべきか、ということが、もう一つの課題と考えられる。

一方、第3の点として、データについては、IoT (Internet of Things) 技術を背景として、あらゆるところからデータを集めてくることができるといふわけであるから、大量のデータが集積されることとなり、AIは、後に言及するように「データ駆動」ということであれば、データそのものの保護が、現在の知的財産制度で十分であるかが、課題となる。

つまり、「データ」に関しては、私見ではあるものの、AI技術や、次の世代のいろいろな技術が発展していくためには、標準的に開発者が利用できる公開されたデータが不可欠であると考ええる。

一方で、当然、各私企業がビジネスを自己に有利なように展開していくことを考える上では、各社のビジネスに直結するデータについての管理に、各社ごとの工夫が、必要になる、と想定される。

1.2 データ駆動型人工知能技術の技術的背景

さて、「データ駆動型人工知能」が技術的に可能となった背景として、一つは「機械学習」と呼ばれる技術の発展がある。特に有名なのは「ディープラーニング (深層学習)」であり、「ビッグデータ」と呼ばれる大量なデータが使えるようになったということ、さらに表立っていわれることが、あまりないものの、非常に「高速に演算できるハードウェア (たとえば、GPU : Graphics Processing Unit)」が実現されてきたという、それらが三位一体となって現在の技術的な応用が広まってきているという背景がある。

さらに、人工知能の応用分野も、どんどん広がってきており、以下の図2では、縦軸が人工知能の寄与度、横軸が社会に対する影響の程度を表している⁽²⁾。上に行くほど人工知能で完全に自動化されているという応用であり、下の方であると人間の意思決定を支援しているということになる。左側は、影響が軽微という言い方がいいかどうかであるのだが、それから、右側では人の命にかかわるようなところまで応用が広がっている、というようなグラフになっている。

(1) 酒井 将行「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」別冊特許第23号 (日本弁理士会中央知的財産研究所 研究報告第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) 大坪 直樹、「XAI (説明可能な AI) そのとき人工知能はどう考えたのか?」リックテレコム発行 (2021) 中の図の記載を基に筆者作成。

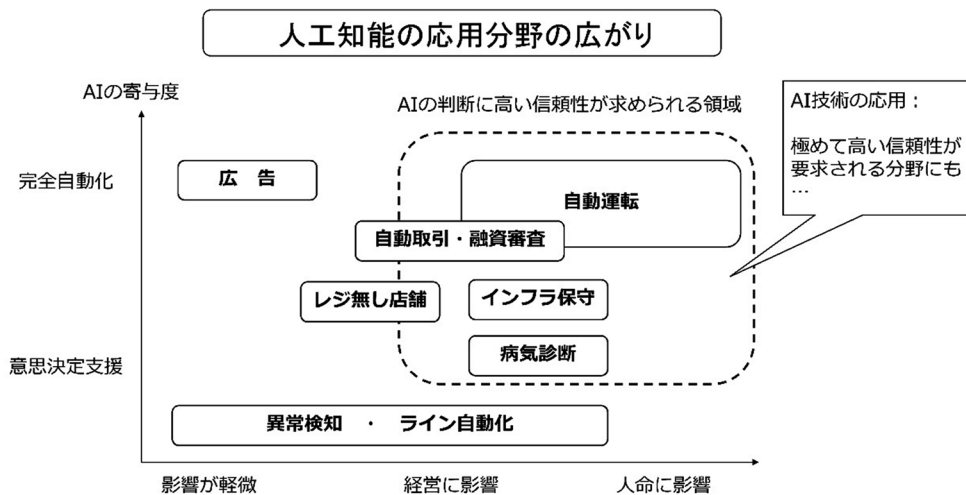


図2 人工知能の応用分野

この図2で、一つ言えることは、例えば「自動運転」という技術は、当然に人間の命にかかわってくるものであるし、かつ自動化をさらに進めようとしている分野になる。あるいは、医療系の事業領域では、病気の診断とか、場合によると病気の治療というところにも、人工知能の技術が入ってこようとしている。したがって、AIの応用として、一般に話題になることが多い、囲碁とか将棋とかいう用途以上のところで、非常に高い信頼性が要求されるような分野にも、人工知能がどんどん応用されようとしているというのが現状である。

そのような中であって、東京パラリンピックが開かれている中で、選手村で自動運転のバスと選手の方が接触事故を起こし、全治2週間ということで試合を欠場ということになられたということが、ニュースとなった。

したがって、これは、現実に、技術の応用をしていくという中での難しいところが、一つ、実際に出てきたところといえる。このようなシステムの信頼性を保証していくということを考えるときには、人工知能の「モデル」、つまり、処理をしている中核のところだけではなくて、モデル生成には、データとしてどういうものを使うのがいいか、あるいは、信頼性を保証するにはどういうテストをしなければならないかなどということは、技術的には、一見、周辺のように見えるけれども、非常に重要ということを示している。

データ駆動型人工知能では、まず、大量のデータを準備しておいて、機械学習という手法でモデルを作っていくという一つのフェーズ（「学習フェーズ」）がある。その学習フェーズで生成されると、出来上がったモデルを使って、新たにモデルに、ある入力があったときに、たとえば、AI囲碁プログラムであれば、次の一手はどこに打つ、ということを出力として出すフェーズと、2つのフェーズに分かれている。後者を「使用フェーズ」と呼んだり「予測フェーズ」と呼んだりいろいろ呼び方はあるものの、大きく2つに分かれているということが特徴的である。要は、大量のデータを使ってモデルを作るところ、作ったモデルを使って何かをするところということである。

したがって、学習フェーズでは、大量のデータを使ってコンピュータの多くのリソースというか負荷を掛けて計算をするので、ここは非常に重い計算処理ということになり、技術的には重要なところということになる。

あと一つ、ここに学習のときの問題点として、「過学習」というものがある。

これは、機械学習で、学習データに対してあまりに忠実に適合しすぎて、本来、データが示唆する傾向から大きく外れてしまう現象をいう。モデルの構造や、学習方法にも依存するものの、そもそも、学習用データの質に依存する部分も大きい。

また、現在最も応用の進んでいる分野である画像認識などの「学習済みモデル」については、「転移学習」

と呼ばれる技術がある。

「転移学習」では、あるタスクで使っていたモデルというものは、まったく別のタイプの画像を認識するというタスクのためのモデルを作るときにも、入力に近い側というのはそのまま流用することができる。要は、入力に近い側は、そのまま、まったく、パラメータの値は固定して変えずに、出力に近い側の層だけ学習させる（パラメータを変えていく）ことが可能で、それで十分性能が出る場合がある。

つまり、異なるタスクについて、一から全部の層について学習すると、また大量のデータと非常に多くの計算をする必要が生じるものの、「転移学習」という技術でモデルの流用をすることで、かなり小規模なデータで学習ができるということが知られている。したがって、「転移学習」は、ある問題を効果的に学習するために、他の学習結果を再利用する技術といえることになる。

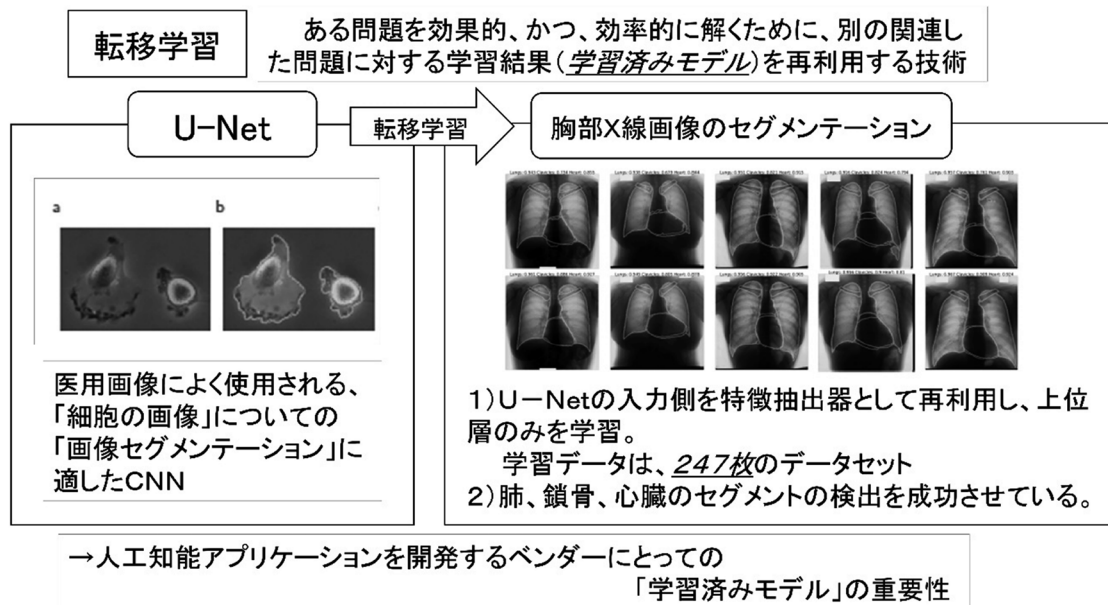


図3 転移学習を応用した胸部 X 線画像のセグメンテーション

例えば、図3に示すように、「U-Net」⁽³⁾と呼ばれていて、公開されている画像認識のためのモデルがある。この図3に示すように、入力としては、光学顕微鏡の画像の中に、細胞があるという画像の中で、どこが細胞なのかを切り分ける作業をする、そのような人工知能になる。これは「セグメンテーション」と呼ばれるタスクに相当する。この既存のモデルを使って X 線の画像をセグメンテーションすることを行った例がある⁽⁴⁾。

図3に示した例では、X 線画像で、どこが肺で、どこが心臓で、どこが鎖骨かということに分けるモデルを作るに際して、前述したように、既存の公開された「U-Net」というモデルの入力側のパラメータはそのまま使っている。学習は247枚の画像で可能であったと論文では記述されている。これはこういうモデルを作るときには破格に少ないデータ数である。

何がいえるかといえば、一般的に、AIモデルをアプリケーションとして作っているベンダーにとってみると、「学習済みモデル」を生成すると、そこで終わりではなくて、それを作ったということが、自身の他のビジネスにも流用できるという、非常に重要な意味があるということになる。

(3) Olaf Ronneberger, Philipp Fischer, Thomas Brox : “U-Net: Convolutional Networks for Biomedical Image Segmentation”, <https://arxiv.org/abs/1505.04597>

(4) Alexey A. Novikov, Dimitrios Lenis, David Major, Jiri Hladůvka, Maria Wimmer, Katja Bühler : “Fully Convolutional Architectures for Multi-Class Segmentation in Chest Radiographs”, <https://arxiv.org/abs/1701.08816>

さて、以上から、「学習済みモデル」が、技術と知財においてなぜ大事かということだが、一つ目はビジネスにおいて、実際マネタイズするという局面で使われるのは、主としては学習済みモデルということがある。2つ目は、今のところ学習用に使うデータは、AIモデル設計・生成を委託する委託元が持っているという場合がほとんどである。それに対してAIモデル生成の業務を受託するベンダーが作成するのは、当然、「学習済みモデル」であるので、ベンダーは、「学習済みモデルは自分たちのものだ」と言いたいという、そのような関係性が生じることになる。

さらに上述したように、学習済みモデルは、ほかのビジネスに流用することができるので、学習済みモデルの権利を持っているということがベンダーにとっては非常に重要で、自分のビジネスの自由度を決めるということになる。

また、学習用のデータが人工知能の技術において重要である理由としては、データ駆動型と呼ばれているAIである以上、学習用のデータがないと、そもそも何もできないことがある。一方で、モデルの性能はデータの量だけではなくて質にも大きく依存する。さらに現在、一番応用が進んでいるのは「教師あり学習」と呼ばれる学習方法で、そのようなモデルを学習させるためには、正解データとAIモデルへの入力データとがセットになった学習用データを、予め準備しておく必要がある。そして、このような「学習用データ（の少なくとも一部）」については、技術の発展という観点からは、広く利用可能にすることが望ましい。

特に、技術の進歩のためには、データを、どこかの会社だけが独占しているという状態は、あまり好ましい状況ではない。しかも、後述するとおり、当然、本当の競争領域（「コンペティティブ領域」（クローズ領域））で、データを共有するというのは、そもそも難しいと思われるものの、いわゆる「オープン領域」とか、「プリコンペティティブ領域」と呼ばれる、競争領域の前段階の領域では、「データの共有」ということが非常に重要になってくると思われる。そのような状況となることを前提としても、データ駆動型の人工知能を使った技術の世界では、最終的に企業の競争力は、「自分のビジネスに適して自分（のみ）が管理できるデータ」が重要になると予想される。

なお、データ駆動型の人工知能技術では、処理がブラックボックスなので明細書を記載するのが難しいというような言及がされる場合がある。

ブラックボックスと明細書の記載との関係に関して言えば、明細書の書き方について、これまでも、すべて基本となる第1原理から発明の動作・作用がわかった上で、発明が把握されて明細書になっていたわけではない場合もあったと考えられる。たとえば、化学分野やバイオ分野では、必ずしも第1原理から対象となる発明の構成や作用の全てが分かっているわけではないものの、発明というものは存在してきたわけであり、現在も存在していると考えられる。

そこで、（モデルの構造とパラメータの値で特定される）モデル内部の処理の部分が第1原理からは説明が不能という場合でも、例えば、化学系の明細書では当然、そのような場合に相当するものでも、明細書に明確性を担保して、発明の具体例を記載することは可能という前提があったわけで、決してそのような分野の明細書と違うことが必要になるというわけではないであろうと、個人的には考える。

さらに、人工知能の特許のクレームが話題となるときに、モデルの入力と出力だけを特定したクレームを作って、モデル中の処理は、「機械学習」ということだけを記載しているクレームや、それに対応するような明細書が、一番、権利範囲が広いということが言われることもあると思う。もちろん当然ながらそのような技術分野というか、そのような技術の進展の状態がその段階にある技術分野も想定される。

これは、私見としては、いわゆる「機能的クレーム」において、当該構成要件が、「周知慣用技術」に属するのか、むしろ、発明の特徴部分に相当するかで、技術的範囲の解釈が変わってくることと、本質的には、相違はないものと理解している。

たとえば、電気・電子のような分野でも機能的にクレームが書いてあるというような場合であれば、クレームの記載だけでは構成が確定できないということは、これまでも度々あったはずであり、AI関連発明といっ

でも、そのような局面で行われてきたクレーム解釈の延長で対処できることが多いと考えられる。これに対して、一方で、サービスを提供する側が複数である点については、クレームのドラフトに、ある程度工夫が必要なだけでなく、このような場合の実施行為をどう解釈していくかということも、検討課題と思われる。

ただし、そもそも、モデルへ特徴量として何が入力されればいいのかということも含めて、本来の意味では「データ駆動」と言うのであればブラックボックスというべきである。つまり、データ駆動型人工知能では、コンピュータが何を特徴として対象を評価すべきか自体を決めているわけであるから、要は、モデルの入力だけを特定することが権利範囲として広いというのは、それなりに限られている状況ともいえる。

ここで、明細書の記載要件を満たすために、具体例を記載することが必要であり、そのように具体例を開示するのでは、広い保護範囲をおさえるようなクレームや明細書が記載できないということであれば、特許出願自体をしない、という選択肢もありうることになる。ただし、「データ駆動型人工知能」では、そのような選択が、必ずしも望ましくはない、というのが、後述するように、筆者の個人的な見解でもある。

2. データ駆動型人工知能の技術を保護するためのクレームドラフトについて

2.1 「学習済みモデル」の保護とこれをクレームできることの重要性

ここで、事例を2つほど挙げることで、「学習済みモデル」をクレームに記載して、それ自身の特許としての権利化を図ることができる、ということがどういうことなのかを、もう一度前提から考えてみたい。学習済みモデルを特許として保護するのがいいのかどうかという場合に、物の発明の知的財産権による保護というときと、人工知能プログラムの知的財産権による保護というときの考え方が、いろいろ対立するというほどではないものの、少し考え方として違うところがあるというのが筆者の印象である。

つまり、物のときはこういう方法で保護するのが望ましいと言われていることが、AIプログラムのときにも、そのまま、あらゆる場合について、そう考えていいのかというと、必ずしもそうではないのではないかと考える。

たとえば、第1の事例としては、化学系の会社が、ある物質を使った何かを生産する方法の発明をしたときに、知財の保護としてはどう考えるかということがある。考慮する点としては、そのような生産方法が、公開される、あるいは、流通されるものと把握できるのか、物を製造販売することは必要だとして、生産方法は公開されるか、などがまず挙げられる。もっとも、生産方法を公開する必要は、ほぼないだろうと考えられる。

また、リバースエンジニアリングとの観点では、物を入手しても、物だけからどうやってそれが製造されているかを明らかにするというのはかなり難しいことであり、ほかの会社が何か同じようなものを作っているとして、それが同じ方法で作られていることを立証するというのは、非常に難しい。その結果、一つの考え方としては、生産方法については特許を出願せずに秘匿化して保護するというのも可能性としては当然あると思う。

それでは、同じような観点で、学習済みモデルについて考えると、AIベンダーがある特殊な方法で学習処理をして独自のモデルを作ったという発明があったとする。例えば、特許出願書類のデータを読み込むと、その内容で特許が取得できるかどうかの可能性を非常に高精度に判断してくれる学習済みモデルが作られたという事例があるとする。そして、学習済みモデルを利用するサービス提供としては、顧客にはSaaS (Software as a Service) として提供することによって、収益を上げるという場合に、では、知財としての保護は、どう考えるか、特許出願して権利を取るのがいいのかというのは検討課題になる。

たとえば、特許庁の特許審査事例⁽⁵⁾には、学習済みモデルのクレームの書き方の例が記載されている。簡

(5) 特許庁「AI関連技術に関する事例について」、https://www.jpo.go.jp/system/laws/rule/guideline/patent/document/ai_jirei/jirei.pdf

単に言えば、クレームの例として、学習用のモデルのまず数学的な構造が記載されている。学習済みモデルは、「構造」と「パラメータ」で特定される場所、まずは「構造」が特定される。そして、それに続いて、モデルの学習方法をクレームに記載して、学習済みモデルを特定するということがされている。

その点では、このようなクレームの形式は、学習済みモデルを、それを生成方法で特定しているというふうに読めるわけで、一見すると、形式的にはプロダクト・バイ・プロセス・クレームのようにして、「学習済みモデル」を特定しているともいえる。このようなクレーム形式についての技術的範囲の解釈⁽⁶⁾はさておくとして、形式上は、プロダクト・バイ・プロセスのようなクレームなのであるから、生成方法がクレームに書かれる以上、その生成方法が具体的に明細書に開示されることが必要になる。とすると、先ほどの物の生産方法の場合と比べて、特許明細書に「生成方法」を開示することを、どう考えるべきなのかが問題になり得る。

一方で、SaaSというサービスの提供方法は、現在、非常に増えているクラウド型サービスであって、ソフトウェアを従来のような「売り切りビジネス」（媒体での販売、ダウンロード販売など）というよりも、アプリケーションプログラムの機能の部分だけをサーバーで提供して、クラウド側のサーバーと協調して、ユーザー側にあるローカルのコンピュータが処理を実行するというビジネス形態である。要は、アプリケーションプログラム自体は、ローカルのコンピュータにはインストールされずに、サーバー側で動いているというサービスである。

そのようなサービスの場合、「学習済みモデル」が存在しているのはサーバー側であって、ユーザー側の方では、ある意味、処理結果を受け取っているだけになる。とすると、先ほどの生産方法の発明との比較で言えば、学習済みモデルを顧客企業に、ライセンスして使用してもらおうビジネスをするとして、SaaS方式であればAIモデルを公開したり、流通させる必要は必ずしもない。それから、当然、学習済みモデルはサーバーの中にあるので、リバースエンジニアリングというのは事実上できないことになり、逆に、特許侵害されたときの侵害立証が可能かということに関しては、近年、査証制度というものができてはいるものの、現実問題としては、かなり難しいのではないかと予想される。

結果として、それでは、技術の保護としては、先ほどの物の生産方法の発明と同じように秘匿化するというのが適切とも考えられる状況もあるかと考えるものの、保護の方法として状況により、特許出願をする方が望ましいという局面も多いのではないかと考える。

ということで、査証制度というものができたことは非常に大きな進歩であるし、この制度が導入されたときの改正の解説では、ソフトウェアなどが査証の対象ということに挙がっているものの、違和感という誤解を招くかもしれないが、査証制度ができたから、それでは、侵害立証も含めた知的財産権による保護について、検討すべき事項が残っていないかということ、そうとも言い切れないというのが、筆者の見解である。

2.2 学習済みモデルを特許で保護するという事

そこで、学習済みモデルの保護方法としては、「秘匿」か？「特許」か？という点について、さらに検討したい。

図4に示すように、物の生産方法の発明の場合は、特許明細書に開示しなければならないというのは、材料に始まってすべて開示しなければならないわけであって、公開のデメリットが、それなりに大きい。

一方で、学習済みモデルというのは、今のところ特許庁の運用で明細書に開示が要求されるのは、図4において、「b（学習方法のアルゴリズム）」と「c（学習モデルの構造）」の範囲だけである。したがって、必ずしも物の生産方法の発明に比べて、公開によるデメリットが大きいかというと、そうでもない可能性もあ

(6) 酒井 将行「AI・IoT 技術によるビジネスモデルに対する知的財産権」別冊特許第20号（日本弁理士会中央知的財産研究所 研究報告第45号『特許クレーム解釈と記載要件』）

ると考える。

また、「学習済みモデル」というときには、おそらく、最大でも、図4中の「b」「c」に加えて「d（学習済みモデルのパラメータ）」までの範囲だと想定される。この状態で、少なくともこのbとcについて記載して特許出願したとして、学習方法のアルゴリズム自体が公開されるのが、技術情報の秘匿の観点から好ましくない、という反論が想定されるものの、そもそも、この技術分野では「アルゴリズムを秘匿すること自体が難しい」ということがあると考える。ここで、秘匿するのが難しいという理由は、要は、営業秘密として、その技術を開発している会社の中で秘密状態を保持するのが難しいという意味ではない。

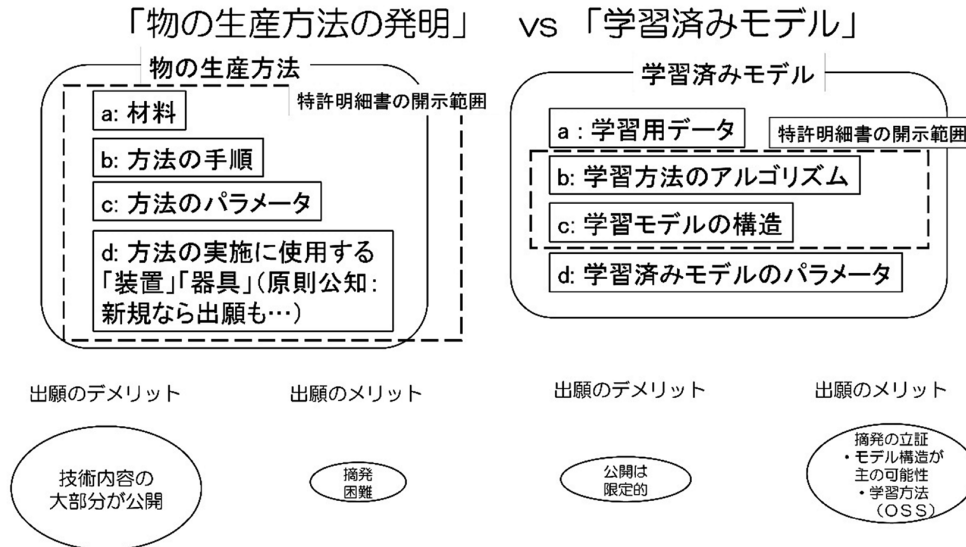


図4 物の生産方法の発明と学習済みモデルの出願のメリットとデメリット

AIの技術分野では、仮に、自分たちがアルゴリズムを秘密にしているとしても、第三者により、どんどん世の中に公開されていくということがある。場合によると特許として公開される、という対象ですらなく、自由に使っていいものとして（オープンソースとして）、インターネット上などで公開されていくことが多い。このため、アルゴリズムの基本部分については、これを秘密にしておくというのが、実効的にどれぐらいできるか、というのはかなり難しいと考える。

しかも、学習済みモデルというか、データ駆動型人工知能の技術としての一番重要なところは、データが集まってくるに従って、学習済みモデルのパラメータが、どんどん更新されていくところにある。うまく運用することで、モデルの性能を向上させていくことができるわけである。したがって、特許出願時点の学習済みモデルのこのbとcの部分が公開されているということが、技術が全部公開されて第三者を利するだけ、ということには、たぶんならないだろうと考えられるわけである。

個人的には、秘密にするべきなのは、むしろ、学習済みモデルのパラメータを更新することができるように、順次蓄積されていく「学習用データ」のところなのだろう、というふうに考えている。

2.3 データ駆動型人工知能に対する権利行使を考慮したクレームドラフト1

それでは、次に、データ駆動型人工知能についてのクレームドラフトについて検討したい。

これは、ことさらにここで言及することではないかもしれないが、図5に記載のとおり、クレームドラフトにおいて、機械的な構造物のときは部材の機械的な結合、あるいは電気・電子の分野であると部品の物理的とか機能的な結合、あるいはネットワークの発明になったときには、ネットワークを構成しているもの同士が通信によってどういうふうに機能的、物理的につながっているかということがクレームには記載されるのが通常である。

一方で、現在、ITで実現されているのは、クレームの対象が、仮に、物であるとしても、どちらかという、その実態は「サービス」であって、クレームに記載されるのも、どちらかというソフトウェアの機能が列挙されている、ということが多いと考えられる。

図5において、「機械的な構造物」から「IT・ネットワーク技術における物」までの範囲では、クレームの背景として、明らかに一体的に「物」として何かが存在しているというのが、前提だと考えるものの、現状では、クレーム対象がサービスの場合は、背後の「物」というものに対する意識が相当薄くなっていると考えられる。

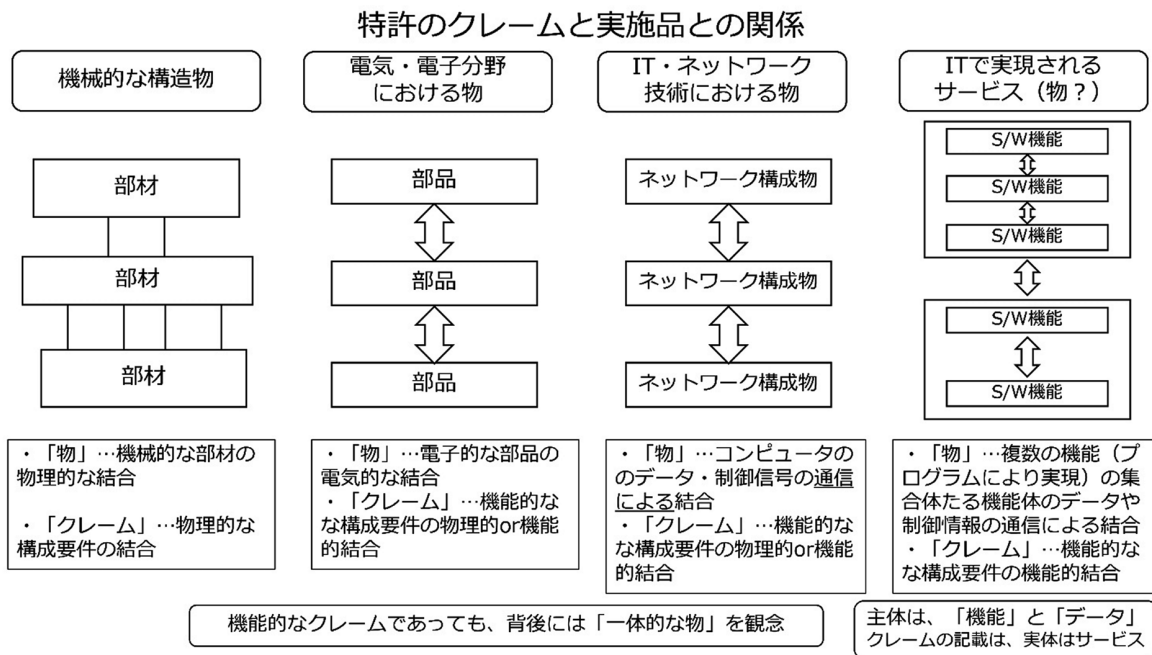


図5 特許クレームの構成要件と実施品との関係の変遷

もう一つ、技術的な進展という点で言うと、「仮想化技術」というものが進展してきている。例えば、物理的な一つのサーバーの上で仮想的に複数のサーバーが動いているかのようにしてサービスが提供される場合や、あるいは、複数のサーバーが、あたかも一つのサーバーのようにして運用されるということが行われている。

したがって、逆に言うと、ユーザーから見ると、どこに物理的にサーバーがあるか、ということは、ほぼ把握は不可能で、場合によると、複数のうち一つのサーバーは国境を越えているということも容易に起きているというのが現状である。仮想化の技術が利用されている状況では、どの物理サーバーで、どのアプリケーションプログラムが動いているかということ自体に、だんだんと技術的には意義がなくなりつつあるというふうに見える。

一方で、クラウドサービスという呼び方をしてきたものの、それ自体も技術的にはどんどん変化してきており、クラウドのサービス提供者側のサーバーというのは、インターネット越しの向こう側にあるだけではない。近年、企業の施設側にも、たとえば、企業の工場の中にそのようなサーバーが分散して置かれるというようなサービスも、もう始まっている。理由はいくつかあると思うが、処理の負荷が非常に大きいということもあり、それを分散して処理をさせようということも理由の一つだと考える。

また、工場の中のデータを、そのまま外に出すのは避けたいという、ユーザーサイドの判断もある。この場合は、工場敷地内のサーバーについては、管理主体は、ユーザーということになる場合も多いと思われる。

結果としてサービスは、クラウドと呼ばれるインターネットの向こう側にあるサーバーで提供されていて、

サービスを受け取るのはユーザーの敷地内にある端末であるというような構図自体、もう技術的には崩れ始めているというのが現状だといえる。

その前提で、従来からネットワーク型発明と呼ばれているものについて、どのようにクレームを作るべきと考えられていたかということ、図6に、簡単にまとめる。

これまでは、サーバーがあってユーザー側の端末があるというのが大前提で、ソフトウェアでサービスが提供されている、というときに、クレームはどうドラフトすればよいのかという点については、全体を含んでいるシステムのクレームとかシステムの方法のクレームだけではだめなので、サーバー側だけの装置クレーム、端末側だけの装置クレームを作るのが必要だとされてきていることが多い。

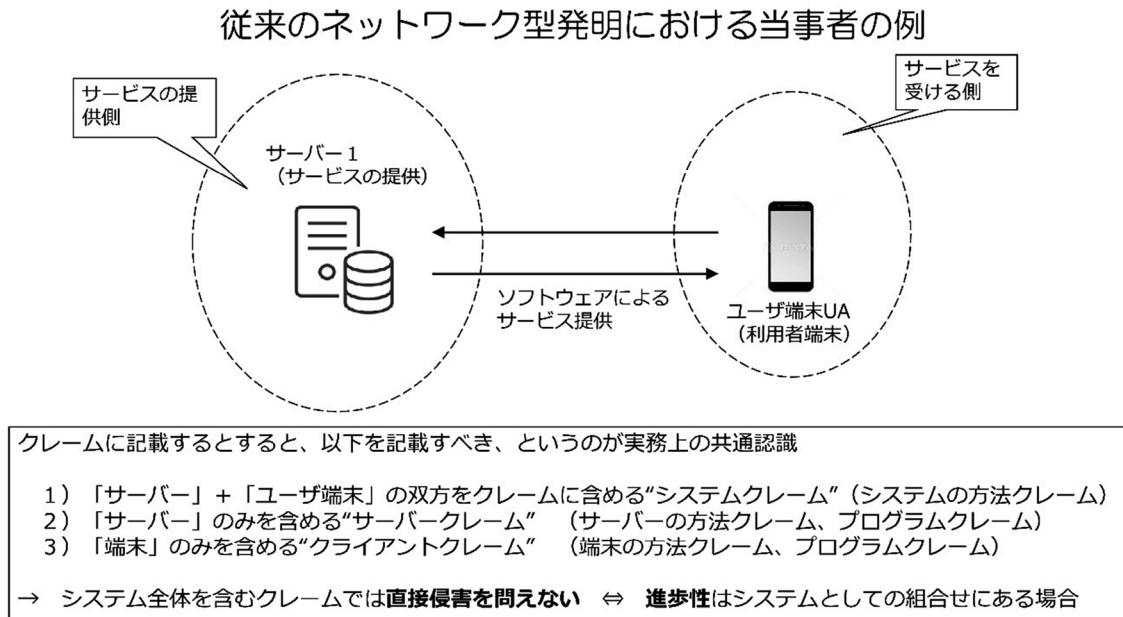


図6 従来のネットワーク型発明の当事者とクレームドラフト

システム全体のクレームでは、サーバー側と端末側とで、実施主体が複数に分かれていることが通常で、直接侵害は問えない、ということになる。一方で、両方合わせないと発明の進歩性というのはクレーム上で主張し難いというときもあって、その両方の兼ね合いの中で、どのようにクレームドラフトすべきかというようなことが議論されてきていることが多いと思う。

次に、クレームが、どのように記載されているときにどういうふう解釈されるかという点で、サーバーとクライアント端末という関係で、その記載が問題になった一つ有名な裁判例に、「インターネットナンバー事件」⁽⁷⁾がある。

この裁判例で重要なことは、図7に示すように、あるインターネットを使ったサービスの全体像のクレームがあって、その中にサーバー側が実行する処理と端末側が実行する処理が、方法クレームの中で、各ステップの中に分かれて記載されていたという点である。

この図7のクレームの中のBやFというステップの処理は、クライアント側で実行することで、サーバー側では、実行してはいないという状況であった。

裁判において、被疑侵害者側としては、まず開示義務違反についての主張を行っている。これは処理を実行している主体が、クレームの中で分かれてしまっているから、発明が明確でないという主張をしたわけで

(7) 知財高判平成22年3月24日(平成20年(ネ)第10085号特許侵害差止等請求控訴事件) (「インターネットナンバー事件」)

あるが、裁判所はそれについては退けるという判断をした。

実際の侵害の判断のところでは、一番重要なのは侵害の主体性だと考える。そもそもこの発明は、クレームの記載でも、「アクセスの発明」ではなくて「アクセスを提供する方法」、つまり、サーバー側の処理として発明が記載されていて、クライアントによるアクセスがないと本件発明にかかる特許権を侵害することができないということではなく、また、クライアントの主体的行為によって、クライアントによる個別なアクセスが本件発明の技術的範囲に属するものとなったり、ならなかったりするものではない、と裁判所は判断した。この結果、この被控訴人の被控訴人方法の提供自体が本件発明の実施行為と評価されるべきということになったという事案である。

この件は、例えば特許誌などでも、いくつか講評が出されている⁽⁸⁾。このような講評では、本件ではクレームが「アクセスを提供する方法」になっていたということが重要という指摘がされ、アクセスの提供方法だから、これはクライアント側ではなくてディレクトリーサーバーといわれるサーバー側の人間の行為として把握できるので、要するにディレクトリーサーバー側の「単数者」によってこの行為が充足されるというふうに裁判所が判断したのが重要である、という評価をされていることが多い。

IT系の事案で侵害行為において、サーバー・端末の記載が問題となった裁判例

知財高裁平成22年3月24日判決（インターネットナンバー事件） （平成20年(ネ)第10085号特許侵害差止等請求控訴事件）

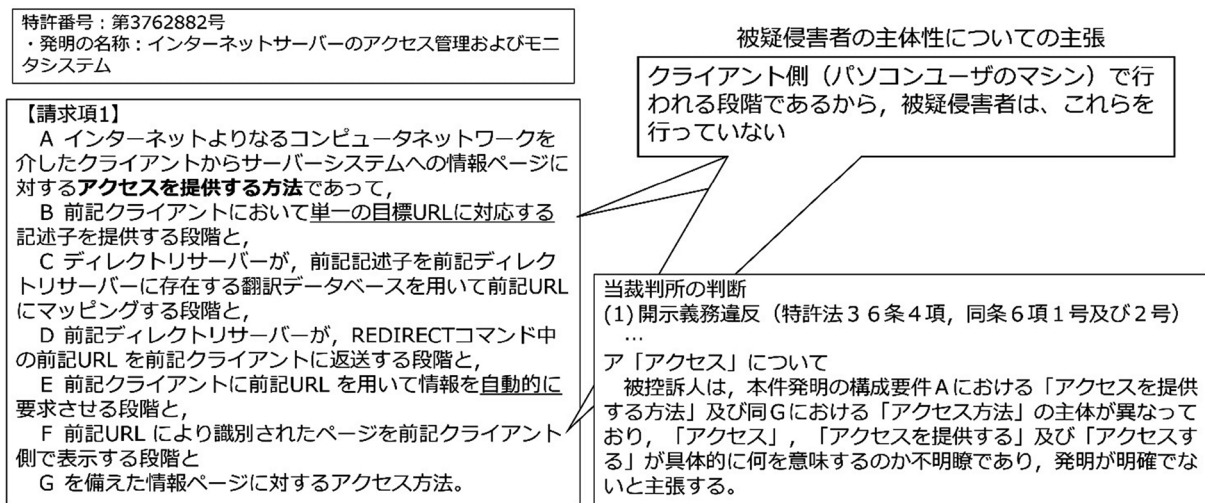


図7 インターネットナンバー事件のクレームと被疑侵害者の主張

しかしながら、その場合でも、例えば、構成要件Bというのは、これは明らかにクライアント側の装置の動作ということが疑問点として指摘された結果、例えば、サーバー側だけで動作する場合はこういうことであるから、こういうクレームにしておくべきだったということで、この元のクレームを書き換えられて、サーバー側だけの処理についての方法クレームというものが提案されるなどしている。

もっとも、前述したように、そもそもサーバー側でどこまでの処理を実行するかということも、技術としては、最近、だんだん流動的になってきているという中で、上記のような指摘は、重要であるものの、あらゆる状況をカバーするような特許クレームを、本当に出願時にドラフトできるのかというのは、また別の問題として出てきているというのが現状と考える。

(8) 鶴谷 裕二「方法の発明における侵害主体性について特許権の侵害を肯定した判決 知財高裁平成22年3月24日判決（インターネットナンバー事件）（平成20年（ネ）第10085号特許侵害差止等請求控訴事件）」特許 2011 Vol.64 No.6 pp.15-20

さらに言うと、この「インターネットナンバー事件」というものを、ネットワークを介して実現される発明の方法について、クライアント側の主体的な動作の関与が不要であれば、これは、客観的に、そのような「ネットワークを使った、そのようなサービスの提供方法」というものが存在するののかという問題ととらえて、まず第1段階では、客体としてそのような方法が現実にあるのかということを確認した上で、それが存在するならば、その提供方法の実施主体、提供主体が誰と考えるのが適切かということを一応分けて判断している裁判例というように、とらえることができる。と考える。

これ以外にも、たとえば、「眼鏡レンズの供給システム事件」⁽⁹⁾というようなものでも、この場合は、方法クレームではなくて物のクレーム、システムのクレームだったものの、こういう複数段階に分けて考えるという点では、同様であろうと考える。

もっとも、この「眼鏡レンズの供給システム事件」は、技術的に重要なところ（眼鏡店側の処理）がどこかというより、システムの各構成要件の充足性での各実施主体の問題と、権利行使の対象となる実施者とを分けて考え、権利行使の対象は「システムを支配管理している者」と認定して、判決がされたと考えられるので、事案の様々な事情を考慮した結果として、例外的に判断がされた部分がある、ともいえるのではないかと思う。

要は、実施主体の認定をするときに、誰が管理しているかということについては、著作権侵害において、すでにいくつかの事案、たとえば、「まねきテレビ事件」⁽¹⁰⁾、「ロクラクⅡ事件」⁽¹¹⁾などがあり、それらの考え方をある程度、特許侵害でも同様に適用されたものともいえる。

AI関連発明のクレームのポイント

1) 技術の中核となる処理のみで、クレームを作成できること。

→ AI技術であれば、「学習済みモデル」は1つの可能性。

2) 中核となる処理を含む「サービスの処理全体」のクレーム。

→ 「間接侵害」としての主張の可能性も残せること。
(中核部分は、もちろん、周辺部分への権利行使の余地。)

3) 実施行為として、サービス提供者側の行為を抑えること。

→ もっとも「不法行為地」については、論点として残る? …

図8 インターネットナンバー事件も考慮したAI関連発明のクレームのポイント

以上を考慮して、AI関連発明のクレームのポイントを図8にまとめた。インターネットナンバー事件のような判断がされた例はあるものの、どの程度の管理支配があれば、その主体の実施行為といえるのかというところは必ずしも明らかではないので、結局のところ「サービス提供者側がどのような行為をしているか」が、重要だとするならば、そのサービス提供者側の実施行為に着目したクレームを作る必要がある。

(9) 東京地判平成19年12月14日(平成16年(ワ)第25576号特許権侵害差止等請求事件)(「眼鏡レンズの供給システム事件」)、大須賀 滋「複数関与者による特許権侵害」パテント2013 Vol.66 No.4 pp.96-109

(10) 東京地判平成20年6月20日(平成19年(ワ)第5765号著作権侵害差止等請求事件)、知財高判平成20年12月15日(平成20年(ネ)第10059号著作権侵害差止等請求控訴事件)、最判平成23年1月18日(平成21年(受)第653号著作権侵害差止等請求事件)(「まねきテレビ事件」)。

(11) 東京地判平成20年5月28日(平成19年(ワ)第17279号著作権侵害差止等請求事件)、知財高判平成21年1月27日(平成20年(ネ)第10055号著作権侵害差止等請求控訴事件・平成20年(ネ)第10069号同附帯控訴事件)、最判平成23年1月20日(平成21年(受)第788号著作権侵害差止等請求事件)(「ロクラクⅡ事件」)。

さらにハードウェアとソフトウェアの関係について、例えば、「ロクラクⅡ事件」の事案であると、ハードウェアとソフトウェアということで、ある程度、どの主体が何を実施する、ということが切り分けられているというような状態であったと思われる。これに対して、AI技術の場合は、ほとんど全部がソフトウェアの処理になっており、どのサーバーで、どの主体が何を実施するのかというところは、かなり柔軟に変更し得るということが前提にある。そうすると、間接侵害を主張できるということも、やはり想定としては考えたクレームドラフトが必要だと考えられる。

したがって、上述した図8では、技術の中核となる処理だけでクレームが作成できることは当然必要であるとの前提で、AI技術であれば「学習済みモデル」は一つの可能性とあえて記載した。もっとも、後述するように、学習済みモデルだけが、必ず、実施行為として、クレームでカバーすべき中核かどうかというのは事案によってたぶん異なっていると思われる。つまり、筆者の見解としては、学習済みモデルの処理は、非常に重要なものであるものの、あくまで、一つの可能性であると考ええる。

さらに、中核処理を含むサービスの処理全体のクレームは、間接侵害を可能性として主張できるためには、やはり作っておく必要は、あるのだろうと思われる。

それから、実施行為としては、サービスの提供者側の行為として押さえられて、ただし、望ましくは、侵害行為が行われる場所としては、できれば技術によっていろいろ変わっていくところでない場所が「不法行為地」というふうに認定できる方が、出願人側としては有利なのではないかと考えられる。

図9に示すように、要は、人工知能技術では、学習フェーズが実施される装置と使用フェーズを実施している装置とが必ずしも同じでないというのが一つの特徴である。そこで、複数当事者の侵害行為という言い方をしたとしても、従来のように、サーバー装置とクライアント装置という切り分け方だけではなく、サーバー側でも、学習を実行する装置・主体と、AIを使用してサービスを提供する装置・主体とが別の当事者ということも考慮する必要が出てくる。

そこで、下記の図9の中で、B（「使用において、サーバーとクライアントの関係で別の当事者（サーバー1, 2, 3と、ユーザー端末UA）」）と記載した点については、これまでもいろいろな議論があったところであり、以後は、A（「学習と使用が別の当事者（サーバー2とサーバー3）」）と記載した点のことを考えてクレームドラフトをどうするかを、次に検討したい。

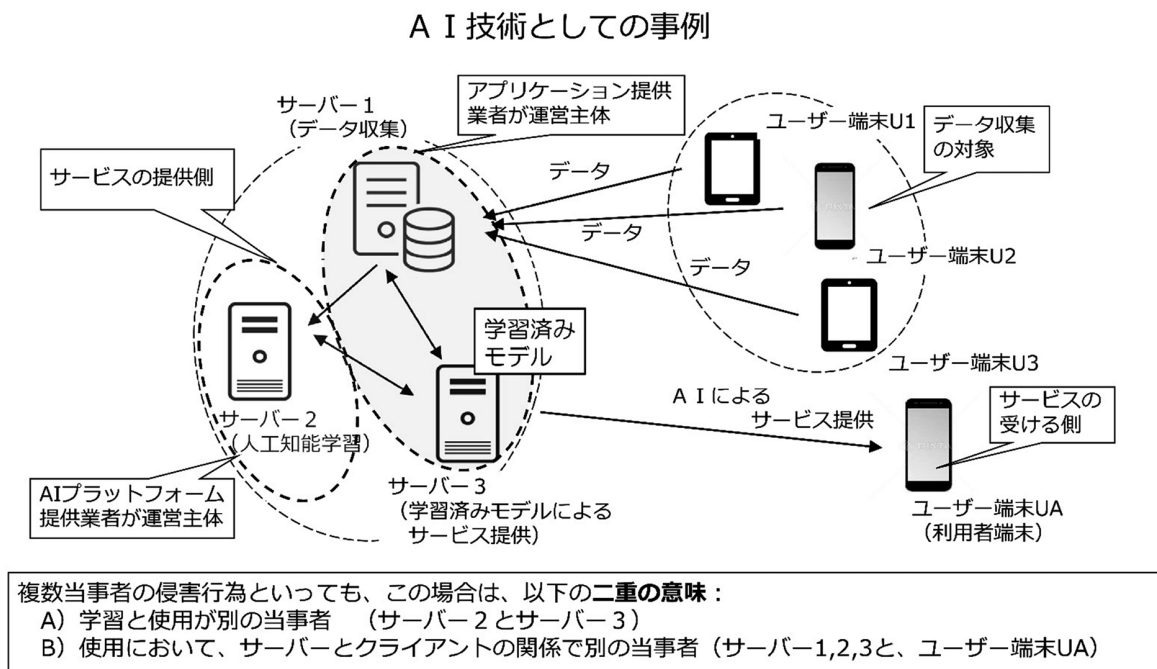


図9 AI技術をサービスとして提供する際の事例

AI技術の事例としては、要は、一つ非常に重要なところとしてデータを集めてくるところがある。そこで、データを集めてくるという処理を実行するサーバーがあって、それに基づいて学習をしてモデルを作るというサーバーが、さらにあって、そのモデルを使ってサービスを提供するというサーバーがさらにあるという例を考えることにする。この場合は、3社の運営するサーバー、あるいは、少なくとも、サーバーとしては3つが想定されるという前提で、ユーザーにサービスが提供されているということを考える。

このときに、今の人工知能技術のサービスの形態だと、データを集めてくるといふことと、人工知能のサービスを提供するといふのは、多くの場合は、(データがAIモデルの生成には最重要という観点からすると)一つの主体が実施しており、一方、人工知能を学習するといふ処理は、AIのプラットフォームの運営主体のサービスを利用しているといふような構造になっていることも多いのではないかと思う。もっとも、場合によると、3つのサーバーの運営主体が全て違うということもあり得る。

ということは、図9のように、サーバー3で学習済みモデルが稼働している構成になることを想定すると、クレームとしてどういうことを考えるかということが一つの課題と思う。

実務的な観点からいうと、図10に示すように、AIアプリケーションサービスを提供するサーバーが、上記のような意味で複数の物理サーバーに分かれ得る点については、日本では学習済みモデル自身を、学習を実行するサーバーとは独立にクレームすることが実務上可能ということに基づいて考察できる。したがって、学習を実行するサーバーについては、学習をする側と学習済みモデルを使用する側とを合わせてコンビネーションと考えた時に、学習を実行するサーバーは一方のサブコンビネーションということで、クレームでは、学習済みモデルが稼働しているサーバー装置を他方のサブコンビネーションとして記載するか、あるいは、「学習済みモデル」そのものをクレームすることが可能になる。

そして、特許性(新規性・進歩性)自体は、一義的には、「学習済みモデル」の記載が担うことになる。

もっとも、この場合、「学習済みモデル」の構成や学習方法そのものが新規である場合もあるし、あるいは、構成や学習方法は既知でも、その使用態様自体などが、新規である、という場合も考えられよう。

このような「学習済みモデル」により提供されるサービスのどこに特許性を求めるかは、発明の把握において、重要なポイントと考えられるものの、以下では、単に、クレームの形式の問題として、さらに検討する。

サービス提供サーバーが複数となる問題 : 1つの対処例

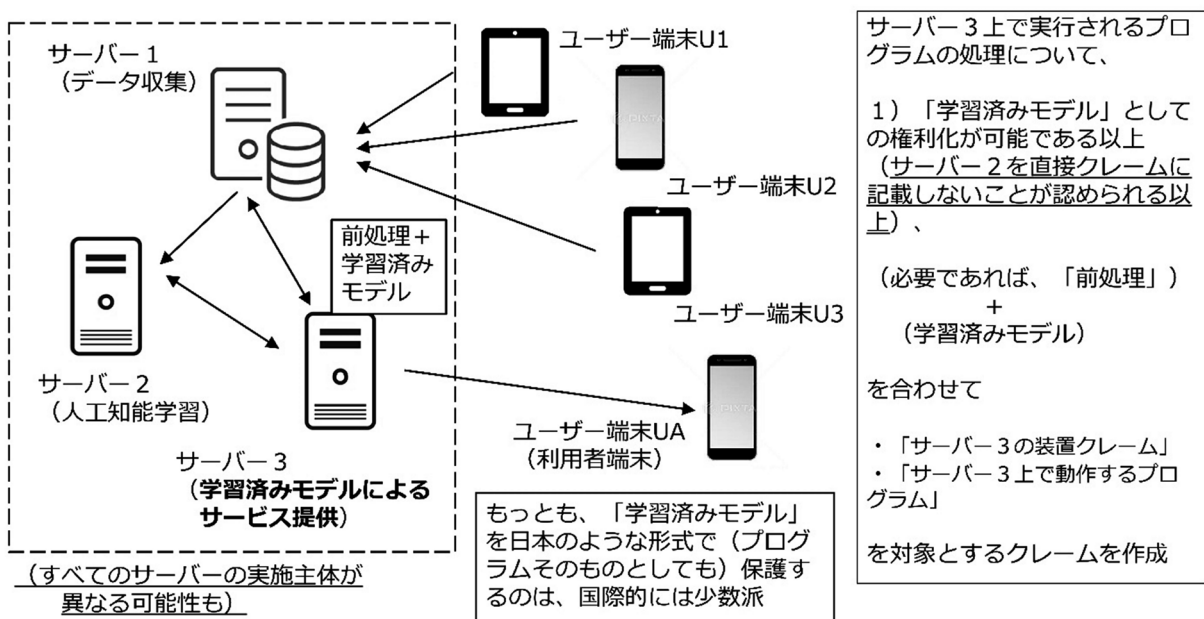


図10 サービス提供サーバー（主体）が複数となるときの対処例

まず、サーバーで実行される AI アプリケーションサービスに限って言うのであれば、学習済みモデル、または学習済みモデルが稼働するサーバーのみを構成要件とするクレームを作るというのが一つの方向性であろう。

一方で、方法クレームで「アクセスを提供する方法」やこれに準じるような発明の名称とするだけでは、どこまで保護されるのか、現時点ではグレーであるので、ドラフト時に「ユーザー側の使用」というよりは、「サービス提供者側の行為」、それを“使用”と呼ぶべきかどうかは検討課題であるものの、実施行為を考慮するのは非常に重要なポイントになる。

もう一つ、実施主体のことをおいて、侵害のことを主張するというのを考えるのであれば、一義的には、物のクレームの方が素直ではあると思われる。これも、北米の裁判の例や日本の「眼鏡レンズの供給システム事件」を考えると、方法クレームであると、各ステップごとに主体を考えることになる傾向があるものの、物のクレームであれば、システム全体を客体として考えるという傾向がある可能性は高い。

ということで、サービス提供サーバーが複数になるという問題については、サーバーが複数というだけではなくて、実施主体が異なっている場合もあるということ想定すると、学習済みモデルが実行している処理について、学習済みモデルとして、そもそもそれだけで権利化が可能であることを前提として、一つの検討対象となるサーバーは、学習処理を実行している側である。そのサーバーは直接クレームには書かないで、学習済みモデルを使用して処理をする「サーバー 3 の装置クレーム」、あるいは「サーバー 3 上で動作するプログラム」というクレームを作ることが、日本では可能と考えられるので、それを行うことが一つの解決方法だと思われる。

ただ、これで全部解決するというのであれば、非常にありがたい話ではあるものの、これは今後の非常に大きな課題であるが、諸外国でも同じように保護してもらえるかということ、必ずしも現状はそうでない。そこで、また別の考えが必要だということは課題としては存在しているというふう考える。その意味では、発明の新規性を担う部分が「学習済みモデルを生成する処理そのもの」などの場合は、「学習を実行するサーバー」と「学習済みモデルが稼働するサーバー」とを両方クレームに記載したシステムクレームを作成せざるを得ない局面もあろう。

2.4 データ駆動型人工知能に対する権利行使を考慮したクレームドラフト 2

以上では、前述したように、学習をさせるサーバーとデータを集めるサーバーと学習済みモデルが稼働しているサーバーというような考え方を前提として、クレームドラフトを考察してきた。

以下では、もう少しだけ違う状況で、かつ、もう少しだけ具体的な例について、検討をしたい。

下記の図 11 に示すように、2017 年にパテント誌の中で、日本弁理士会特許委員会の中のソフトウェア部会の先生方が、具体例を作って、いろいろ検討された内容⁽¹²⁾を参考に、この具体例について、アイデアをより人工知能の例に近くなるように変更して検討したい。

(12) 平成 27 年度特許委員会第三部会（ソフトウェア部会）「クラウド時代に向けた域外適用・複数主体問題」パテント 2017 Vol.70 No.1 pp.39-53

「クラウド時代に向けた域外適用・複数主体問題」(パテント 2017 Vol. 70) の仮想事例

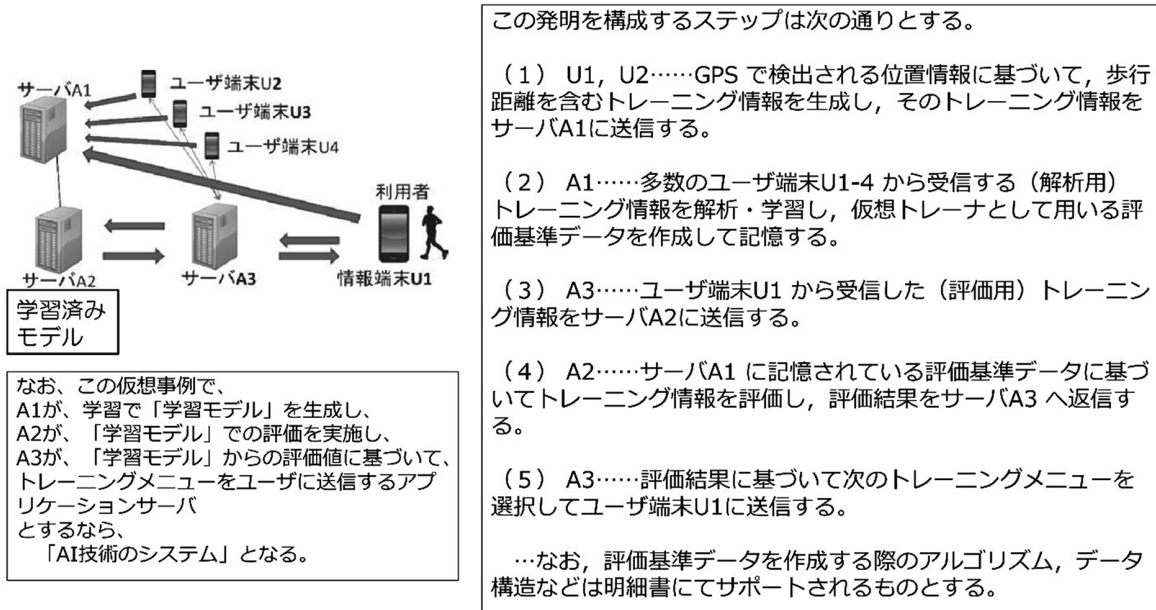


図 11 「クラウド時代に向けた域外適用・複数主体問題」(パテント誌 2017Vol.70) の事例

つまり、この場合、AI 処理を実施した結果が、直接ユーザー端末には提供されずに、ほかの提供者のサーバー処理が介在する場合の具体例であり、以下は、そのような例を、検討する。

要は、先ほどの例では、学習処理をしてモデルを生成するのがプラットフォーマーの別のサーバーであり、サービスとして「学習処理をする技術基盤」が別主体により提供されているという場合であったのに対して、以下の例では、「学習済みモデルの処理結果」を提供する当事者も別に存在しうる構成であって、最終ユーザーへのサービス提供という全体からみると、学習済みモデルの稼働を行うという行為については、このサービスの直接の提供者とは別の当事者が実施しうる場合ということになる。

パテント誌に書かれたときは、必ずしも、人工知能とまで言える状況ではなかったと想定されるものの、要は、スマホ端末をユーザーが持っていて、ほかのユーザーも同じようにスマホ端末を持っているという状況で、このユーザーの方々が、何かトレーニング、例えば、ジョギングなどをしたときのデータを、この図 11 内のサーバー A1 が収集する。

収集をした結果、「仮想トレーナー」というサービスで用いる「評価基準データ」をサーバー A1 が生成するということがもとの事例では記載されていた。さらに、サーバー A2 で利用者から送られてきたデータを受け取ると、この評価基準データに基づいて評価値を返すという処理をしており、サーバー A3 は、利用者の端末に、今の状況からすると、こういうトレーニング方法が望ましい、というトレーニングメニューを利用者の端末に返すという処理をしているという例が開示されていた。

したがって、AI 技術として考えるのであれば、図 11 中でサーバー A2 は、学習済みモデルが動いているサーバーと想定でき、サーバー A1 が学習済みモデルを作るサーバーで、サーバー A2 が学習済みモデルで評価値を返すサービスを実施するところで、サーバー A3 がこの学習済みモデルからの評価値をもらって、トレーニングメニューをユーザーに送るという処理をしているというふうに、元の想定を少しずつ読み替えると、この構成は人工知能の技術を使ったサービスシステムということになる。

つまり、サーバー A2 で、学習済みモデルが稼働しているというシステム構成として読み替えることが可能だと思われる。

そう考えた上で、このシステム構成のどこまでを請求項に記載するのかというのは難しいところではあるものの、一応、一つの例として、図 12 にしめすようなシステムのクレーム、トレーニング支援システムを

クレームに記載することができると思う。

「学習済みモデルが稼働するサーバ」のクレーム

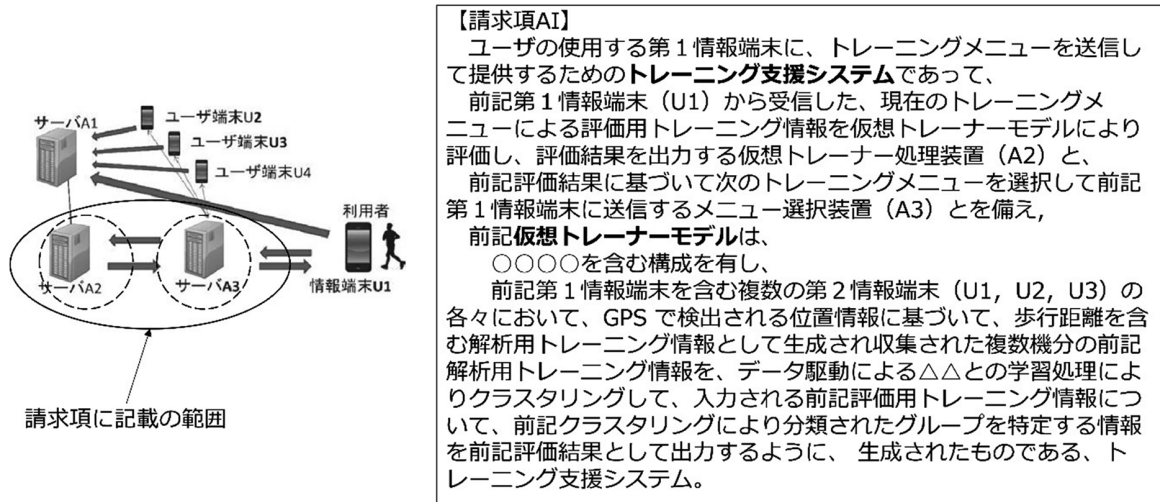


図 12 学習済みモデルが稼働するサーバとして起案したクレームの例

図 12 の中で、サーバ A2 が仮想トレーナーモデルという人工知能のモデルで評価結果を出すという処理をしている処理装置で、サーバ A3 というのはそれに基づいてメニューを送っている装置ということでクレームに書いた上で、サーバ A2 で稼働している仮想トレーナーモデルというのは、前述のとおり、「学習済みモデル」の構成として、モデルの構造を簡略に記載し、どのように学習がなされてそのモデルが生成されたかという記載をその後ろにすることで、仮想トレーナーモデルを限定することができる。

日本の特許庁の運用では、学習済みモデルの生成処理または学習済みモデルの構成（モデルの構造や入出力の構成を含む）に、新規性・進歩性があれば特許として認める、ということであるので、この例では、サーバ A1 側で実行されるモデルの生成処理のところに新規性・進歩性があるという状況で、サーバ A1 を直接クレームに記載しないとしても、この形のクレームとして特許査定を受けられるはずだと考えられる。

従って、少なくとも学習するところは、クレームのボディの中では間接的に記載しているに過ぎないので、どこかのプラットフォーマーが学習処理のサービスを提供している場合や、あるいは他の当事者が学習処理の部分だけを分担しているという場合も、形式上はクレームの技術的範囲の中に含まれると判断できると考える。

もっとも、サーバ A2 とサーバ A3 との処理を分けて記載しているものの、そもそも、そう分ける必要があるのかということも含めて流動的、ということもある。したがって、サーバ A2 と A3 での処理を、サーバ A1 での学習処理との関係で、サブコンビネーションのクレームとして記載すれば、問題は、すべて解決なのかということ、必ずしもそうとは言えない可能性がある。権利の行使に際しての交渉の対象という観点では、サーバ A1 側の処理が問題になるということも考慮が必要となる場合もありうる。また、サーバ A2 と A3 で実行される処理との関係では、間接侵害のアプローチというの考えなければならない場合もありうる、というふうにも思われる。

なお、上記のサーバとしてのクレームを、学習済みモデルのプログラムのクレームとして記載すると、以下の図 13 のようになると考える。

このプログラムクレームでも、サーバ A2 と A3 での処理ステップを、上記のようなステップの分け方で記載するのが妥当かということは、そもそも論としてあるが、一応こういうモデルのことを記載することで、この形式のクレームを作ること自体は可能だと考える。

(参考) 「学習済みモデルを用いたサービス提供のためのプログラム」のクレーム

【請求項AI-2】
 ユーザの使用する第1情報端末に、トレーニングメニューを送信して提供するための**トレーニング支援プログラム**であって、
 演算装置と記憶装置とを有するコンピュータに実行させたときに、前記プログラムは、
 前記記憶装置に格納された仮想トレーナーモデルに基づいて、前記演算装置が、前記第1情報端末(U1)から受信した、現在のトレーニングメニューによる評価用トレーニング情報を仮想トレーナーモデルにより評価し、評価結果を出力する仮想トレーナー処理ステップ(A2)と、
 前記演算装置が、前記記憶装置に格納されたトレーニングメニューと前記評価結果とに基づいて、次のトレーニングメニューを選択して前記第1情報端末に送信するメニュー選択ステップ(A3)とを実行させ、
 前記**仮想トレーナーモデル**は、
 ○○○○を含む構成を有し、
 前記第1情報端末を含む複数の第2情報端末(U1, U2, U3)の各々において、GPSで検出される位置情報に基づいて、歩行距離を含む解析用トレーニング情報として生成され収集された複数機分の前記解析用トレーニング情報を、データ駆動による△△との学習処理によりクラスタリングして、入力される前記評価用トレーニング情報について、前記クラスタリングにより分類されたグループを特定する情報を前記評価結果として出力するように、生成されたものである、**トレーニング支援プログラム**。

図13 学習済みモデルのプログラムクレーム

3. 学習済みモデルの特許による保護を前提とした実施行為の把握に対する一提案

もう一つ、私見ではあるものの、こういうサーバーがいくつかある中で侵害行為を追及するというときに、実施行為というものをどう考えるかというのは、間接侵害ということに限らず、非常に重要な論点と考える。

こういう問題のときには、特に、学習済みモデルを生成する基盤を提供するプラットフォームというよりも、むしろ、ユーザーへ直接的にサービスを提供する業者側が競業者というか競争相手になるということがほとんどであるので、多くの場合、権利行使、例えば、差止めの対象は、従来の観点でいえば、サービスを提供している当事者のサーバーであるという扱いになると想定されることが多かったと考える。

もっとも、下記の図14のとおり、一方で、こういうサービスのときにユーザー端末側から送られているのはコマンドとデータだけという場合が、ほとんどである。

さらに、話をむしろ複雑にしているのは、ユーザーへの直接のサービスの提供自体が、学習処理のサービスを提供するプラットフォームを提供する提供業者のシステム上で、(学習処理を実行するサーバーとは別のサーバー上での場合も含めて)稼働している場合があるということも想定しなければならないことであろう。

言い換えると、このようなサービス全体がプラットフォームの提供業者のシステムの上で動いているということがあり得ることになる。

先ほど技術の背景というところで申し上げたとおり、仮想化技術などの技術が進展しているので、権利行使を物理的なサーバーに対して行うというようなイメージで特許権を取得して権利行使を考えるのが適切といえるのか、いつまでも妥当性があるのかというのは、少し考えなければならないところとも思われる。

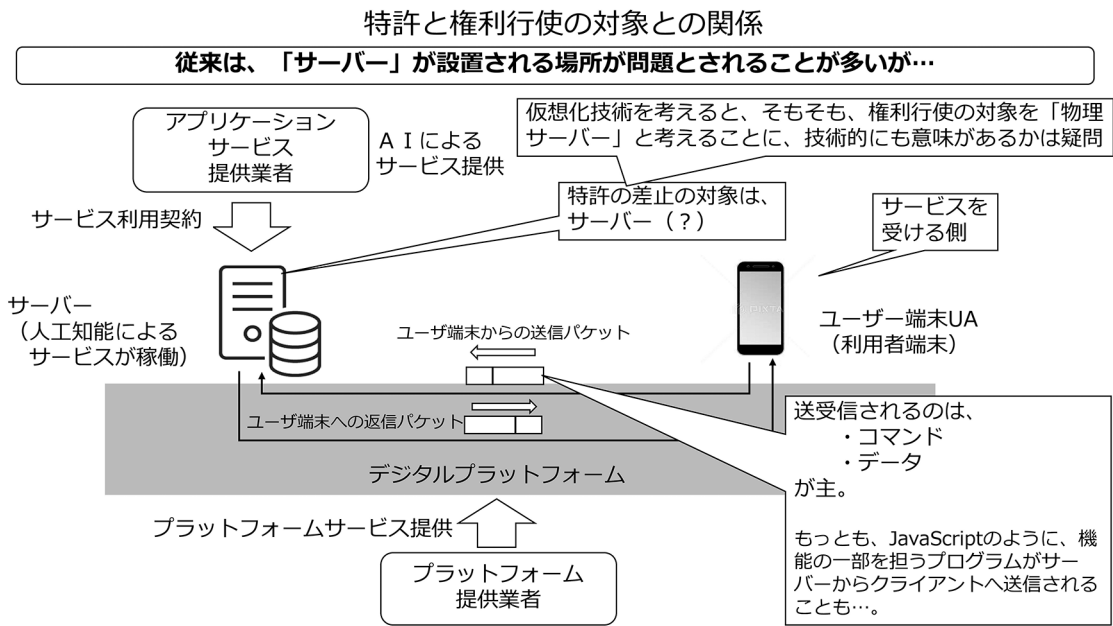


図 14 AI サービスがクラウドサービスとして提供される際の権利行使の対象

そこで、以下では、「プログラムの特許発明」についての実施行為を、典型的な「物の発明」の実施行為と対比して、図 15 にまとめる。

図 15 に記載の通り、物を譲渡したり、使用したりするということと、プログラムの特許法上は「譲渡等」という表現で、ややその範囲を広げたとも考えられる表現であるところ、プログラムの「譲渡等」ということ、あるいは、プログラムの「使用」ということを整理して考える。

物の場合の使用と譲渡、貸し渡すということに対応して、プログラムの場合には、どのように考えられるであろうか。「使用」という場合は明らかに自分が保有している、あるいは、自分が所有しているコンピュータに、著作権者だけではなく、そのプログラムについての権利者から使用を許諾されたプログラムをインストールして、自分の操作でそのプログラムを使うということが「使用」というように、一番典型的には把握されてきたと考えられ、これが「プログラムの使用」に該当すると考えられる。しかしながら、クラウドサービスの形態になったときに、プログラムについて、このときの「使用」という考え方を、そのまま延長して考えるのが妥当な場合だけではないのではないかというふうにも考える。

つまり、こういうクラウド側のサービスで、サーバーで何をしているかといえば、ユーザーが使う「機能を時間貸し」というか、排他的に、仮にあたかも複数のユーザーからの処理を同時に実行しているように見えたとしても、実際は時分割で排他的にサービスを提供しているというのが実態である。

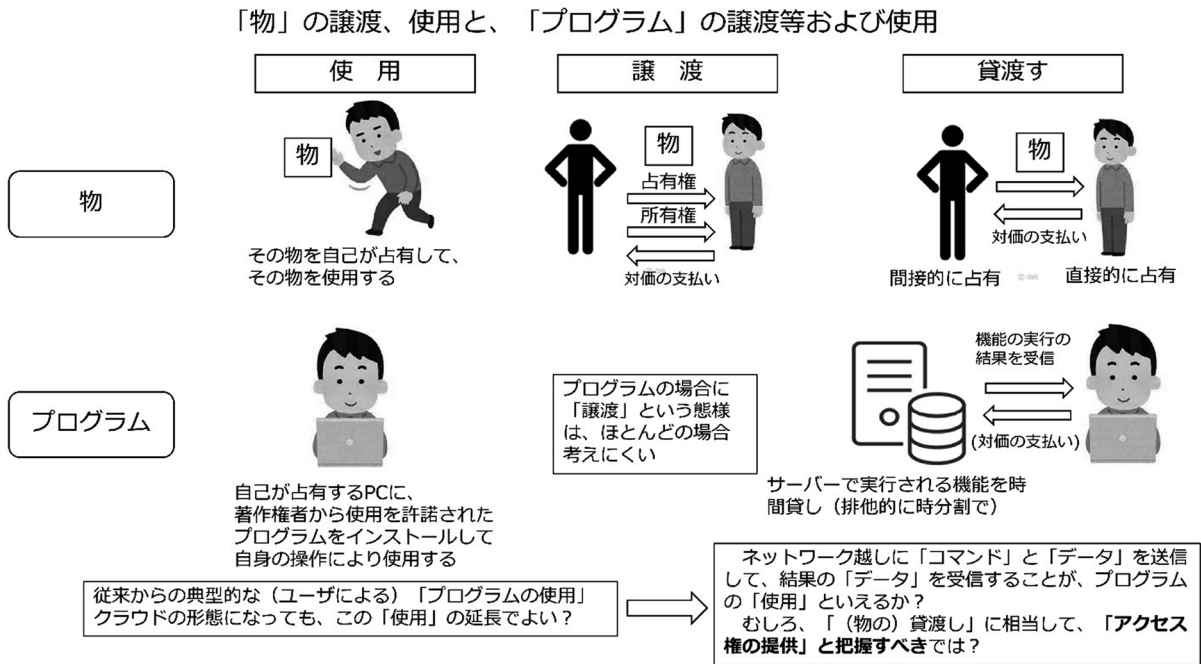


図 15 物の使用、譲渡と、プログラムの譲渡等および使用の関係

一方で、ユーザーがやっていることはアプリケーションプログラムと言いつつも、何かコマンドないしデータみたいなものを自身のコンピュータ側からクラウド側へ送って、処理がされた結果を単に受信している、ということである。そう考えると、上述した従来のプログラムの「使用」という概念とはかなり状況が違うのではないかと考えられる。むしろ、どちらかというところ、物のときは「貸渡し」と呼んでいたところ、この「貸渡し」に近い概念の方に把握する方がいいのではないかと考える。

そこで、図 16 に記載の通り、プログラムについて、クラウド型サービスを提供することは、「アクセス権の提供」と考えることが可能性としてはあり得る。

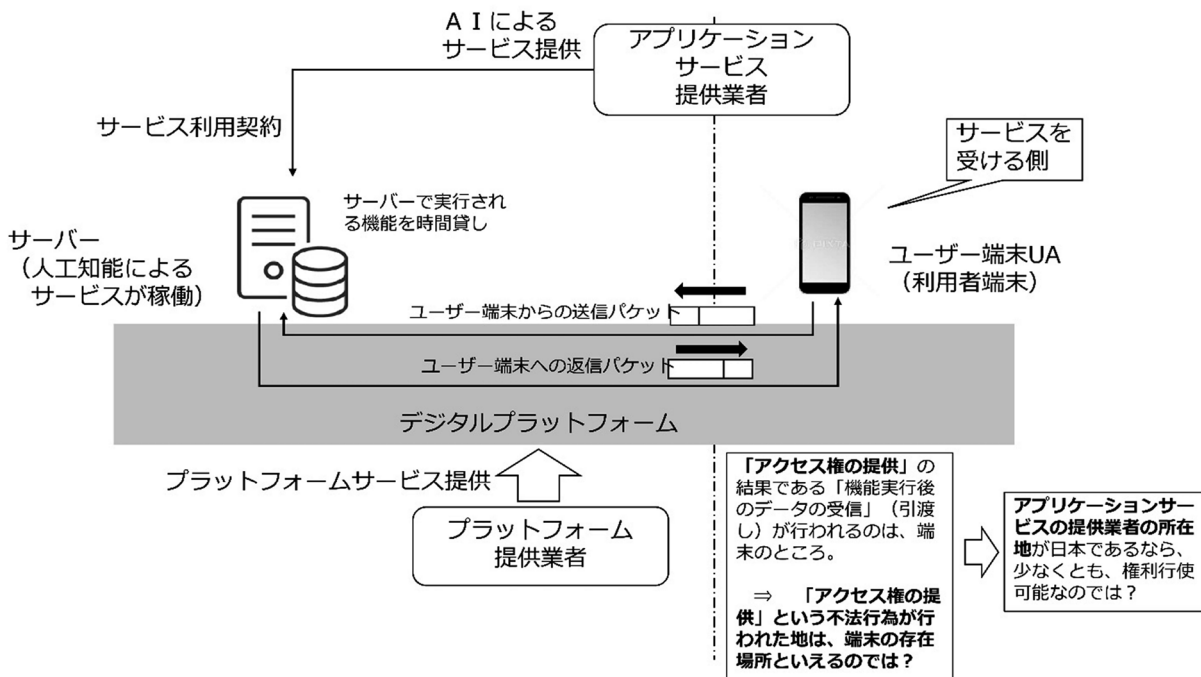


図 16 プログラムについてクラウド型サービスを提供する際の実施行為の把握

図15の「使用」との実施行為であると、あくまでユーザー側の行為であるのに対して、「アクセス権を提供する」というのは、当然サーバー側の管理者の行為であるので、そのような把握ができる場合もあり、むしろ、そのように把握する方が望ましい場合もあるように考える。

図16で言うと、アプリケーションサービス提供者がいて、サーバーを使って、人工知能のサービスを提供しているということになっているものの、これを「アクセス権の提供」という実施行為、つまり、実際のところ貸渡しに近いことなのだというふうにと考えると、こうすると、「貸渡し」で得られる結果が引き渡されているのは、あくまで端末のところであるので、アクセス権が提供されるという不法行為が行われたところ（不法行為地）を、端末が存在するところだ、というふうにと把握する可能性はあるのではないかと考えられる。

したがって、アプリケーションサービスの提供者の所在地が日本であるなら、サーバーがどこに設置されているかに関わらず、権利の行使については、日本で不法行為が行われていて、日本のサービスの提供者に権利行使を行うということもあり得ると考える。

不法行為が日本国内で行われたと言えるかという点では、たとえば、「物の譲渡の申出」という実施行為の例ではあるものの、被疑侵害製品が一審被告ウェブサイトに掲載されていた等の事情から、「譲渡の申出」行為又はその受領が日本国内で行われていたと判示して、外国法人に対する日本特許侵害訴訟につき、不法行為地に基づく我が国裁判所の国際裁判管轄を肯定した事例がある⁽¹³⁾。

また、少なくとも不法行為、権利が侵害されている行為が行われているというのであれば、SNSなどのサービスを提供するプラットフォームについては「プロバイダ責任制限法⁽¹⁴⁾」と呼ばれる法律がすでに存在しており、権利侵害されている情報を削除することを、プロバイダーの側に請求することが認められている。

この観点からすると、AI技術を使ったサービスの場合も、サーバー装置に権利行使をする、うんぬんというより、被疑侵害サービスの提供が運営されているプラットフォームに、そのサービスを削除するということを請求することが、ある条件の下では可能とするということも、権利行使の困難性を緩和するための一つの解決方法としてはあり得るのではないかと考える。

このような「クラウドサービス」における利害関係者と特許権との行使との関係を、図17に示す。

(13) 知財高判平成22年9月15日（平成22年（ネ）第10001号特許侵害予防等請求控訴事件）（「モータ」事件）

なお、本判決は、むしろ、国際裁判管轄の問題として言及される事例であるものと理解する。ただし、本稿では、「サーバー」自体が国外に存在するときに、日本国内での特許権侵害行為が発生しているといえるか、という問題の検討のために参考とさせていただいた。したがって、たとえば、「サーバー」の設置場所の問題ではなく、AIサービスの提供者自身が、外国法人である場合などは、別途、まったく異なる観点からの検討が必要になる。

もっとも、国際裁判管轄の検討においても、一般には、知的財産権の侵害は、「不法行為」の問題として把握され、準拠法などの判断にあたっては、一義的には「結果発生地」が判断の根拠と想定されることから、本稿において、AIサービスの提供者が日本法人である限り、日本の特許法に基づいて、当該AIサービスの提供者の行為に対して、特許権侵害の訴えを日本の裁判所に提起することは、単に、「サーバー」の設置場所が日本でないことにより、何ら妨げられるものではないと考える。なお、差止請求を実効あるものとするための方策については、本文中にも記載のとおり、現行の「プロバイダ責任制限法」が、クラウドベンダを介した「AIサービスの提供行為」にまで、適用されるものかは、別途検討が必要であろう。「プロバイダ責任制限法」では、発信者とは、特定電気通信設備の記録媒体に情報を記録し又は情報を入力した者となっているが、「AIプログラム」が「情報」に相当するのか、などが、そもそも論点となると考える。

(14) 平成十三年法律第三十七号 特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律

なお、商標権侵害の事例ではあるものの、東京地判平成18年4月26日（平成17年（ワ）第24370号プロバイダに対する発信者情報開示等請求事件）は、権利者が、ウェブページ（商標権侵害等が行われていたとするウェブページ）を開設していた者（発信者）の情報開示を、いわゆるプロバイダ責任制限法に基づき求めた点につき、これが認められた事案である。もっとも、この場合、レンタルサーバーの保有者に対する損害賠償責任については認められなかった。むしろ、この事案は、プロバイダ責任制限法が、その正式名称を見ても分かる通り、プロバイダ等の特定電気通信役務提供者の損害賠償責任の制限と発信者情報の開示について定めたものであることを、まさに示したものともいえる。したがって、逆に、権利者側の立場から見ると、①一定の要件を満たす特定電気通信役務提供者に対しては権利者からの損害賠償請求が認められないこと（同法第3条第1項、同法第4条第4項）、②侵害情報を停止してもらうためには権利侵害に関する情報を特定電気通信役務提供者に正確、適切に知らせること（同法第3条第2項）、③一定の要件を満たした侵害情報を発信した者の情報開示を請求できる権利を有する（同法第4条第1項）ということが主に規定されているといえる。

クラウドサービスにおける利害関係者

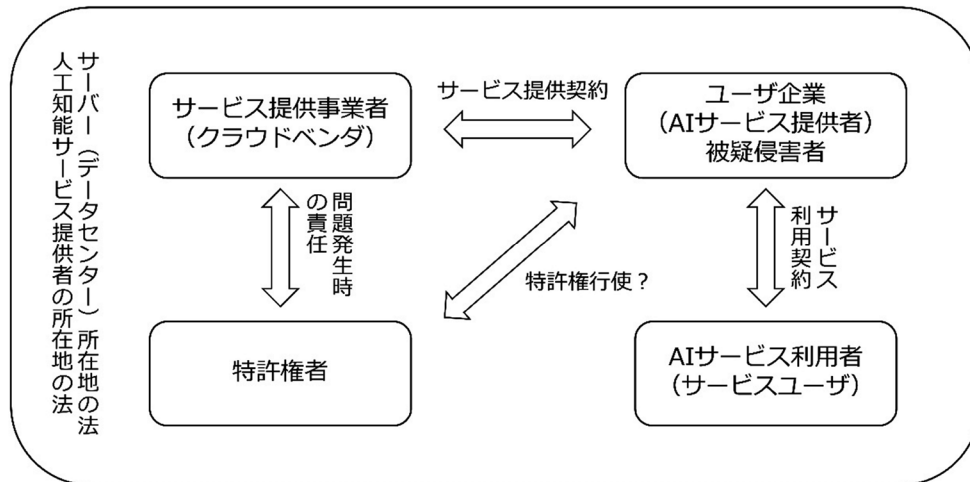


図 17 クラウドサービスにおける利害関係者と特許権の行使

すなわち、クラウドサービスを提供する事業者（クラウドベンダ）のサービスを利用してユーザー企業（AIサービス提供者）が、AIサービス利用者に、自身のAIサービスを提供しているという構図である。

このときに、特許権者の権利をAIサービス提供者が侵害している可能性がある場合、これまでは、権利行使の対象としては、AIサービス提供者が想定され、その結果、AIサービスのためのプログラムが稼働しているサーバーが、権利行使の対象と想定されてきたことになる。

ただし、図17に示すように、そもそも、被疑侵害者であるAIサービス提供者のプログラムがクラウドベンダのサービス内で稼働している場合は、特許権者は、クラウドベンダに対して、サービスの提供の停止を請求できることとなれば、少なくとも、どのサーバーでAIサービスプログラムが稼働しているかは関係なく、実質的に、差止めを行うことができることになる。一方で、クラウドベンダとしては、仮に、自身のサービス内で、特許権を侵害するようなプログラムが稼働している場合であっても、クラウドベンダ自身には、一定条件下で損害賠償等の責任がないことを規定するという意味で、「クラウドベンダ責任制限法」とでも呼ぶものが必要と考えられる。

以上を、もう一度整理すると、今の特許法では、物にプログラムが含まれていて、実施行為としては、「生産、使用、譲渡等」になっており、プログラム等である場合には「電気通信回線を通じた提供を含む」という規定ぶりになっている。つまり譲渡等については、プログラムの場合には「提供も含む」になっているわけで、この「電気通信回線を通じた提供」という行為を、プログラムをダウンロード販売するというだけでなく、ネットワーク上に存在するプログラムの「アクセス権を提供する」ということを含む、という解釈の仕方ができれば、いろいろな意味で、権利行使の局面で、柔軟な解釈の可能性が出てくるのではないかと考える次第である。

ところで、「電気通信回線を通じた提供」というのはダウンロード販売のことだけを言っているととらえられがちであるものの、平成14年法改正で「電気通信回線を通じた提供」との表現が入ってきた当時のいわゆる「逐条解説」⁽¹⁵⁾には、『ネットワークを通じたプログラム等の送信や、ネットワークを通じたASP（Application Service Provider）型サービス』について、送信者やサービス提供者の手元にも元のプログラム等が残るといふ有体物にはない性質であるとの言及がある。特に、後者は基本的にクラウドサービスに相

(15) 平成14年法律改正（平成14年法律第24号）逐条解説

https://www.jpo.go.jp/system/laws/rule/kaisetu/sangyozaisan/document/sangyou_zaisanhou/h14_kaisei_1.pdf

当すると考えられ、そのようなサービスの形態がありうるものが、記載されていたことになる。とすると、こういうクラウド型のサービスの形態が発明の実施行為に含まれることも改正時には考慮されていた、ともいえるように考える。

その意味で言うと、前述した解釈は、あくまで可能性としてはあるけれども、平成14年法の改正の時点で、ある程度想定されていたのではないかと考えられる。

しかも、観点としては、若干それる懸念もあるものの、薬機法⁽¹⁶⁾という法律があって、「製造販売」ということについて、この法律上は、医療機器プログラムについては、「電気通信回線を通じて提供すること」が含まれるという、ほぼ特許法の文言その通りのようなことが規定されている。

現在、医療機器というものは、医療機器というハードウェアだけではなくて、医療機器プログラムそのものも、承認の対象になるというふうな運用が広がっている。最初のうちは、医療機器プログラムについて、特許法の対象の歴史的経緯と同じように、媒体でプログラムを提供する、ダウンロード販売で提供することが、主たる対象であった。ところが、特にAI医療機器プログラムというようなクラウド型のサービスというものが普通に提供されるようになってきており、薬機法の世界では、「医療機器プログラムの使用权（アクセス権）の提供」というのは「医療機器の製造販売とみなす」、つまりアクセス権を提供することも製造販売だという扱いが、実はもう行われているところである。これは、別の法律の中での扱いではあるが、特許法でも、同様の扱いとすることは、あながちあり得なくはない運用なのではないかと考える。

そのようなことを前提として、実施行為を「プログラムのアクセス権の提供」というような把握の仕方をする、先ほど、図12において、具体例ということで挙げたものについて、仮想的にクレームを作ってみて、間接侵害がどういうふうな解釈できるかを検討してみる。

すでに、「トレーニング支援システム」のクレームと、「トレーニング支援プログラム」のクレームについては、図12、図13に示したので、以下の図18に、「トレーニング支援方法」のクレームを示す。

上述した通り、サービスを提供しているサーバー群（A2とA3）のクレーム、サーバー群のプログラムのクレーム、サーバー群で実行される方法のクレームを仮想的に作成することが可能である。

（参考）「学習済みモデルを用いたサービス提供のための方法」のクレーム

【請求項AI-3】
 ユーザの使用する第1情報端末に、トレーニングメニューを送信して提供する処理を、演算装置と記憶装置とを有するコンピュータに実行させるトレーニング支援方法であって、
 前記記憶装置に格納された仮想トレーナーモデルに基づいて、前記演算装置が、前記第1情報端末（U1）から受信した、現在のトレーニングメニューによる評価用トレーニング情報を仮想トレーナーモデルにより評価し、評価結果を出力する仮想トレーナー処理ステップ（A2）と、
 前記演算装置が、前記記憶装置に格納されたトレーニングメニューと前記評価結果とに基づいて、次のトレーニングメニューを選択して前記第1情報端末に送信するメニュー選択ステップ（A3）とを有し、
 前記仮想トレーナーモデルは、
 ○○○○を含む構成を有し、
 前記第1情報端末を含む複数の第2情報端末（U1, U2, U3）の各々において、GPSで検出される位置情報に基づいて、歩行距離を含む解析用トレーニング情報として生成され収集された複数機分の前記解析用トレーニング情報を、データ駆動による△△との学習処理によりクラスタリングして、入力される前記評価用トレーニング情報について、前記クラスタリングにより分類されたグループを特定する情報を前記評価結果として出力するように、生成されたものである、トレーニング支援方法。

図18 学習済みモデルを用いたサービス提供のための方法のクレーム

(16) 昭和三十五年法律第百四十五号 医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律、厚生労働省医薬・生活衛生局医療機器審査管理課 渡利 彰浩「最近の医療機器行政の動向について～ AI技術による医療機器開発の留意点等～」<http://www.pref.kyoto.jp/yakumu/documents/koueniti.pdf>

そして、各カテゴリのクレームについて、「AI 学習済みプログラムを使用してサービスを提供している者（サーバー A2 の運営者）」（この者が「学習済みプログラムを生成した」と想定する）が独立している場合に、この者に対して、これらのクレームにより、間接侵害の適用の可能性を検討すると、以下の表 1 の通りとなると考える。

たとえば、間接侵害の規定（特許法第 101 条第 5 号）上で考えると、AI プログラムである学習済みプログラムを前提とすると、方法クレームについては、それを方法の使用に用いるものだと考えて、間接侵害の規定上は、「その方法の発明が特許発明であること」という前提の下で、「その方法の発明の実施に用いられる」というのは、これは「支援方法の実施に用いられる」であるから、支援方法に用いられることを知りながら業として譲渡等、貸渡しをする行為、すなわち、「アクセス権を提供する行為」が、間接侵害行為になるという把握の仕方ができる可能性がある。

表 1 のとおり、この場合、筆者としては、条文へのあてはめという点で、「方法のクレーム」による間接侵害との解釈が、素直な解釈とも考える。

そのように仮定するのであれば、サーバー A2 と A3 が複数当事者により実施されている場合（状況により、A1, A2, A3 の間で複数の実施者がいる場合）に、直接の侵害者がいないということになってしまうことをおくこととし、直接侵害行為がなくても間接侵害行為を問うことは可能だというふうに考えられるのであれば、このような形式のクレームで間接侵害の適用を想定できるのではないかと考える。

表 1 各カテゴリのクレームと間接侵害の可能性

サーバー群のクレーム	サーバー群プログラムクレーム	サーバー群の方法クレーム
二 特許が「人工知能サービスを提供するサーバー（群）」の発明についてされている場合において、その「人工知能サービスを提供するサーバー（群）」の生産に用いる「AI プログラム（A2）」（日本国内において広く一般に流通しているものを除く。）であつてその発明による課題の解決に不可欠なものにつき、その発明が特許発明であること及びその「AI プログラム（A2）」がその発明の実施（すなわち、サーバー群の使用（?））に用いられることを知りながら、業として、その生産、譲渡等若しくは輸入又は譲渡等の申出をする行為	二 特許が「人工知能サービスを提供するサーバー（群）」のプログラム」の発明についてされている場合において、その「人工知能サービスを提供するサーバー（群）」のプログラム」の生産に用いる「AI プログラム（A2）」（日本国内において広く一般に流通しているものを除く。）であつてその発明による課題の解決に不可欠なものにつき、その発明が特許発明であること及びその「AI プログラム（A2）」がその発明の実施（すなわち、サーバー群としてのアクセス権の提供）に用いられることを知りながら、業として、その生産、譲渡等若しくは輸入又は譲渡等の申出をする行為	五 特許が方法の発明についてされている場合において、その方法の使用に用いる「AI プログラム（A2）」（日本国内において広く一般に流通しているものを除く。）であつてその発明による課題の解決に不可欠なものにつき、その発明が特許発明であること及びその「AI プログラム（A2）」がその発明の実施（すなわち、「方法」の使用）に用いられることを知りながら、業として、その生産、譲渡等（すなわち、アクセス権の提供）若しくは輸入又は譲渡等の申出をする行為
1) 日本では、知財高裁の一郎事件判決により、コンピュータにコンピュータプログラムをインストールする行為は、「生産」に当たるとされている。 2) AI アプリケーションサービス提供者側は、「サーバー群」の「アクセス権の提供」はしておらず、「サーバー群の使用」をしていると把握すべきかが、論点。	複数のサーバーで分散して実施されるプログラムの一部である「AI プログラム（A2）」（学習済みモデルのプログラム）は、『「人工知能サービスを提供するサーバー（群）」のプログラム』の生産に用いるといえるか、が論点。 すなわち、複数のコンピュータに分散して、クレームの各機能ステップに相当するプログラムがインストールされ、ネットワークで相互に連携することはプログラムの「生産」か？	「その発明の実施」とは、「方法の発明の実施」すなわち「方法の使用」であるが、サーバーが分かれていることにより、使用している主体（サービス提供者）が複数であるなら、直接的に、「当該方法を使用する者」が存在しないという前提で、「アクセス権の提供」の行為について、間接侵害のみを認めるという構成になる。

もっとも、実際、実務上このようなクレームの構成が具体的事案に対して最適であるか、あるいは、その他の可能性として、サーバーとしてのクレーム、物としてのクレームの方がいいのかということについては、検討課題だとは想定されるものの、この「アクセス権の提供」ということを提供者側の実施行為と考えることで、侵害行為と認定できる範囲が、ある程度柔軟に考えられる余地が出てくるのではないかと考える。

4. 「学習用データ」の保護

4.1 「データ」と知的財産としての保護

前述した通り、たとえば、「データ駆動型社会」というような言葉も最近使われるようになってきている。

ところが、そこでの「データ」という言葉を聞いたときに、それがどういうものなのかというイメージが、実は各人各人で、未だ、かなりばらばらなのではないかと考えている。従って、「データ駆動型社会」において、データが、どれくらい重要なのかということに対する認識も、たぶん人によってばらばらなのが現状だと考えられる。

結果として、技術としての「データ駆動型人工知能」といったときに、「データ」とは何なのかということに対してのイメージも、実は決して、技術上の共通認識といえるものが存在する状態ではないと予想される。認識がばらばらなので、知的財産としてどう保護するのがいいのかとか、データにかかわる技術の特許をどのように考えるべきかというようなことも、各人によって、実は今のところ、まだ意識や印象が違っていると予想される。

例えば、データを利用するための契約について解説をしている書籍などでは、「情報」というものは電磁的に記録された以外のものも全部含んでおり、その中で特に電磁的に記録されたものを「データ」と呼ぶ、というような記載になっている場合がある。

したがって、この場合は、情報が上位概念で、その下位概念の中にデータがあるというようなイメージとして把握されていると考えられる。

また、最近、データ解析、データサイエンスがいろいろなところで取り上げられており、統計学の専門家がテレビや書籍の中などで、さまざまな解説をされる場合もある。このときには、おそらく、これまでの統計学がそうであったからと考えられるが、目的を設定してデータを準備して何らかの予測なりを、コンピュータを利用して実行し、その結果の解釈を、統計学の専門家の知見を利用して、意思決定をしていくということが重要だと紹介されている場合もある。

このとき、この統計解析の状況において、最初の「目的の設定」とか「課題の構築」などは、人間がやることとして、そこから話が始まっているという場合が多い。そのような背景もあってか、たとえば、「データというものは、仮説を検証するために用いるものであって、仮説を構築するのに用いるものではない」というような指摘がされている場合もあるようである。

もっとも、「データ駆動型人工知能」の技術的背景でも説明した通り、データ駆動型人工知能の「学習済みモデル」というのは、まさに「データから作られた仮説」というべきものである。したがって、仮説をデータから構築して、その仮説であるモデルが、本当に正しいのかを技術的に検証することも別のデータで行っているというのがデータ駆動型人工知能ということになる。

一方で、データ駆動型の人工知能を開発したり、あるいは、しようとしている研究者の感覚について伺った話としては、生データ、例えば、何月何日に何が売れたというデータが時系列で並んでいるのが「データ」であって、それに対してだんだん内容を集約して行って、最終的に、(販売実績として評価するには、どれだけの個数のデータが必要なのかは、おいておくとして)「新製品の方が従来品より人気がある」のように、だんだんと、データから抽出されてまとまっていくというのが「情報」であり、むしろ、「データから情報が生まれる」という言い方がされる場合もある。つまり、「初めにデータありき」「データから情報が生まれてくる」というのが、こういうデータ駆動という考え方の一番基本的なところなのだと考える。

データ駆動型人工知能における技術要素の重要度

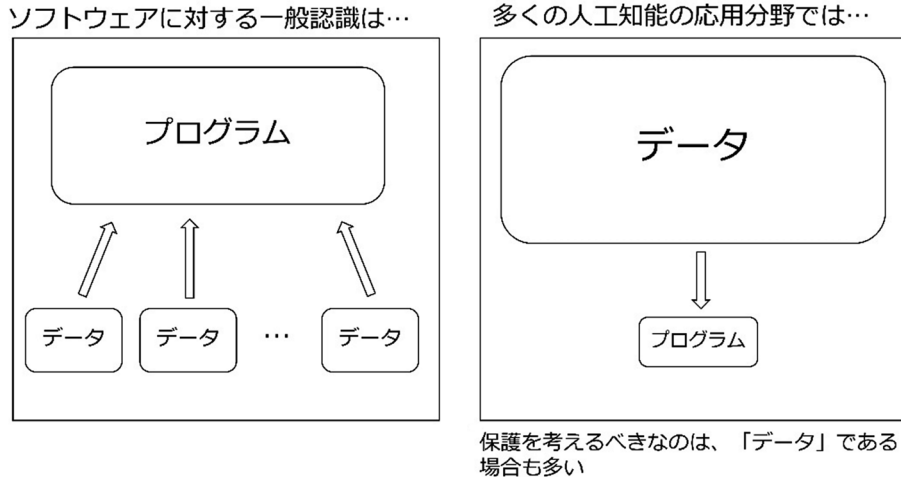


図 19 データ駆動型人工知能技術における「プログラム」と「データ」

ということで、図 19 に示すように、ソフトウェアの技術に対する一般的な認識というのは、プログラムがすごく重要なものとしてあって、それにデータが入力されてプログラムが動いて何か処理がされているという、イメージが多いと思う。しかし、現代の人工知能の分野では圧倒的に、まずデータがあって、そのデータに基づいてプログラムが動いているというのが実情だといえる。したがって、「保護を考えるべきなのはデータである場合も多い」、あるいは、もっと言えば、「データである場合の方が多い」と言った方が正しいかもしれない。

データ駆動型の人工知能では、現実の社会から「データ」が取得されるという場合がほとんどであり、これは、「リアルワールドデータ」と呼ばれることも多い。強化学習と呼ばれる方法で、コンピュータ上で仮想的に学習するということは、あり得るものの、多くの場合、強化学習においても現実の世界からデータが取得される。

さらには、デジタルツイン (Digital Twin) と呼ばれ、現実の世界から収集した様々なデータを、まるで双子であるかのように、コンピュータ上で再現する技術も盛んに研究されている。もっとも、この場合でも、コンピュータ上では、収集した膨大なデータを元に、限りなく現実に近い物理的なシミュレーションを可能とすることとされており、「リアルワールドデータ」の重要性が低下するというものでは決してないものと理解する。

なお、リアルワールドからデータを収集してくる部分やモデルへの入力データの処理 (前処理等) というところ⁽¹⁷⁾は、実は技術的にも特許的にも大変重要な部分であって、特許を取得することを考えるときに、「学習済みモデル」が、中核的で重要なところの一つと言及した際に、「重要なのは、そこだけではない」との筆者の意見を述べたのは、ここに理由がある。

もっとも、こと「データ」に関して言うのであれば、現状では、知財の世界ではデータは、コンピュータが動作するための単なる「情報」とか「パラメータ」のように扱われることが多く、その点は、以下に説明するとおり、課題が多いと認識する。

(17) 酒井 将行「研究開発型企業のための知的財産保護—技術シーズの事業化のために—」パテント 2021 Vol.74 No.6

この論文では、音声認識技術などを例にとり、「生データ」から「学習用データ」を作成する際に使用される「雑音除去技術」のような技術は、データに対する「前処理の技術」に相当して、「学習を実行するためのソフトウェア」や「学習モデルの構造・構成」より、むしろ、技術の有効な期間が長くなる傾向にある、ことを指摘した。また、学習処理は外部からは何い知れない環境で実施されるとしても、「前処理の技術」は、「使用フェーズ」でも使用されることから、侵害の摘発性の観点からも、特許として意義がある可能性がある、との指摘も行った。

4.2 データの特性、データの階層性

データ駆動型人工知能の知的財産保護という点をデータの側で考えると、重要なところは、「データ」と一言で言っても実は階層的な構造を持っているということがある。以下の図 20 は、このようなデータの階層構造を示す。図 20 では、横軸にデータが有効な時間スケールを取り、縦軸にデータの量をとっている。

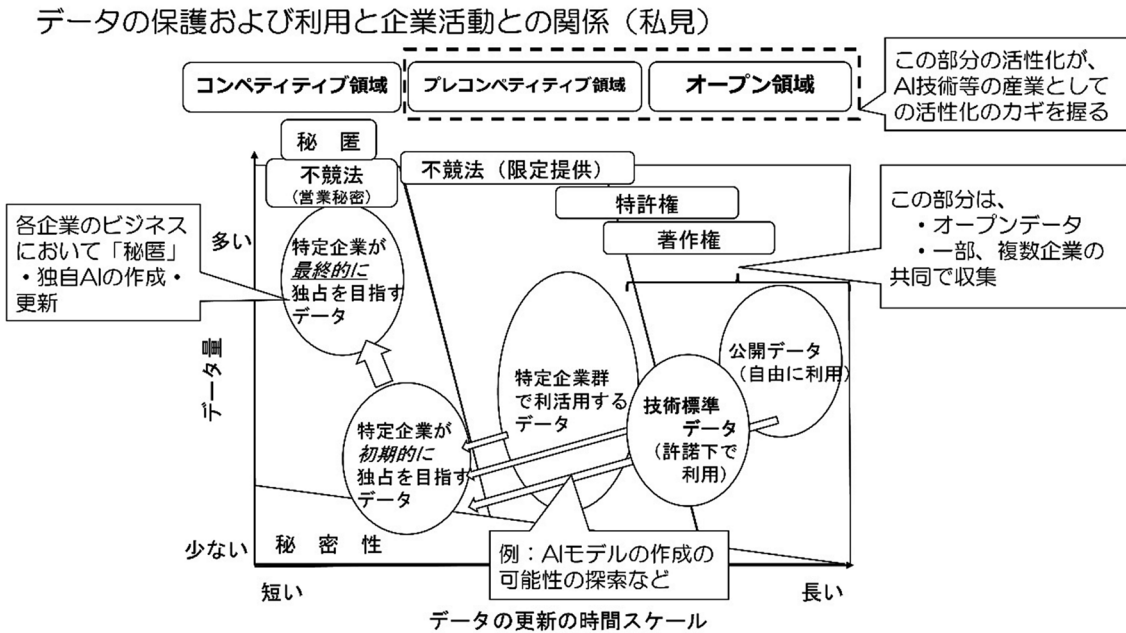


図 20 データの秘匿の程度とデータ利用・データによる競争力の関係

この図 20 の右側に示すように、今現在でも、すでに公開されて自由に使えるデータが存在し、あるいは、今後も、たとえば、技術標準のような形で公開されて使われていくデータというものもある。

さらに、この図 20 の中ほどに示すように、完全に公開されてはいないけれども、特定企業間だけで共有して使用するデータというものもある。さらには、この図 20 の左側に示すように、最終的にはある企業が、自分のビジネスのために独占していきこうとして、データを集めていく部分もある。

ここで重要であるのは、このような「秘匿すべきデータ」は、一般には、公開されて広く使用されるデータに比べると、有効な期間、言い換えると、アップデートされるスパン、つまり、更新される時間的な間隔が、短い傾向にあることである。逆に言えば、このようなスパンが短いゆえに、特定の企業にとっては、他社に対する競争力となり得る、ともいえる。長期にわたって変わらないデータというものは、いずれは、他にも獲得する競合他社が生まれてくると想定されるからである。

このように、知的財産として自社の強みとする部分を、「長期にわたり固定的な領域」ではなく、「常時、変更・変更される領域」に置くということは、いわゆる「オープン・クローズ戦略」として、「物」に対する特許権など他の知的財産でも行われることである。たとえば、「Intel Inside」戦略などと呼ばれる場合も、CPU の仕様を、比較的短い周期で定期的に変更することで、下流側のマザーボードや組み立てられた PC、周辺機器などを、自己のコントロール下に置こうとする戦略とみることもできる。ただし、「ソフトウェア」そして、さらには「データ」では、このような傾向がより顕著に表れるといえるであろう。

したがって、このようなデータの階層構造が、現実存在する状態となったとして、特許出願という観点から考えれば、例えば、明細書の開示要件としては、公開データ等を利用して開示を要求することも想定されうる。公開データにより証明されうる「人工知能モデル」の「モデルの構成」や「学習方法」の発明などであれば、開示要件としては、第三者の実施可能性を担保することも可能となる。一方で、真にビジネスとして他社に対する競争優位を有する状態の「人工知能モデル」そのものについては、秘匿領域にある

データにより学習されるものと考え、特許出願による公開と、自社技術の秘匿という一見矛盾する要請にも応える道が存在するといえよう。

一方で、技術開発の促進、より大きく言えば、産業の発達への寄与という観点でいうのであれば、いわゆるオープン領域と呼ばれているところに、自由に利用できる公開データがあることを前提にして、企業間の連携でいろいろな技術開発が行われる領域として、プレコンペティティブな領域というものが想定される。つまり、このプレコンペティティブな領域では、いくつかの企業で、実際に、データ駆動型人工知能の実現可能性を試してみることで、POC（Proof Of Concept）を実証するということが必要になる。その上で、POCの後には、実際に各社で独自にビジネスを展開するコンペティティブな領域というのに分かれていくと考えられる。

現在、オープン領域に対応しているデータについての知的財産権としては、著作権、特許権がある。それから、プレコンペティティブな領域については、不競法では、「限定提供データ」という制度が設けられた。完全に秘密な領域であれば、営業秘密ということで保護することになる。

データ駆動型の人工知能技術であれば、この図で、プレコンペティティブ領域というところで人工知能のモデルが本当に実用化できそうかを、いくつかの会社が共同で探索し、その後、最終的に、各社独自のビジネスに対応した人工知能技術を、この秘匿すべきデータ領域で開発する、という流れが産業上ないしは産学連携上で必要と考える。

4.3 既存の知的財産権によるデータについての保護

(1) 秘匿領域（コンペティティブ領域）でのデータの保護

上述のとおり、データ保護の観点から「秘匿領域」と呼ぶべき領域では、「学習用データ」は、「営業秘密」ということで保護される。この場合には、「学習用データ」について、「秘密管理性」「有用性」「非公知性」が要求されることになる。

「秘密管理性」については、いわゆる「アクセス制限」の問題となるものの、不競法の平成30年改正により、技術的制限手段についての「保護対象」に、「情報（電磁的記録に記録されたものに限る。）の処理及び記録」が追加されており（不競法第2条第8項）、秘匿する「データ」についての「技術的制限手段によるアクセス制限」が、権限の無い者による不正取得や不正使用に対抗するためだけでなく、転得者に対抗するためにも「学習用データ」について重要と考えられる。

「有用性」については、少なくとも、すでに、その「学習用データ」から、「人工知能モデル（学習済みモデル）」が生成され、一定の事業用途に使用されている状態が想定されることから、肯定されるものと考えられる。

そして、営業秘密である「学習用データ」については、不正取得者自身による使用だけではなく、不競法第2条第1項第10号において、以下の通り規定されている。

十 第四号から前号までに掲げる行為（技術上の秘密（営業秘密のうち、技術上の情報であるものをいう。以下同じ。）を使用する行為に限る。以下この号において「不正使用行為」という。）により生じた物を譲渡し、引き渡し、譲渡若しくは引渡しのために展示し、輸出し、輸入し、又は電気通信回線を通じて提供する行為（当該物を譲り受けた者（その譲り受けた時に当該物が不正使用行為により生じた物であることを知らず、かつ、知らないことにつき重大な過失がない者に限る。）が当該物を譲渡し、引き渡し、譲渡若しくは引渡しのために展示し、輸出し、輸入し、又は電気通信回線を通じて提供する行為を除く。）

上記の行為も、不正競争行為と規定される点に注意が必要と考える。すなわち、「学習用データを不正に

取得した者」が、当該データを用いて、学習済みモデルを生成し、当該学習済みモデルを組み込みプログラムとして搭載した機器、当該学習済みモデルを組み込んだプログラムを譲渡等する行為も、不正競争行為となると考えられる。学習済みモデルは、「プログラム」と把握されるところ、不競法でも、第2条第11項で、『この法律にいう「物」には、プログラムを含むものとする。』と規定されているからである。

なお、ここでも、上述した特許法における実施行為と同様に、「電気通信回線を通じて提供する行為」を、「アクセス権を提供する行為」と把握するのであれば、不正取得した「学習用データ」を使用して生成した「学習済みモデル」をクラウド環境下においてサーバー側で使用してサービスを提供する行為も、不正競争行為と把握できるものと考えられる。

(2) プレコンペティティブ領域でのデータの保護

上述した通り、データ駆動型の人工知能技術の開発であれば、プレコンペティティブ領域では、まさに、「他者との共有を前提に一定の条件下で利用可能な情報」として、「学習用データ」を利用することになるので、不競法の平成30年改正による「限定提供データ」としての利用となると想定される。

その場合、アクセス権のない者による取得・使用・開示や、取得時に悪意の転得者の使用や開示が不正競争行為とされたことで、たとえば、一旦、コンソーシアムなどが組成された後に、共有されているデータについての保護が、積極的に図られることとなっている。

ただし、重要なもう一つの側面としては、データ駆動型人工知能の技術開発のために、そもそも「コンソーシアム」などの「データを共用する体制」をいかにして組成するのか、ということが、本質的と考える。

その意味では、不競法第2条第1項第14号にいう「限定提供データを保有する事業者（以下「限定提供データ保有者」という。）からその限定提供データを示された場合」、すなわち、「限定提供データへのアクセス権のある者」による「学習用データ」の使用や開示、及びその使用により得られる成果（たとえば、「学習済みモデル」）に関する権利の帰属など、むしろ、「学習用データ」を共有することになる、当事者間での「契約」の存在が、極めて重要ということになると考える。

たとえば、第14号の規定は、以下のとおりである。

十四 限定提供データを保有する事業者（以下「限定提供データ保有者」という。）からその限定提供データを示された場合において、不正の利益を得る目的で、又はその限定提供データ保有者に損害を加える目的で、その限定提供データを使用する行為（その限定提供データの管理に係る任務に違反して行うものに限る。）又は開示する行為

特に、「限定提供データへのアクセス権のある者」については、いわゆる「図利加害行為」の範囲が、「その限定提供データの管理に係る任務に違反して行うものに限る。」との限定がされている。

このことは、言い換えれば、コンソーシアムを組成したメンバーにとっては、共有する「限定提供データ」を利用した「学習済みモデル」の生成行為に一定の自由度があると、解釈することもできる⁽¹⁸⁾。

一方で、法律上は、このような自由度のある解釈が可能であるがゆえに、プレコンペティティブ領域での「データ利用」に必要以上の制限が、コンソーシアムなどの契約において規定されてしまうと、そもそも、「新

(18) 経済産業省「限定提供データに関する指針」（平成31年1月23日（最終改訂：令和4年5月））では、「なお、取得したデータを使用して得られる成果物（データを学習させて生成された学習済みモデル、データを用いて開発された物品等）がもはや元の限定提供データとは異なるものと評価される場合には、その使用、譲渡等の行為は不正競争には該当しない。」（第22頁）と記載されており、この記載からも「自由度」が認められていると解釈できる。もっとも、逆に、そうであるがゆえに、「学習用のデータセット」を作成する行為や、「学習済みモデル」を生成する行為を、「データ共有」の契約の中でどのように扱うかが、重要になると考える。https://www.meti.go.jp/shingikai/sankoshin/chiteki_zaisan/fusei_kyoso/pdf/20220517_2.pdf

規な人工知能についてのPOCを実現する」ことを目的としている場合は、むしろ、弊害となる。これは、「データ駆動型人工知能技術」である以上、実際にデータを集積して、学習処理を実行してみないことには、実用化に資するような、いかなる「学習済みモデル」を生成できるかが、事前には確定できないゆえに、コンソーシアムでの「データ共有」を実行しようとしたことと、そもそも、矛盾してしまうことになると考えられるからである。

とすると、たとえば、契約において、「データの利用目的」を規定する場合などには、そもそも、コンソーシアムでの「開発項目」について、技術的観点からの「事前の開発ステージの設定」を行うことが一つの解決方法とも考えられる。たとえば、「医療ビッグデータ」をコンソーシアムで共有して、医療上の有用性を有するような「学習済みモデル」のPOCを実施するような場合を考えてみる。

「利用目的」を設定するという観点からは、一つの可能性は、たとえば、コンソーシアムメンバー間で、「対象疾患」を限定しておくなども考えられるであろう。

ところが、データ駆動型人工知能の技術開発ということであれば、このような「従来の観点にとらわれた利用目的の設定」が必ずしも適切でない場合も想定される。

たとえば、『白い巨塔』という小説とドラマ、映画についてはよく知られているものと思う。ここで、主人公は、財前教授であり、物語の終盤で彼は亡くなってしまうことになるものの、実は、そのドラマの視聴者がどの年代であるかによって、財前教授の病名の記憶が、たぶん異なっているという話がある。筆者などにとっては、財前教授を演じられた俳優としては田宮二郎さんが、一番印象が強く、その病名は「胃がん」であったと思う。唐沢寿明さんが、テレビドラマで主演されておられたときには、亡くなる原因となったもともとのがんは「肺がん」ということであったと思う。さらに、岡田准一さんが、やはり3夜連続ドラマで主演されたときは「すい臓がん」が主人公のがんの病名であったと思う。

ここで、あまり疑問に思わないところかもしれないが、「がん」というのは、なぜこういうふう臓器の名前で呼ばれるのかということを考える。これは、実は、がんを分類するということをしたときに、人間が持っていた情報（データ）が、その「がん」の病巣の存在するのが、人間の目を見てどこだと分かるかということに限られていたので、こういう病名が付いているということなのだと思える。ただし、現在は、その患者の遺伝子を解析して、その遺伝子ごとにその患者に一番適した薬が何であるかというようなことが判断されるという時代になりつつある。

つまり、「がん治療薬は臓器ごとから遺伝子ごとに選択する時代」になってきていることになる。要は、どういうデータが集まってきて、そのデータをどういうふう解析できるかという技術が変わってくると、つまりデータが、あるいはデータの解析技術が、あるいはそのデータに対する知見が変わってくることによって、我々がそれまでやっていたことのスキームというか、社会的な物事のやり方・認識の仕方自体が変わってしまうという一つの例だと思える。あまり意識はないかもしれないものの、実は、ほかの分野でも同じようなことは十分起こり得るのではないかというふう考えられる。

そう考えると、先ほどのコンソーシアムでの「共有データ」の「利用目的」について、状況によっては、従来の慣習に従って「対象疾患」で限定するということは、そもそも、「データ駆動型人工知能技術」の新規開発の芽を自ら摘んでしまっている可能性すらあることになる。

とすると、「利用目的」については、むしろ、それが実用化されたときに、「どれぐらいオープンに利用することが想定されるか」というような観点から規定した方が望ましい場合もあることになる。たとえば、「医療ビッグデータ」の集積により、新たな疾患の診断手法（たとえば、先ほどのがんの例であれば、単に「がんであるか」だけではなく、それが、どのようなタイプの「がん」に分類できるのか、のような手法）と、そのような新規な手法に基づく個別の「治療手法」までの開発が想定されるのであれば、「診断方法」については、基本的には、「オープンな利用」を想定して共同で開発利用することを前提とした「限定提供データ」の利用目的の範疇とし、個別の「治療方法（例えば、特定の薬剤の治療応答性など）」については、個別企

業による「クローズ領域での開発」というようにして契約を締結をすることも、十分に想定される。

(3) オープン領域でのデータの保護

現在、オープン領域に対応しているデータについての知的財産権としては、著作権、特許権があるといえる。

以下の図 21 にも記載のとおり、平成 30 年の不競法改正時の経産省の解説⁽¹⁹⁾では、「公開データ」に対する知的財産権としては、典型的に「著作権」が挙げられている。

ここで、重要な課題は、このオープン領域からプレコンペティティブ領域に移行していく道筋というのが、今のところ、十分に整備されているとは言えないことである。この道筋の部分の活動が活性化すると、AI 技術とか、それを使った技術開発が進んでいくのではないかと考える。

- データは複製・提供が容易。不正な流通が生ずると被害は急速かつ広範囲に拡大するおそれ。
- 価値のあるデータであっても、①著作権法の対象とはならない、又は、②他者との共有を前提とし「営業秘密」には該当しない場合、その不正流通を差し止めることは困難。

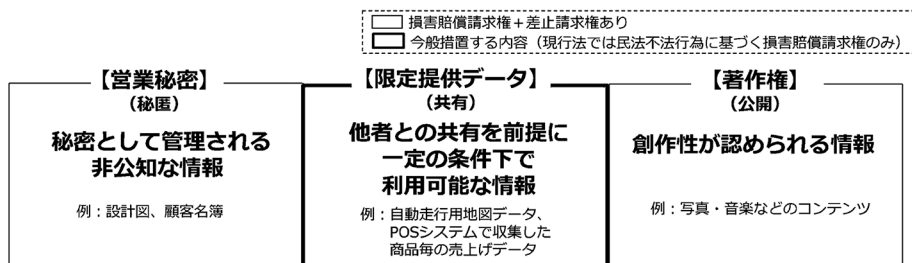


図 21 平成 30 年改正の不正競争防止法における限定提供データ

たとえば、そもそも、ある発明が特許として保護される対象になる理由は、うまくいくかどうか分からないところに挑戦していった何かやったという部分とその公開への代償という側面も考えられるところ、実は、データ駆動型人工知能の技術で、初期的に一番コストが掛かるのは、最初にデータを集めるところだと考えられる。そして、「データ駆動型人工知能」ということであると、データを集めてモデルを作ってみないと、本当にうまくいくかどうか分からない、という前提で、開発に着手していることも加味することが必要と考える。

つまり、このオープン領域やプレコンペティティブな領域に対して、貢献や努力をするということは、知財としての保護の観点からも、ある程度、評価がされなければならないのではないかと考える。

さらに言えば、むしろ、日本企業の中には、データを集めてくる部分のハードウェアに、強みを持っている企業も多いと考えられる。オープン領域やプレコンペティティブな領域のところの技術内容を自分たちで決めていく(たとえば、「技術標準化」していく)ということに積極的に関与していくということも可能なのである。とすれば、日本企業、ないしはもちろん日本企業に限らないものの、オープン領域やプレコンペティティブな領域のところに強みを持っている企業にとってみれば、これらの領域での知財保護というのは非常に重要と考える。

日本ではデータ構造の特許権というのは認められているものの、後述するような理由で、それ以外にも、何らかの方策を考える必要があるのではないかと、個人的には考える次第である。

(19) 経済産業省 知的財産政策室「不正競争防止法平成 30 年改正の概要(限定提供データ、技術的制限手段等)」
https://www.meti.go.jp/policy/economy/chizai/chiteki/H30_kaiseigaiyoutext.pdf

4.4 データ保護のための新たな知的財産の必要性について

一般論として、知的財産として保護される理由として何があるかといえば、「創作」だから保護する、「標識」だから保護する、「投資」があったので保護する、というような分類がありえる。そして、この投資については、不競法で保護がされるという部分があると思われる。ただし、この「投資」に関連する部分については、特に、データ駆動型人工知能技術については、もう少し検討がされてもいいのではないと思われる。

そこで、まず、特に著作権の場合を一つ例にとると、著作権については、平成30年法改正があって、下図22のとおり、第1層、第2層、第3層と分け、人工知能の学習にデータを使うためにも、著作権による制限がなるべくないように、改正がされてきているといえる⁽²⁰⁾。

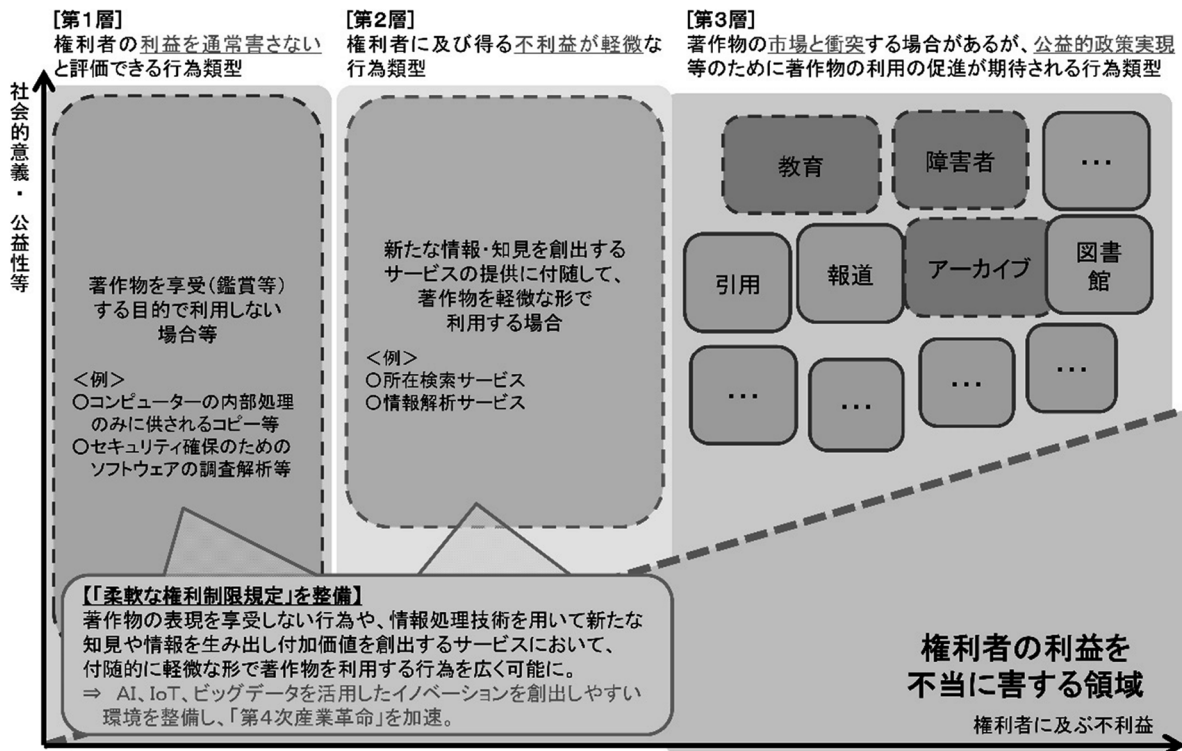


図22 平成30年改正 著作権法における「データ」の保護

研究者などによっては、日本が、人工知能の学習に関しては、少なくとも著作権に関しては世界で一番やりやすい国になっているということを、言われる場合もある状況である。

典型的には、例えば、著作権法第30条の4という条項が改正法によりできたことで、重要なところとしては、それまでは自身が学習済みモデルを作るためにデータを使用することが例外として規定されていたところ、それだけではなく、収集した学習用データを他人に提供する、利用させることが例外として可能となっている。

ただ、そのような状況の中で、著作権法だけではなく、学習用のデータを流通させるというときに、今の知的財産制度だけで十分かを、以下、具体例を挙げて検討する。ある意味、マニアックな例のようにも見える可能性があるものの、AIの学習用データという観点からは、十分に現実問題として想定されうるものと考えている。

(20) 文化庁「著作権法の一部を改正する法律 概要説明資料」

https://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/pdf/r1406693_02.pdf

そこで、たとえば、図 23 に示すように、ある A 社という会社が学習用データのデータベースを作ったとする。「タグデータ」と呼んでいるのは正解のデータで、それに続いて「データ項目」と書いてあるのが、人工知能の学習のために入力されるデータである。つまり、この「入力データ」については、各レコードについて、これが「正解データ」である、というものを多数並べたデータベースを作ったとする。

この「タグデータ」は、たとえば、自動運転における街路の画像データ（動画データを含む）であれば、人間が、その画像中に含まれる物・人が、「車」であるとか「通行人」であるというようなものであり、ここに創作性は、ないものとする。

一方、著作権の保護について考察しやすくするために、この例の場合、各項目も著作物ではないと想定することにしよう。もっとも、各項目は著作物ではないとしても、実際は、データを多数集めてくるには、かなりの時間とコストがかかるということが、特に現段階での人工知能の学習用データについては確かなこととしてある。

一方で、著作権の問題として考えることとしたいので、このデータベースのレコードの中で、データ項目 1, 2, 3 というものを集めて、このような編集の仕方をしていること自体には創作性があるというふうに見えるものとする。

そのような状況において、A 社は、このような学習用データベースを作ったので、先ほどの著作権法の改正なども考慮して、B 社にサブライセンスができるような条件でライセンスをしたところ、B 社は、人工知能の研究を、自社でも実施しているので、自社でのこの学習用データベースに基づいて、一部が異なる構成のデータベースを作成したものとする。

「（学習用）データ」の流通における知的財産権の課題

→ 特に、AIでは、「学習用データ」が重要

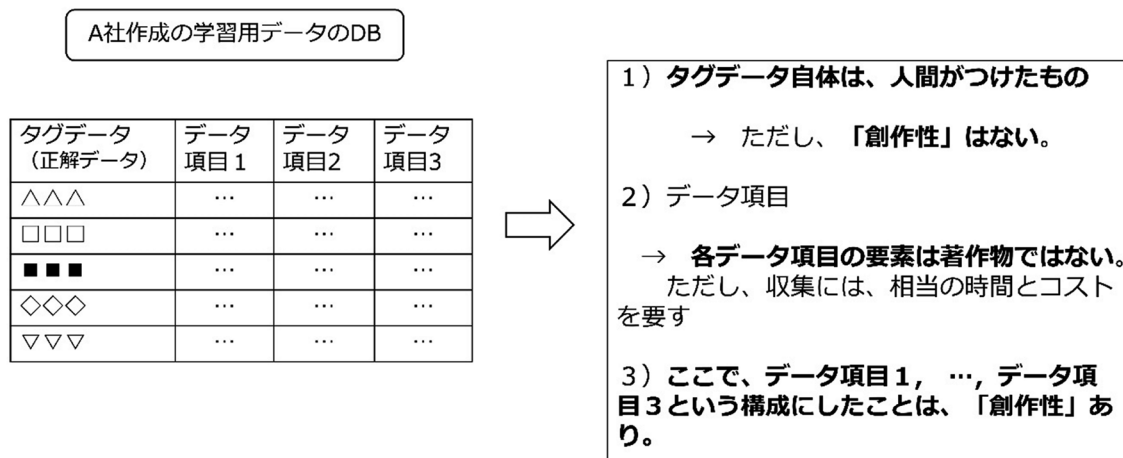


図 23 「学習用データ」の流通における課題を説明するためのデータベースの例

A 社は、B 社にデータベースの複製は許可しているものの、通常は、データベースのライセンスのときに「翻案」は禁止しているということが多いのではないかと思います。B 社に、データベースの内容については勝手にいろいろな変更をされないようにするためである。

このとき、図 24 に示すように、タグデータのところに B 社が手を加えて新しい学習用データベースを作ったとする。B 社は、このタグデータを変えると、元々のデータベースでやっていたことよりもさらに機能アップしたような、人工知能を生成できるということを見つけたとする。ただ、著作物としては、創作性があるところはまったく手を加えていないという状況である。

そうであれば、問題点として、タグデータに手を加えると著作権の侵害になるのかとか、言い換えると、

翻案に当たるのかどうかや、翻案は許諾されていなかったものの、著作権法第30条の4も存在するので、B社自身の行為についての扱いとしてはどうなのかということが、現実の実務上は、論点として挙がり得ると考える。

また、契約で、翻案をすることは禁止していたときに、上記の変更をしたものを、B社が第三者にサブライセンスするというのは契約違反になるのか、という論点も考えられる。

「(学習用) データ」の流通における知的財産権の課題

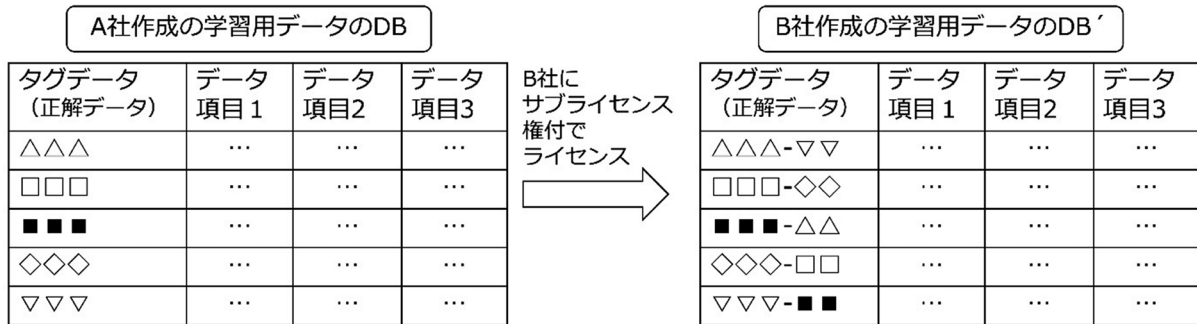


図 24 データ流過程での B 社による「学習用データ」の改変

実態としては、データの流通において、サプライチェーンという呼び方がいいかどうか分からないものの、こういう改良により、新しいものを作った以上、自社が支払っているライセンス料よりも、もっと高額で下流側には、サブライセンスしていくことがあると考えられる。しかし、もともとの A 社は、それに対して何も言えないのかというようなことも、実務上は論点になる可能性がある。

そもそも、学習用データを誰かが作成して流通させる、ということまで考えると、説明したように、流通の途中で何らかの変更があるということは十分に想定される。ただし、著作権法上で現在カバーしている範囲でのみ、その論点がカバーできる場合だけではない、と思われる事情も、実務的には、出てくる場合があり得る。

一方で、最初にデータベースを作った者にしてみれば、創作的な寄与（または、投資行為としての寄与）については何らかの経済的なりターンがあるというのは想定されるべき、との要請があるものと考えられる。

上述したように、データには階層構造があるというふうに図 20 に示した際にも言及したとおり、公開データだけではなく共有データについても、それを技術標準にするとか、その先のことも含めて、データを集めて学習用データを生成していくという動機づけも想定はされるものの、学習用データの作成者という主体に対する知財制度上でのインセンティブというのは何らかの方法で、考えるべきであろう。

しかも、もう一つ重要なところなのであるが、データを集めて公開するということは、例えば、国の研究資金などにより、公的な研究ファンドの実施期間中に研究成果として、公開をするということは、ごく普通に実施されている。むしろ、最近では、そのような公開を実施しないと、公的な研究資金の提供を受けるといことが、そもそもできない、という事情がある。しかしながら、問題となるのは、公的な研究資金の提供というのは、例えば、3～5年の研究期間の終了とともに、終わってしまうということである。

すなわち、データを集めてデータベースを公開するところまでは公的な資金でまかなわれたとしても、研究ファンドの期間が満了した後、そのまま公開を続けるための、資金的・財政的などところをどうするのか、という点は、現実問題として残る。仮に、研究資金の提供が終了しても、民間企業であれば、技術的な標準を自分の側がつくっていくということをメリットだと考えるというのも、一つの方向性ではあるものの、経済上のリターンのことを想定したときに、データを知財として保護するということが一切いらぬの

かという点、そこは、何かあった方がいいのではないかと、というのが筆者の意見である。

データの保護を知財として考えたときには、著作権、データ構造の特許、データの生成方法の特許などが想定されるものの、著作権は、データベースのうち、創作的な表現のところでないとは保護されない⁽²¹⁾。一方、データ構造の特許といっても、日本の場合はハードウェアとソフトウェアの協働関係がクレームの記載上は必要ということになっているので、データのところだけを保護してもらえないということにはなっていない⁽²²⁾。

また、学習用データを生成する方法の特許というのも、もうすでに日本国内で多数出願されていると推察

(21) 知財高判平成 28 年 1 月 19 日（平成 26 年（ネ）第 10038 号著作権侵害差止請求控訴事件）（「旅ネスプロ事件」）。判決骨子としては以下の通りである。

- ・著作権法第 12 条の 2 第 1 項は、データベースで、その情報の選択又は体系的構成によって創作性を有するものは、著作物として保護する旨規定しているところ、情報の選択又は体系的構成について選択の幅が存在し、特定のデータベースにおける情報の選択又は体系的構成に制作者の何らかの個性が表れていれば、その制作過程において制作者の思想又は感情が移入され、その思想又は感情を創作的に表現したものととして、当該データベースは情報の選択又は体系的構成によって創作性を有するものと認めてよいものと解される。

- ・リレーショナルデータベースにおいては、データベースの一部分を分割して利用することが可能であり、また、テーブル又は各テーブル内のフィールドを追加したり、テーブル又はフィールドを削除した場合であっても、既存のデータベースの検索機能は当然に失われるものではなく、その検索のための体系的構成の全部又は一部が維持されていると評価できる場合があり得るものと解される。

ここで、「学習用データ」については、独自の体系的構成を有しているとは言えない場合も多いと想定される。しかも、仮に、「学習用データ」が、外見上データベースとしての構成を有しているとしても、本文において、例示したように、必ずしも、「創作性」のある部分の変更が問題となるとは限らない。

この判決では、「リレーショナルデータベースの複製・翻案の判断方法」として以下の通り判示されている。

『(オ) 被告 CDDB（新版）と共通する原告 CDDB のテーブルに係る部分の体系的構成及び情報の選択の創作性等について原告 CDDB のうち、被告 CDDB（新版）と共通するテーブルに係る部分には体系的構成及び情報の選択における創作性があるから、被告 CDDB（新版）において原告 CDDB の上記部分の表現上の本質的特徴を直接感得することができるというべきである。

- a 原告 CDDB における主要な検索を可能とするテーブルに係る部分の体系的構成及び情報の選択の創作性
 - (a) 体系的構成の創作性 …
 - (b) 情報の選択の創作性 …
- b 被告 CDDB（新版）と共通する原告 CDDB のテーブルに係る部分について …

以下のとおり、上記 3 テーブルを欠くことにより、原告 CDDB 主要部分の体系的構成における創作性が喪失したものとはいえず、原告 CDDB 共通部分には、主要な検索を行うための体系的構成が依然として維持されているから、原告 CDDB 共通部分の体系的構成は創作性を有するものであり、被告 CDDB（新版）には、原告 CDDB 共通部分の体系的構成の表現上の本質的特徴が維持されている。…

(c) 以上のとおり、被告 CDDB（新版）には、原告 CDDB 共通部分の体系的構成及び情報の選択における創作的表現が認識可能なかたちで残っているから、被告 CDDB（新版）に原告 CDDB がないテーブルやフィールドがあり、また、データが追加されたとしても、被告 CDDB（新版）において原告 CDDB 共通部分の表現上の本質的特徴を直接感得することができるというべきである。』

すなわち、①被告データベースにおいて、原告データベースのテーブル、各テーブル内のフィールド及び格納されている具体的な情報（データ）と共通する部分があるかどうかを認定し、次に、②その共通部分について原告データベースは情報の選択又は体系的構成によって創作性を有するかどうかを判断し、さらに、③創作性を有すると認められる場合には、被告データベースにおいて原告データベースの共通部分の情報の選択又は体系的構成の本質的な特徴を認識可能であるかどうかを判断する、という手順を採用することを示した。このようにまず共通点を認定してから著作物性を判断する手法は「濾過テスト」と呼ばれる（最判平成 13 年 6 月 28 日（平成 11 年（受）第 922 号損害賠償等請求事件）（「江差追分事件」））。

「学習用データ」が流通するという場合、元の「学習用データ」に、何らかの変更が加えられることは、高い頻度で想定される。このような流通において、著作権者の権利が、上記のようなデータベースの考え方で保護されうるかは、学習用データでの「正解データ」に創作性があるといえる場合は、むしろ稀とも考えられ論点となりうると考える。

(22) 知財高判平成 26 年 9 月 24 日（平成 26 年（行ケ）第 10014 号審決取消請求事件）（「知識データベース事件」）。争点となったクレームの記載は、以下のようのものであった。

【請求項 1】

知識ベースシステムであって、
 コンピュータによる論理演算の対象となる知識ベースを記憶している記憶部を備え、
 前記知識ベースは、物を識別する物識別子と、前記物をもつ少なくとも一つの属性であって、当該物の物識別子と対応づけられた属性とを含み、
 前記属性には、当該属性を識別する属性識別子が 1 対 1 に対応づけられ、
 前記属性識別子には、属性を表す少なくとも一つのデータである特徴データ、及び属性を表す言葉に対応付けられ

するものの、その生成方法が特許になっていけば学習用データにまで、権利が及ぶかどうかについては、少なくともドイツの判例⁽²³⁾とかなを見る限りは、なかなか難しい可能性がある。少なくとも日本では、まだ、判例の蓄積等はないものの、おそらく、同様に困難が想定される。

したがって、学習用データは非常に重要であるにもかかわらず、現状、知的財産権による保護が十分に与えられているとはいえない、とする方がむしろ妥当であろう。

ただし、一方で、ただの情報にすぎないものにあまり強すぎる権利を与えるというのも、そもそも難しい

たデータである識別データのうちの少なくとも一方が対応づけられ、
前記物識別子は、物を表す言葉ではなく、かつ、それ自体で物の意味を持たない記号で構成され、
前記属性識別子は、属性を表す言葉ではなく、かつ、それ自体で属性の意味を持たない記号で構成され、
前記特徴データは、対応する属性の実体であり、
前記識別データは、対応する属性を識別するためのデータである
知識ベースシステム。

【請求項 26】

知識ベースシステムのためのコンピュータ読み取り可能な記録媒体であって、
請求項 1～23 のいずれか 1 項に記載の知識ベースまたは請求項 25 記載のプログラムが記録された記録媒体。

判決では、以下の通り、判示された。

『(2)・・・(略)・・・特許法第 2 条第 1 項は、「発明」とは、「自然法則を利用した技術的思想の創作のうち高度のもの」をいうと規定し、発明は、一定の技術的課題の設定、その課題を解決するための技術的手段の採用及びその技術的手段により所期の目的を達成し得るという効果の確認という段階を経て完成されるものである。

そうすると、請求項に記載された特許を受けようとする発明が、特許法第 2 条第 1 項に規定する「発明」といえるか否かは、前提とする技術的課題、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の技術的意義に照らし、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か否かによって判断すべきものである。

そして、上記のとおり「発明」が「自然法則を利用した」技術的思想の創作であることからすれば、単なる抽象的な概念や人為的な取決めそれ自体は、自然界の現象や秩序について成立している科学的法則とはいえ、また、科学的法則を何ら利用するものではないから、「自然法則を利用した」技術的思想の創作に該当しないことは明らかである。また、現代社会においては、コンピュータやこれに関連する記録媒体等が広く普及しているが、仮に、これらの抽象的な概念や人為的な取決めについて、単に一般的なコンピュータ等の機能を利用してデータを記録し、表示するなどの内容を付加するだけにすぎない場合も、「自然法則を利用した」技術的思想の創作には該当しないというべきである。』

『ウ さらに、本件補正発明の上記構成から導かれる効果についてみると、本願明細書には、「本願技術を用いることにより、物および物が持つ属性を、物や属性の意味を表わさない識別子で表わし、保持できる。」(段落【0056】)との記載はあるが、これは本件補正発明の構成自体を別の表現で表現したに等しいものであって、前記アのとおり、そのような形式でデータを保持する技術的な意義が明らかではなく、それによって、どのような効果が生じるのかについて明らかになっているとはいえない。

エ 以上を総合して検討すれば、本件補正発明については、そもそも前提としている課題の位置付けが必ずしも明らかではなく、技術的手段の構成としても、専ら概念の整理、データベース等の構造の定義という抽象的な概念ないしそれに基づく人為的な取決めに止まるものであり、導かれる効果についてみても、自ら定義した構造でデータを保持するという本件補正発明の技術的手段の構成以上の意味は示されていない。また、その構成のうち、コンピュータ等を利用する部分についてみても、単に一般的なコンピュータ等の機能を利用するという程度の内容に止まっている。

そうすると、本件補正発明の技術的意義としては、専ら概念の整理、データベース等の構造の定義という抽象的な概念ないし人為的な取決めの域を出ないものであって、全体としてみて、「自然法則を利用した」技術的思想の創作に該当するとは認められない。』

日本では、審査基準の「第Ⅶ部」「第 1 章 コンピュータ・ソフトウェア関連発明」「2.2「発明」であること」「2.2.1 基本的な考え方」の記載「ソフトウェアとハードウェア資源とが協働した具体的手段によって、使用目的に応じた情報の演算又は加工を実現することにより、使用目的に応じた特有の情報処理装置(機械)又はその動作方法が構築されること」が支配的な判断基準(ここでは演算・動作構築基準とよぶ)であることから、データ構造をクレームした場合には、データ構造自体でコンピュータが動作するものではないので、上記演算・動作構築基準に合致せず、その結果、発明に該当しないと判断されることが多く、コンピュータの動作をクレームに規定するように求められる。

- (23) 最高裁判所判決 X ZR124/15 (2016 年 9 月 27 日判決)。Franz-Joseph Zimmer, Olivia Nemethova, Andreas Kayser (中村 敏夫(訳))「診断データは方法の『直接製品』として保護されるか?」AIPPI63 巻 1 号 26-32 頁 (2018)。

『最高裁判所は、下記理由によって、被告の行為がドイツ特許法 9 条 3 号の侵害をしていないことを確認した。ドイツ特許法 9 条 3 号に従って、特許方法の直接結果物である製品は、上記の通り、物の特許による製品の保護と同様に保護される。しかし、特許方法が、特許方法によって与えられた特性を示す製品を生産する場合にのみ、ドイツ特許法 9 条 3 号が特許方法の結果物に適用される。従って、特許保護された方法は、原則として、それ自身で物の特許の対象となりうる何らかの物に導かれるものでなければならず、それは裁判所の考慮事項となる。

上記考慮事項を本件の背景事実に適用すると、チェコ共和国で生産された試験結果は、裁判所によってドイツ特許法 9 条 3 号の意味で直接製品とはみなされない。試験結果は、物の特許によって物として保護されることがなく、むしろドイツ特許法 1 条 3 項 4 号の下、特許保護対象から排除される「情報」を単に構成する。』

のではないかと考える。かえって、自由にデータを流通させたいという意味では問題があるというふうにも考えられるし、データに仮に独占排他権を付与したとしても、その侵害摘発の困難さから、実効的に、それで侵害行為を抑えられるかというのも、なかなか難しいところがあるというふうに認識する。

もっとも、学習用データを収集・生成した者に対する何らかのインセンティブというのは、やはり必要であり、特に「公開されるデータ」について、第三者までデータが流通していくということを考えたときに、それが元々は誰のデータであるかということについての権利は、一定程度は必要なのではないかと想定する。

そこで、全然別の知財の考え方をを用いる可能性も、もちろん想定されるものの、たとえば、半導体集積回路の場合は、「回路配置利用権」というものがあることを参考に、「データ利用権」と呼ぶべき制度を考えるというのも一つの方向性ではないかと考える。もちろん、「回路配置利用権」は、創作だから保護されるという立て付けではあるものの、可能性としては、あり得るものとも考える。

「回路配置利用権」については、すでによく御存知のところかもしれないが、知的財産権としては、実務上は、通常、あまり遭遇することがない権利のようにも思うので、以下に簡単にまとめる。

回路配置利用権

1. 保護対象

法第2条に以下の定義規定がある。

- 1) 「半導体集積回路」とは、半導体材料若しくは絶縁材料の表面又は半導体材料の内部に、トランジスタその他の回路素子を生成させ、かつ、不可分の状態にした製品であって、電子回路の機能を有するように設計したものをいう。
- 2) 「回路配置」とは、半導体集積回路における回路素子及びこれらを接続する導線の配置をいう。

2. 保護の体系

半導体集積回路配置法は、申請された半導体集積回路の回路配置を設定登録することによって「回路配置利用権」という権利を発生させることを規定する。

「回路配置利用権」に基づいて他人による回路配置の模倣を禁止することが可能である。

3. 申請と登録手続き

設定登録の申請があった場合、経済産業大臣から委託された一般財団法人ソフトウェア情報センター(SOFTIC)が、以下の事項について審査する。

- ① 申請者が申請に係る回路配置の創作者またはその承継人であること
- ② 創作者またはその承継人が2人以上ある場合において、それらのものが共同申請していること
- ③ 創作者またはその承継人が申請に係る回路配置について申請の日から2年以上遡った日前に業として譲渡・輸入などの行為をしていなかったこと
- ④ 申請書が方式に適合していること

4. 登録の効果

- ① 設定登録をすることにより回路配置利用権を取得。回路配置利用権の存続期間は設定登録の日から10年。
- ② 設定登録をすることにより回路配置を業として利用する権利を専有。
- ③ 回路配置利用権者は専用利用権を設定することが可能。
- ④ 回路配置利用権者は他人に通常利用権を許諾することが可能。
- ⑤ 権利侵害者に対して損害賠償請求や差止請求をすることが可能。
- ⑥ 設定登録後の移転、処分の制限、質権設定等は登録をすることにより第三者に対抗することが可能。

5. 権利の性質

- ① 回路配置利用権の効力は、他人が創作した回路配置の利用には及ばない。すなわち、絶対的な排

他権ではなく、著作権のような相対的排他権である。したがって、第三者の独自に創作した回路配置である限り、回路配置利用権の侵害には該当しない。

- ② 解析又は評価するために設定登録を受けている回路配置を用いて半導体集積回路を製造することも、回路配置利用権の侵害には該当しない。したがって、解析等のみを目的とするリバースエンジニアリングは権利侵害には該当しない。

もっとも、残念ながら、「回路配置利用権」の登録は、極めて低調なようである。

しかしながら、むしろ、「学習用データ」の方が、たとえば、技術標準などとするを目的として、公開して第三者に利用を促す、などの需要は、あるものと想定される。そこで、たとえば、回路配置利用権になぞらえて、「データ利用権」として想定されるものを、そのうち、主要な点に関する案を、例示すると、以下のようなものが想定される。もちろん、あくまで、一案に過ぎないものであり、権利のイメージのためにあえて、例として、記載させていただくものであることとお断りしておく。

データ利用権（案）

1. 保護対象

- 1) 「データ駆動型人工知能」とは、収集されたデータに基づいて、コンピュータ自身が機械学習処理を実行し、学習した結果に基づいて推論・予測を実行する「人工知能」をいう。
- 2) 「学習用データ」とは、教師あり学習のための正解データのあるデータセット又は教師なし学習のための正解データのないデータセットであるかを問わず、データ駆動型人工知能の機械学習処理のために予め収集されコンピュータへの入力に適した形式とされたデータであって、業として第三者に提供する情報として電磁的方法（電子的方法、磁気的方法その他の知覚によっては認識することができない方法をいう。）により相当量蓄積され、及び管理されている技術上又は営業上の情報をいう。
- 3) 「学習用データの作成者」とは、データを収集して、コンピュータに、教師あり学習又は教師なし学習により、第三者が機械学習処理により人工知能モデルを生成する用に適したデータセットとして、学習用データを作成した者をいう。

2. 保護の体系

データ利用法（仮）は、申請された「学習用データ」を設定登録することによって「データ利用権」という権利を発生させることを規定する。

「データ利用権」に基づいて他人による学習用データの模倣を禁止することが可能である。

「データ利用権」の設定された「学習用データ」については、少なくとも、その一部（10%以上）を無償で公開することを要する。その公開にあたっては、学習用データの全体について利用を希望する第三者に提供するためのライセンス条件を併せて提示することを要する。ライセンス条件は、学術用途または商業用途に応じて、個別に条件の設定が可能であり、学習用データの作成者の選択により、対価について有償又は無償とすることができる。ただし、利用を希望する第三者の利用目的のみを理由として、ライセンスの許諾を制限することはできない。

3. 申請と登録手続き

設定登録の申請があった場合、経済産業大臣から委託された〇〇〇〇が、以下の事項について審査する。

- ① 申請者が申請に係る学習用データの作成者またはその承継人であること
- ② 作成者またはその承継人が2人以上ある場合において、それらのものが共同申請していること
- ③ 作成者またはその承継人が申請に係る学習用データについて申請の日から2年以上遡った日前に

業として譲渡などの行為をしていなかったこと

- ④ 申請書が方式に適合していること

4. 登録の効果

- ① 設定登録をすることによりデータ利用権を取得。データ利用権の存続期間は設定登録の日から〇〇年。
- ② 設定登録をすることにより学習用データを業として利用する権利を専有。
- ③ データ利用権者は専用利用権を設定することが可能。
- ④ データ利用権者は他人に通常利用権を許諾することが可能。
- ⑤ 権利侵害者に対して損害賠償請求や差止請求をすることが可能。
- ⑥ 設定登録後の移転、処分の制限、質権設定等は登録することにより第三者に対抗することが可能。

5. 権利の性質

- ① データ利用権の効力は、他人が独自に作成した学習用データの利用には及ばない。すなわち、絶対的な排他権ではなく、著作権のような相対的排他権である。
- ② 解析又は評価するために設定登録を受けている学習用データを用いて人工知能モデルを作成することのみでは、データ利用権の侵害には該当しない。したがって、解析等のみを目的とするリバースエンジニアリングを仮に実行したとしても権利侵害には該当しない。

なお、データである以上、公開してひとたび第三者に提供してしまえば、そもそも、侵害の摘発が困難ではないか、という疑問もあろう。しかしながら、たとえば、技術標準などとして、一定の認証や承認の対象の試験のために必要となるようなデータであって、定期的に内容がバージョンアップされるデータについては、そもそも、そのようなデータを、独自に収集して学習用データを作成するよりも、経済的には、権利者から取得する方が有利ということは、十分に想定される。むしろ、有用なデータベースというものは、それを生成するために、多くの投資が必要であればあるほど、権利者に対して、その利用の許諾を求める、という傾向が強い⁽²⁴⁾。この点では、侵害行為が発見され摘発されるまでは、侵害者による侵害行為が継続することの多い特許権などとは、むしろ好対照といえるであろう。

逆に、学習用データを提供する側としては、このような「一定の認証や承認」のためのデータを、制定しバージョンアップする権利を自身が保有することは、競争上、極めて有利という側面もある。

例えば、「プログラム」であって、「データ」ではなく、また、ビジネスにおける顧客層が全く異なると思われるので、単純には比較できないものの、パソコンの自作業者や企業が、パーツ性能を数値化するために利用することが多い「ベンチマークソフト」などを提供する企業のビジネスモデルなども、参考になるところは多いと考えられる。

それから、「限定提供データ」についても、何らかの方法で、もう少し、それが使いやすく、例えば、一定の制限条件の下で、個人情報保護法上の例外を認めるなどのような制度的な改定も必要なのではないかとも考える。

5. まとめ

以上説明したとおり、データ駆動型人工知能の知的財産保護という観点では、日本の特許制度において、「学習済みモデル」そのものを特許の対象とする、との運用については、このような技術の保護にとってクレー

(24) 筆者が知的財産部門を担当した株式会社国際電気通信基礎技術研究所では、その1986年の創設のころから、音声認識、言語翻訳、音声合成技術などの開発を行っていた。その技術に適用される「音声データベース」「対訳コーパスのデータベース」については、むしろ、利用を希望する第三者からの打診が顕在的に多かったという実務上の経験がある。

ムドラフトの自由度ないしは保護範囲の拡大という観点から、極めて重要と考える。もっとも、このような「学習済みモデル」そのものの特許による保護という点では、まだまだ国際的な調和という観点から、課題が多い。

一方で、「データ駆動型人工知能」のような技術によるサービス提供という観点からは、特許法上の実施行為の概念について、「プログラム等の譲渡等」については、「ネットワークを介した、プログラムのアクセス権の提供」自体も、含まれるような解釈が望ましいと考える。

もっとも、実は、「データ駆動型人工知能」を知的財産権で保護することにより、産業の発達を図る、という観点からは、筆者としては、特に、オープンに学習用データとして利用可能なデータについて、何らかの新たな知的財産権としての保護の必要性を感じており、立法論とはなるものの、その可能性について提案させていただいた。