

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

弁理士 酒井 將行

目次

1. はじめに
2. プラットフォーム型ビジネスモデル
 - 2-1. 「プラットフォーム型ビジネスモデル」とは？
 - 2-2. 「プラットフォーム型ビジネスモデル」における知財の役割
 - 2-3. 「プラットフォーム型ビジネス」が他の「プラットフォーム型ビジネス」を駆逐するとき
 - 2-4. 人工知能技術とプラットフォーム型ビジネスモデル
3. データ駆動型ビジネスモデル
 - 3-1. 「データ駆動型ビジネス」とは？
 - 3-2. 「データ駆動型ビジネスモデル」は、すでに存在しているか？
 - 3-3. 「データ駆動型ビジネスモデル」における知的財産の役割
4. 「データ駆動型人工知能」を保護するための知的財産とは？
 - 4-1. 「データ駆動型人工知能」に関連する技術の明細書・クレームの実例
 - 4-2. 「データ駆動型人工知能」についての特許庁ハンドブックの記載と検討課題
 - 4-3. 「データ駆動型人工知能」の明細書・クレームをいかに記載すべきか？
5. まとめ

1. はじめに

ビジネスにおいては、「製造業のサービス化（“〇〇 As a Service”との名称で呼ばれることもある）」や、「モノの所有」から「モノの使用」へと顧客の価値が変容していく中で、ビジネスモデルとしても、従来の製品たる「モノ」を中心として、生産者から消費者に向かって直線的に価値の創造が行われる「パイプライン型ビジネス（サプライチェーン）」という考え方から、「プラットフォーム型ビジネス」が中心的な役割を果たすようになってきていることが指摘されている^{(1), (2), (3)}。

事実、ICT（Information Communication Technology）技術においては、現在、「プラットフォーム型ビジネスモデル」が、支配的な影響力をもっているといわれ、このようなビジネスモデルに関しては、いわゆる“GAFA（Google, Amazon, Facebook, Apple）”の独壇場ともいえる状況となっている。もっとも、後述するように、「プラットフォーム型ビジネスモデル」は、世界的超巨大企業だけの話とはいえない。

一方で、ICT技術において、ディープラーニングなどの、いわゆる「データ駆動型人工知能」技術が、様々な局面で、大きな成功を収めている。この結果、AI（人工知能）のビジネス適用が本格化してきており、

-
- (1) アレックス・モザド（著）、ニコラス・L・ジョンソン（著）、藤原 朝子（翻訳）「プラットフォーム革命——経済を支配するビジネスモデルはどう機能し、どう作られるのか」英治出版（2018/2/7）
 - (2) ジェフリー・G・パーカー（著）、マーシャル・W・ヴァン・アルスタイン（著）、サンジート・ポール・チョーダリー（著）、& 2 その他、妹尾 堅一郎（監訳）、渡部 典子（訳）「プラットフォーム・レボリューション PLATFORM REVOLUTION 未知の巨大なライバルとの競争に勝つために」ダイヤモンド社（2018/8/23）
 - (3) 妹尾 堅一郎「イノベーションが変えるビジネスモデル～産業生態系の変容が競争力原理を変える～」講演資料、知財人材育成研究分科会 第32回例会（大阪工業大学 梅田キャンパス）（2018/2/16）

PoC (Proof of Concept : 実証実験) 段階を経て⁽⁴⁾, 「業務効率化」「新規ビジネスの創造」に取り組む企業も増えている状況である。結果として、「人工知能」の処理を実施するためのソフトウェアと両輪として働き、「データ駆動型人工知能」の学習に資するものとして、「データ」そのものの重要性が、特に、認識されるにいたっている。状況によっては、人工知能の学習や予測・分類処理を実行するための「プログラム」自身は、既成のものであって、「データ」にこそ、新規な機能を実現するための「技術上の価値」が存在する場合も多くみられる。

その結果、もともとは、ソフトウェア技術等における技術用語であった「データ駆動」あるいは「データドリブン」という用語が、単に技術の世界だけではなく、「データドリブンマーケティング」「データドリブン経営」などというように、経営学の世界でも使用されるようになっていくだけではなく、さらには、「データ駆動型社会⁽⁵⁾」「データ駆動型ビジネス」などというような、より広い概念を表現することを意図する言葉も使用されるにいたっている。

一方で、「知的財産推進計画 2017⁽⁶⁾」において、「データ・人工知能の利活用促進による産業競争力強化に向けた知財制度の構築」ということが謳われ、i) データ利用の契約ガイドラインの策定、ii) 不正競争防止法改正 (データの不正取得等の禁止等)、iii) 著作権法改正 (柔軟性のある権利制限規定の整備)、iv) AI 学習済みモデルの特許化の具体的な要件や保護範囲の検討 などが課題として挙げられ、これらは、いずれも法律・ガイドライン・審査ハンドブックなどの形で具体化された。

特に、著作権法の改正は、人工知能の学習処理のために必須である「学習用データ」について、「インターネット上で公開された情報」(サイバー空間での情報) または「公表された著作物の情報」(現実世界の情報の一部) を、一定条件下、収集・利用・提供することを可能とする点で、大きな前進と言える。

さらに、“多種多様かつ大量のデータの分野横断的なデータ流通を実現するため、個人の関与の下でデータ流通・活用を進める仕組みである「PDS (Personal Data Store)⁽⁷⁾」や「情報銀行」、データの需要と供給のマッチングや適切な利益還元を促進するための「データ取引市場」などの社会実装に向けて、官民が連携した実証実験等に取り組む”とされ、「情報銀行」については、実際の運用が開始されようとしている。

ただし、この「知的財産推進計画 2017」では、“知財制度と別の観点の検討が必要なデータ⁽⁸⁾を除き、データの収集・蓄積・保管等に関する投資インセンティブを付与する必要性の観点から、民間の投資等により生成された「個人に関わらないデータ」及び「匿名加工されたデータ」を「価値あるデータ」として主な検討対象とし、その利活用促進のための知財制度の在り方について検討することが必要である。なお、新たな法的な枠組みを検討する場合には、予見可能性や取引の安全等の見地から許容されるかという許容性について

(4) もっとも、実際に AI の導入を検討してみたが、結局、できなかった/失敗したという声も聞かれ、いわゆる「POC 疲れ」という状態の企業プロジェクトもあるようである。 <https://special.nikkeibp.co.jp/atcl/NXT/19/ycd0408/>

(5) 「未来投資戦略 2018—「Society 5.0」「データ駆動型社会」への変革—」(平成 30 年 6 月 15 日)

https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/miraitousi2018_zentai.pdf

(6) 内閣府知的財産戦略本部、<https://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku20170516.pdf>

(7) 「2022 年から本格始動 日本のデータ連携活用基盤」、大和総研 (2019/4/15)

https://www.dir.co.jp/report/research/policy-analysis/human-society/20190415_020750.pdf

なお、ここでは、PDS とは、「個人が企業に提供したパーソナルデータを自ら手元に集約し管理しデータ別に利活用条件を容易に設定するためのシステム」と規定されている。PDS、情報銀行の運用については以下の通り。

“パーソナルデータの第三者提供や利活用条件等については、個人が自ら判断することが基本であるが、一般には、個人が個人情報保護に関する法律知識を十分には持ち合わせていないケースや、頻度が多い場合には都度の判断が煩雑となるケースも想定される。…、あらかじめ個人が自らのパーソナルデータの利活用条件を設定した上で、データの一部または全てを、情報銀行を運営する事業者に一括して信託し、個人が PDS 等を用いて行う権利処理を代行することも可能としている。経済産業省及び総務省が出した指針によれば、情報銀行は、具体的には、情報銀行が管理するパーソナルデータの取得を希望する企業等との間で、第三者提供について本人同意済みのパーソナルデータを提供する際の切り出し範囲 (データの一部または全部) や提供形式 (個人特定が可能なパーソナルデータの状態のままか、匿名加工情報か) の確定、パーソナルデータ提供先となる企業内における情報セキュリティ及びプライバシー保護対策の遵守状況やガバナンス体制構築状況の確認、企業とのデータ提供価格の折衝等も担う…”

(8) たとえば「個人情報保護法」の観点から、「個人情報」となるデータについては除外するとの趣旨である。

の検討も必要である。”との下で、表1の範囲での検討が行われている。

表1. 知的財産推進計画2017における検討対象とされるデータ

【データと主な検討対象のイメージ】

原資	データの種類		
	個人に関わらないデータ	匿名加工されたデータ	個人情報を含むデータ
民間の投資等	今回の主な検討対象		
公的資金			

たとえば、上述したような「ビジネス適用」の局面では、社内（または状況によっては社外）にそれまでの各業務部門ごとの業務システムにおいて利用するために、オンプレミスまたはクラウド上に分散して保存されている「データ」を一元的に集約して利用する環境を構築することの重要性が指摘されている。ただし、現在のところ、このような「ビジネス適用」の局面では、一般には、利活用されるのは、公開データを除けば、まだ社内データに限定されているのが現状といえるであろう⁽⁹⁾。

一方で、いわゆるインターネットサービス事業者などは、広く不特定多数から収集したデータを利用している状況ではあるものの、これらの事業者が、現在までに集積してきた「強みがあるデータ」とは、「インターネット上で」「文字、音声または画像として収集された」「当該インターネットサービス業者のサービスのユーザ」から収集されたデータが主たるものと考えられる。そして、「インターネット上で」「文字、音声または画像として収集された」データ⁽¹⁰⁾とは、基本的には、データを収集するハードウェア（H/W）の特性には、あまり依存していないデータであって、データが入力されるコンピュータやスマホ等の仕様に依存することがなく、データフォーマットとしては既存の統一的なものが収集される状況であることに注意が必要である（テキストデータや画像データについては標準的な仕様が存在しており、データを収集する側は、そのフォーマットについてほとんど考慮する必要がない。）。また、データを収集する際に、ユーザがどの程度、自覚的に認識しているかはおくとしても、「サービスの提供」と引き換えに、データを提供するユーザの同意の上で、インターネットサービス業者がデータを収集しているとの構図となっている⁽¹¹⁾。

とすれば、今後、収集の対象となる（データ駆動型人工知能での「学習用データ」として使用することが想定される）データとは、もちろん、a) 上記のようなネットワーク経由で収集されるデータがあるのは当然ながら、b) データを取得するH/Wと技術的に密接な関連性を有するデータ、c) 「個人に関する情報」についてのデータ（個人情報保護法にいう「匿名加工」までの処理がされているか否かにも大きく依存する。特に、ヘルスケア分野でのデータについて留意が必要なことが顕著）、それから、d) 公的な資金により取

(9) 【失敗事例に学ぶ】IBMの人工知能エキスパートが説く、AI活用「第2章」の歩き方、ビジネス+IT、(2019/7/26)、<https://www.sbbit.jp/article/bitsp2/36375?ref=191230btsw#enquete>

(10) 2020年1月14日に、Googleは、クッキーにより収集した個人のネット閲覧履歴情報を、外部のネット企業などに無料で提供していたサービスを、2022年までに停止すると発表した。個人情報の保護の観点との発表ではあるが、一方で、むしろ、個人に関するデータの大手IT企業への寡占が進む、との懸念もある。

(11) たとえば、ニューヨークタイムズ紙の調査では、米国において、すでに、デバイス、アプリ、サービス、ウェブサイトが収集した膨大なデータが、民間のデータ収集企業の手におわり、巨大データベースに蓄積され、（一応の匿名化はされてはいると考えられるものの）これらのデータが広告企業にわたっている、とのことである。しかも、（匿名化されていても）特定の端末についての「位置情報の履歴」をトレースすることも不可能ではない、との指摘もされている。日本でも同様の事態となっているかはさておき、スマートフォンについては、（たとえば、アプリから要求される）「位置情報の記録・送信」を「許可」状態としているユーザは、相当数に上ると予想される。

<https://www.lifehacker.jp/2020/01/204647your-location-data-isnt-as-anonymous-as-you-think.html>

<https://lifehacker.com/your-location-data-isnt-as-anonymous-as-you-think-1840540886>

<https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>

集されたデータ（または、今後収集が予定されるデータ）の利活用をどうするか、という観点からの検討が必須と予想される。そして、今後は、これらのデータを制する者または有利な立場で利用できる者が、関連する市場で圧倒的に有利な立場をもつとの構図となるものとも予想される。

また、筆者の見解としては、後述するように、「データ駆動型ビジネス」と呼ばれるものの萌芽は、すでに各所に表れているものの、真の意味で「データ駆動型“ビジネスモデル”」といえるものが生まれているかについては、まだ、検討の余地があると考え。ICT分野において、いわゆる「プラットフォーム型ビジネスモデル」により、巨人といえるような企業によるビジネスの寡占化が進んだものの、この意味で「データ駆動型ビジネスモデル」が生まれるときは、いわゆる「破壊的イノベーション」⁽¹²⁾が生じる可能性もあり、まさに、それは現在進行中であるのかもしれない。

本稿は、『「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』を検討するにあたり、以下の諸点の筆者としての検討を基に、今後の「知的財産保護の制度」の検討を目指すものであるが、もちろん、「知的財産全般」については、筆者の手に余るところであり、中でも、特に、「特許制度」に焦点を絞ることとする。そして、「データ駆動型ビジネスモデル」が生まれることに備えて、少なくとも、現状の知的財産制度およびその延長において、どのような知的財産（あるいは知的資産）の保護を追求すべきかを考察することは有用であると考え。そこで、本稿では、「データ駆動型人工知能」を保護するための特許について、検討を行いたい。特に、検討対象としたいのは、以下の点である。

論点 1 「超スマート社会（Society 5.0）」において、中核的に重要となると考えられる「（データ駆動型）人工知能技術」に関連して、どのようなクレームの記載として、特許権の取得を検討すべきか？

論点 2 そのような「（データ駆動型）人工知能技術」に関連する特許出願において、どのような「特許要件」を想定すべきであるか？

上記の「どのようなクレームの記載として、特許権の取得を検討すべきか」という点については、技術内容に対する理解はもちろんであるが、現在、圧倒的な影響力を有している「プラットフォーム型ビジネスモデル」に対する理解、さらには、将来的に展開されると予想される（「データ駆動型ビジネスモデル」に至らないまでも）「データ駆動型ビジネス」として想定されるものを考慮する必要があるものと考え。

そこで、特許制度その他の知的財産制度について考察する前提として、まずは、「プラットフォーム型ビジネスモデル」および「データ駆動型ビジネス（モデル）」について検討することとする。

2. プラットフォーム型ビジネスモデル

2-1. 「プラットフォーム型ビジネスモデル」とは？

いわゆる、「（技術）プラットフォーム型ビジネスモデル」とは、以下の図1に示すように、プラットフォームが自身の技術を「共通基盤（プラットフォーム）」として提供し、これを利用して、複数の「プロデューサー（生産者）」が、プラットフォームの機能を利用して、複数の「ユーザ（消費者）」に提供することで、対価を得るという構成である⁽¹³⁾。

その価値を生み出している（そのプラットフォームの売り物となるものを作っている）のは、一義的には、プラットフォームを利用している「プロデューサー」と呼ばれるヒト・企業である。プラットフォームは、そのような「プロデューサー」と、価値を消費する「（プラットフォームの別の意味の利用者である）ユーザ」とをつなぐ、「場」を提供していることになる。

(12) クレイトン・クリステンセン（著）、玉田 俊平太監修、伊豆 原弓（訳）「イノベーションのジレンマ—技術革新が巨大企業を滅ぼすとき 増補改訂版」翔泳社（2001/7/3）

(13) 妹尾 堅一郎氏は、前掲注3で、このような関係を「 $N \times 1 \times N$ 」のビジネスモデルと呼んでおられる。最初のNは、参加するプロデューサーの数であり、真ん中の1は、プラットフォームであり、最後のNは、参加するユーザの数を意味している。ただし、本稿では、プロデューサー側をNとして、ユーザ側がMであるものとして記載することにする。

プラットフォーム(PF)ビジネスモデル

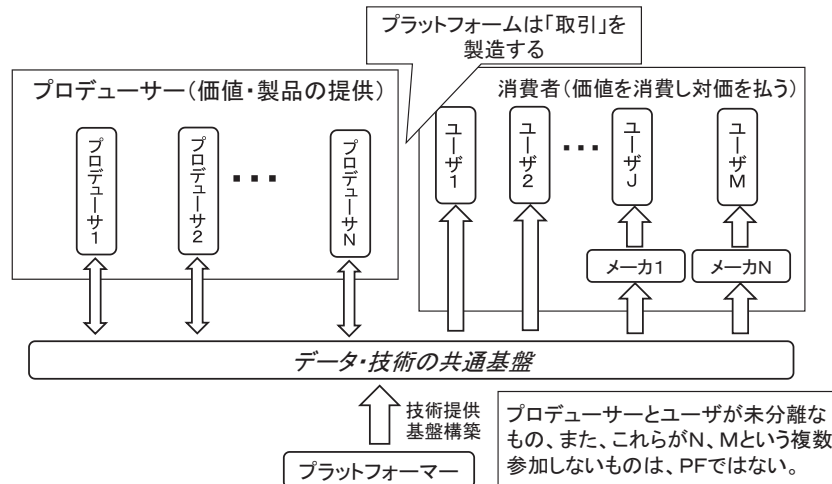


図1. 技術プラットフォーム型ビジネスモデルの概要

そして、プラットフォームは、ある程度以上の規模となると、まさにその規模の大きさが、「プロデューサー」や「ユーザ」にとってのメリットとなるため（ネットワーク効果）、よく言われるように、いわゆる「独り勝ち（Winner takes all.）」の状態を生み出しやすいといわれる。「パイプライン型ビジネス（サプライチェーン）」では、供給サイド側の規模の経済により効率化が進むことが一般であるが、「プラットフォーム型ビジネス」では、「需要サイドの規模の経済」も、重要な役割を果たすことになる。

プラットフォームの目的は、生産者と消費者を引き合わせて、情報、製品やサービス、通貨という3つの形の交換に関与させることにある。ここで、「通貨」⁽¹⁴⁾とは一種の比喩的な表現であり、製品サービスへの対価と言いつけられる。このような対価としては、たとえば、クレジットカードでの決済や、キャッシュレス決済、仮想通貨による決済などが、プラットフォーム内で実現されることが多い。これにより、（たとえば、自動車配車ウェブサイトおよび配車アプリを提供するウーバー（Uber）社などがそうであるように）サービス・製品の提供自体が、プラットフォーム外でなされる場合であっても、プラットフォームは、プラットフォーム上での取引に対して、たとえば、取引手数料を徴収する立場を確保できることになる。ただし、「対価」としては、必ずしも、このような具体的な経済上の価値の交換に限らず、ユーザからの「いいね！」などのような肯定的評価が、対価の役割を果たしている場合もありうる。

いずれにしても、プラットフォーム自体は、参加者をつなげるインフラと、簡単に互いに満足のいく交換を実現させるツールとルールとを提供するのである。

(1) YouTube の例

たとえば、現状のテレビ放送局を「パイプライン型ビジネス」とするならば、YouTube は、「プラットフォーム型ビジネス」ということになる。テレビ放送局は、スタジオを構え、制作スタッフを雇用して、タレントと芸能事務所経由で契約して、番組を制作し、自前の放送設備を介して、番組を放送する。YouTube は、自身で番組を制作することはなく、（ある意味視聴者でもある）YouTuber が作成したコンテンツを活用して、ビジネスを展開している。しかも、どのテレビ局よりも多くの視聴者を獲得している。たとえば、広告主たる企業が YouTube に広告宣伝料を支払い、YouTuber は、コンテンツに対する再生回数に応じて、YouTube から広告料の分配を受け、視聴者は、その広告の対象となる製品やサービスを購入することで、

(14) なお、Facebook は、2019年6月18日に、それまでの「仮想通貨」とは一線を画し、法定通貨と一定比率で交換できる「ステーブルコイン」としての「Libra（リブラ）」を2020年に実用化する旨を発表している。

間接的に、YouTuber に対するコンテンツ使用料を負担していることになる。ただし、この最後の部分は、民間放送と異なるわけではない。ただし、YouTube の方が、コンテンツ制作・提供のための負担は、圧倒的に少ないにも関わらず、提供できるコンテンツの量は、放送局とは比較にならないことになっている。

(2) Google による Android の例

Google は、スマートフォンなどで使用される OS (Operating System) として、Android を提供している。この Android は、オープンソースソフトウェア (OSS: Open Source Software) であり、そのソースコードは公開されている⁽¹⁵⁾。

スマートフォンメーカーは、このような OSS で提供される OS を用いて (つまり、OS そのものにはコストをかけずに) 自社のスマートフォンをリリースできるだけでなく、Android では、Google Play というアプリを介して、アプリ開発者がアプリを販売し、ユーザは、これを利用できるという環境を提供するプラットフォームが構築されている。この場合、OS 上で提供されるアプリの数が増えれば増えるほど、ユーザは、その OS を搭載したスマートフォンに魅力を感じるようになる。一方で、その OS を使ったスマートフォンのユーザが増えれば増えるほど、アプリ開発者にとっては、その OS 上で動作するアプリを提供しようとする動機が高まることになる (このような供給側と需要側の双方からのネットワーク効果のことは「ツーサイド・ネットワーク効果」と呼ばれる)。

YouTube などは、「コミュニケーション・コンテンツ型プラットフォーム」、Google などは、「オープン OS 型プラットフォーム」などとも分類される。プラットフォームには、さらに、ウーバーなどのような「仲介型プラットフォーム」や、医療画像診断支援 AI 統合オープンプラットフォームの AMIAS⁽¹⁶⁾ などの「ソリューション提供型プラットフォーム」に分類されるものがあるとの指摘もある⁽¹⁷⁾。

2-2. 「プラットフォーム型ビジネスモデル」における知財の役割

「プラットフォーム型ビジネスモデル」における知財の役割を検討するにあたり、まずもって、重要な点は、後述するように、「開発に巨額を投じた技術」、すなわち、従来なら典型的には、「特許により保護されるべき技術」と想定されてきたものと、プラットフォームが真に収益を上げている対象とは、必ずしも一致しないという点であろう。一方で、知財マネジメントにおいては、「オープン&クローズ戦略」の重要性が指摘され、数多くの書籍⁽¹⁸⁾が出版されているのが現状であり、IoT の世界でも、知財戦略として、「オープン&クローズ戦略」の重要性が指摘されている⁽¹⁹⁾。

それでは、このようなプラットフォーム型ビジネスモデルでは、知的財産 (権) はどのような役割を果たしているのだろうか？

上述した Google による Android OS では、これは OSS なのだから、この OS そのものの販売の対価によって収益を上げるという状況ではないことになる。また、Android OS 自体は、Linux ベースのカーネルの上で、

(15) Amazon が自社製品として提供しているタブレットである“Fire HD”は、Android ベースの OS で動作している。しかし、後述するような“Google Play”などのアプリは、プレインストールされていない。

(16) <https://www.optim.co.jp/medical/amias/>

(17) 「日本型プラットフォームビジネス」ということで、既存のプラットフォームを利用することで自社のビジネスを拡大する「既存プラットフォーム連携戦略」や、既存プラットフォームとの間でのすみわけにより自身がプラットフォームとなる「セグメンティッド・プラットフォーム戦略」の2つの戦略があることが、以下の文献で議論されている。

小宮 昌人 (著), 楊皓 (著), 小池 純司 (著)「日本型プラットフォームビジネス」, 日本経済新聞出版社 (2020/1/25) ここでは、「プラットフォーム」が巨大企業だけのものでないことが強調されている。

(18) 嚙矢をなすものとして、以下が挙げられよう。

i) 小川 絃一 (著)「オープン & クローズ戦略 日本企業再興の条件 増補改訂版; 第2版」翔泳社 (2015/12/4)

ii) 徳田 昭雄 (著), 小川 絃一 (著), 立本 博文 (著)「オープン・イノベーション・システム—欧州における自動車組込みシステムの開発と標準化」晃洋書房 (2011/7/1)

(19) 小川 絃一 (著)「オープン & クローズの戦略思想を必要とする時代の到来 IoT 時代に我が国知財マネジメントをどう方向付けるか」講演資料, 特許業務法人 磯野国際特許商標事務所 50 周年記念講演 (2017/6/14)

Apache License 2.0 のライセンスで構築された OS であって、Linux の OSS ライセンスの GPL (General Public License) や、Apache License 2.0 では、下流側への特許権の行使は制限されるということもあり、上流の開発者は、自己の特許権を下流側に対しては行使できない関係にある。Google にとっては、Android OS に関する技術で仮に特許を取得したとしても、直接的には、当該技術について（下流側の）他者の使用を禁じるというわけにはいかない、という関係にあると考えられる。

ただし、たとえば、Android OS 上でのアプリを作成するためのツールである SDK (Software Development Kit) は、Google が提供するものであり、アプリ開発者は、Google が提供するプラットフォーム上でなければ、Android OS 上でのアプリを開発できない。また、アプリ販売の“Google Play”や“Gmail”などというような標準アプリは、Google が提供するプロプライエタリなソフトウェアであり、プレインスツールで販売するには、Google の（有償の）ライセンスが必要といわれている。

ソフトウェア・サービスをビジネスとするための知的財産

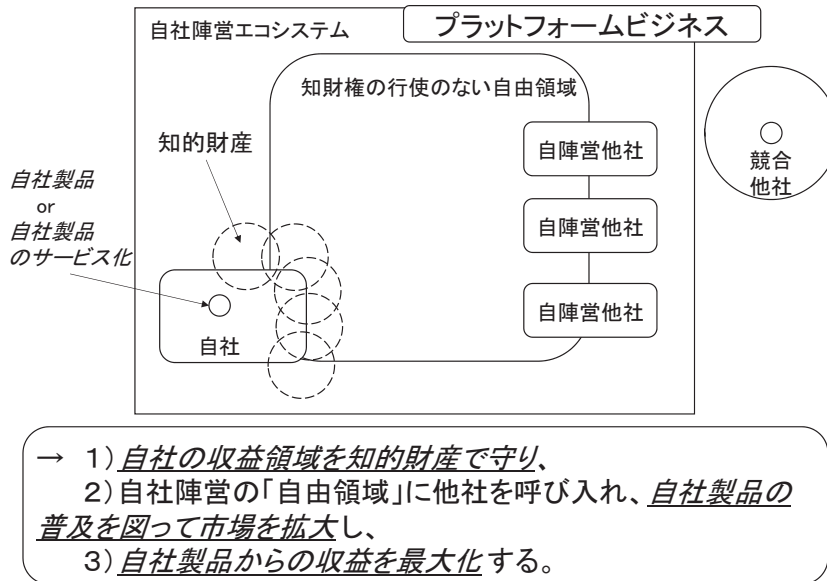


図2. ソフトウェアやサービスをプラットフォーム型ビジネスで実現する際の知的財産

さらに、OS に対する“Android”との商標も Google が所有している。この結果、Android OS のバージョンアップをする権限（改版權）は、Google が所有しているという関係にある。ICT の分野では、バージョンアップ後に（標準的に）使用される技術の内容を自身が決定できる、ということは、次期技術の準備を誰よりも早く実行できるという点で、ビジネスの展開上、圧倒的に有利な立場を持てることを意味するので、ある意味、この「改版權」は、もっとも重要な権利の1つといえる⁽²⁰⁾。

すなわち、プラットフォーム型ビジネスモデルでは、ビジネスの全体を「特許権」等でカバーして他社が入ってこられないようにするという意味で、知的財産権が使用されることはまれである。むしろ、プラット

(20) たとえば、携帯電話機の国際規格において、ヨーロッパの GSM 規格が、i) 携帯端末側と基幹システム側において、携帯端末側の規格は、標準化として完全に公開されていたのに対して、基幹システム側は基本的にクローズであったこと、および、ii) GSM 規格の改版權を有していたのが、携帯端末側と基幹システム側の双方を自社に持つヨーロッパ企業であって圧倒的な競争力を維持する一方、携帯端末側しか製造販売しない外国企業（日本企業を含む）は、結局、市場参入が困難であった点で、「改版權」が、市場を独占する有力なツールであったことが指摘されている。
 小川 紘一（著）「国際標準化と事業戦略 日本型イノベーションとしての標準化ビジネスモデル」白桃書房 2009 年初版発行。

また、たとえば、Adobe 社が、電子文書データの閲覧用のソフトウェアである Adobe Reader を無償で配布している状況などは、ソフトウェア分野において「オープン領域」を提供しつつ、「改版權」を確保するとの戦略であるといえるであろう。

フォーマーは、自社の収益を生む対象となるビジネス領域では、自己の収益源となる一部の領域を知的財産権で守るもの、自社のプラットフォームを利用する他社（自陣営他社）に対して、「知的財産権の行使のない自由領域」を提供して、自己のプラットフォームの利用を促進して技術を広めるとの戦略をとっていることになる。そして、（他の技術フォーマーなどの）「競合他社」に対しては、（知的財産権も手段の1つではあるものの）「ネットワーク効果」等を、当該ビジネス領域に対する最も有効な「参入障壁」として構築する、というような知的財産権の利用の仕方をしていることになる。

2-3. 「プラットフォーム型ビジネス」が他の「プラットフォーム型ビジネス」を駆逐するとき

上記のとおり、ビジネスの独占を生む新たな手法の1つとして台頭している「プラットフォーム型ビジネスモデル」ではあるものの、これまでには、先行して成功していた「プラットフォーム型ビジネス」が、後続の他の「プラットフォーム型ビジネス」に置き換えられてしまうということも、何度も発生している。このような事例には、一定の法則がみられることも知られており、後述するように、現在が、新たなビジネスモデルと考えられる「データ駆動型ビジネスモデル」が生まれようとしている時期にあるとするのであれば、過去のこのような事例を振り返っておくのは、意義があるものと考えられる。

i) ICT技術の分野で、初期にプラットフォームと呼べるものが現れた後に、これが他のプラットフォームに置き換えられてしまった事例としては、「ウェブブラウザ」がある。

最初のウェブブラウザ WorldWideWeb がティム・バーナーズ＝リーによって1991年に公開された後、商用のブラウザとして最初に広まったのは、ネットスケープ社から1994年にリリースされた Netscape Navigator である（一時は、シェア90%を超えたといわれる）。これに対抗して、1995年にマイクロソフト社が、Internet Explorer をリリースして、いわゆる「ブラウザ戦争」が引き起こされることとなった。そして、2002年には、Internet Explorer の利用率は、95%となったといわれており、完全に、Netscape Navigator を圧倒してしまったことが知られている⁽²¹⁾。

一般の認識では、Internet Explorer がこのような勝利を飾ったのは、マイクロソフト社が圧倒的なシェアを有していた PC 用の OS（当時で90%ともいわれる）である Windows に、Internet Explorer を同梱し（初期には、OS と分離不能でさえあった）て販売したことが原因であるとされている。しかも、このような販売方法は、独禁法の観点から、しばしば非難にさらされることにもなった。

ただし、そもそも、ネットスケープ社は、単純に、ブラウザを提供するというだけの戦略ではなかったとされる。すなわち、コンテンツを Web 上で提供するための Web サーバソフトと、ブラウザの両方の製品を市場に提供し、無償でリリースされる機能限定版のブラウザの使用によりユーザを拡大する。その上で、代金を払う用意があるユーザには、機能限定のないブラウザを有償で販売するとともに、基本的に、収益は、Web サーバソフト側から得るとのビジネスモデルであったといわれている。この点では、ネットスケープ社のビジネスモデルは、「Web 上でのコンテンツ提供」という技術プラットフォームを提供する一種のプラットフォーム型ビジネスであったとみることができる。

そして、マイクロソフト社がウェブブラウザのシェアを伸ばし始めたのは、マイクロソフト社自身が、1996年に、Web サーバ製品である IIS (Internet Information Service) をリリースした後である。このとき、マイクロソフトは、これもサーバ OS に、IIS を同梱するというビジネスを開始した。

つまり、マイクロソフトは、自社が圧倒的なシェアを有する OS を有償で購入するユーザには、ブラウザも、サーバ側のソフトウェアも無償で提供することで、これらを有償としているネットスケープ社からシェアを奪うことに成功した、ということになる。すなわち、自社の収益部分を確保したうえで、競合先の収益

(21) 根来 龍之 (著) 「スマートフォン革命で挑戦を受ける業界盟主マイクロソフト プラットフォーム包囲戦略の弱点」, <https://diamond.jp/articles/-/14738?page=2>

部分を無償で提供する、との戦略をとったことになる。

ii) 他の事例として、例えば、日本の携帯電話キャリアの NTT ドコモのビジネスモデルであった“i-mode”が、2019年に新規受付を終了することが発表された。

これは、見方によれば、スマートフォン（特に、iPhone）の登場により、圧倒的な先行するプラットフォームが、後続の他のプラットフォームにより置き換えられた例とみることもできる。

すなわち、“i-mode”は、携帯電話網を介して（つまり、有料の通信料金の下で）、インターネットに接続することにより、メールサービス、ネットワーキング、着メロ配信などのサービスが提供される技術プラットフォームといえる。この場合、取引手数料は、プラットフォームたる携帯電話会社が、携帯電話料金として徴収するのであって、典型的なプラットフォーム型ビジネスモデルといえる。もちろん、単純にただ一つの原因を特定できる、というような話ではないものの、スマートフォンが、WiFiを経由して（つまり、携帯電話網を経由することなく）、コンテンツなどのダウンロードが可能であり、このダウンロードされるコンテンツ（初期的には、音楽データ）の料金は、ダウンロード元のサイト運営者（つまり、スマートフォンの製造元）に支払われる構成となることで、コンテンツを抑えている側に向けて、急速に、サービスの置き換えが進んだ、とみることもできる。ここでも、後続するプラットフォームは、自社の収益部分を確保したうえで、競合先の先行するプラットフォームの収益部分を無償で（または、競合先よりも圧倒的に低額で実現して）提供する、との戦略をとったことになる。

このような後続するプラットフォームが、先行するプラットフォームを置き換える事例は、他にも多数、事例があり、特に、ICTの分野では、一つの典型的手法ということが出来る。この意味では、現在、「データ駆動型ビジネス」が顕在化する過程にあることは、これが「ビジネスモデル」にまで昇華することで、プラットフォームの交代も生じうる状況となっているとも、見ることが出来るであろう。

2-4. 人工知能技術とプラットフォーム型ビジネスモデル

i) Google の Android OS のビジネスモデル

まず、上述した Google の Android OS のビジネスモデルを、図1に示した「プラットフォーム型ビジネスモデル」の図に従って記載すると、図3のようになるであろう。

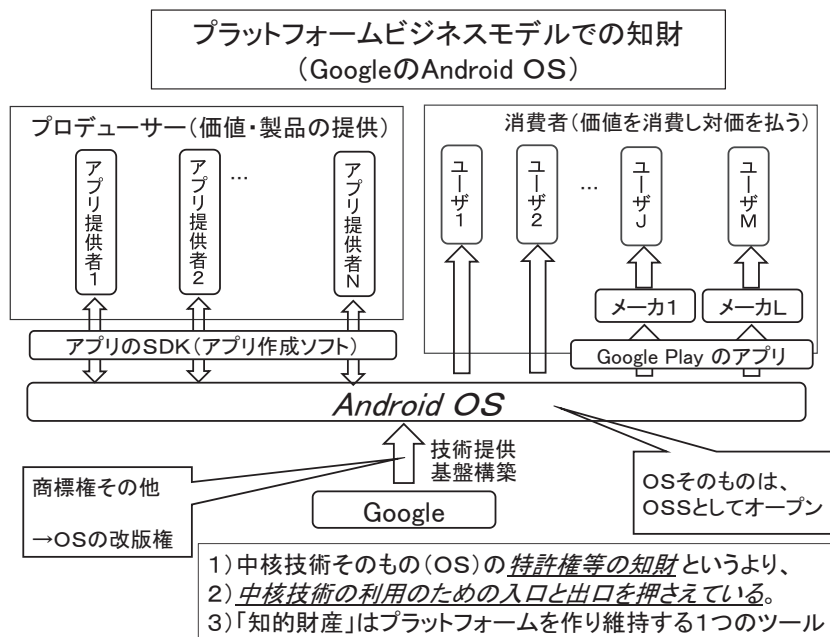


図3. Android OS についてのプラットフォーム型ビジネスモデル

「プロデューサー」としては、もちろん、スマートフォンのメーカーを含めて考えることも可能であるが、ここでは、ソフトウェアの方を中心に考えることとして、「プロデューサー」は、「アプリ製作者」であるものとして、スマートフォンメーカーは、むしろ、最終ユーザとの間で、プラットフォームから提供されるアプリを使用する側にあるものとして記載している。

図2で説明した「知的財産権の行使のない自由領域」をソフトウェアの世界で実現する方法としては、「技術標準」とすることもありうるが、ここでは、OSSが利用されていると考えることができる。

そして、「アプリ製作者」側はSDKにより、「ユーザ」側はGoogle Playにより、いずれも、プラットフォームの提供者がコントロールできる体制となっている。特に、アプリに関して言えば、Google Playを介しての販売に限定することとすれば、プラットフォームは、取引手数料として、収益を上げる構造とすることができることになる。また、アプリ内決済にも、Googleが一定の手数料を徴収しているとされる。

ii) 人工知能技術に関する「プラットフォーム型ビジネスモデル」とは？

2016年の“This year’s Founder’s Letter”において、GoogleのSundar Pichai CEOは、以下のようなコメントを述べておられる。

“Looking to the future, the next big step will be for the very concept of the “device” to fade away. Over time, the computer itself - whatever its form factor - will be an intelligent assistant helping you through your day. We will move from mobile first to an AI first world.”

このようなコメントに従って、人工知能技術を上記のようなプラットフォームに組み込むために、以下のような施策が、次々と実行に移されてきている。

- a) 人工知能技術において、最も計算負荷の重い「学習処理」を実行することが可能なサービスとして、“Google Cloud Platform”上でのツール等が公開されている（一定条件下で有料）。なお、Googleは、機械学習の計算に適したH/Wとして、一般に使用されているGPU（Graphical Processing Unit）よりもさらに高性能であるTPU（Tensor Processing Unit）を開発をしている。
- b) 機械学習に用いるためのソフトウェアライブラリが、OSSで“TensorFlow”として公開されている（OSSのライセンスとしては、Apache 2.0）。さらに、モバイル機器用に、“TensorFlow Lite”も、Android, iOS, Raspberry Pi向けのコードが公開されている。これにより、モバイル機器では、H/Wに依存せずに、Googleの人工知能技術が利用できる環境となっている。
- c) さらに、Googleは、IoTなど端末側での人工知能演算のために、エッジ向けTPU「Edge TPU」の外販も開始している。

このような構成は、Google以外の他社が提供する人工知能ソフトウェア開発プラットフォーム⁽²²⁾によっても、多かれ少なかれ、実現され、あるいは、実現されようとしているものである。いずれにしても、人工知能学習においては、クラウドベースでのPaaS（Platform as a Service）の形式となって「従量課金」が行われ、また、端末側の処理において、推論等の処理を実行しやすい環境を整えることで、自社の「人工知能技術」の使用を広める戦略がとられていることになる。

ただし、現在のところ、「人工知能ソフトウェア開発プラットフォーム」では、人工知能の学習処理を実行するソフトウェア、学習処理の結果得られる「学習済みモデル」の実行環境を提供するソフトウェアなどに、その重点があり、すでに、人工知能技術は、コモディティ化した、との見解まで現れる状況となってい

(22) たとえば、マイクロソフト社は、Microsoft Azureとしてサービスを提供しているし、Amazon社は、Amazon AWSとしてサービスを提供している。日本でも、株式会社Preferred Networks（PFN）からは、Chainerがリリースされている。もっとも、PFNは、2019年12月5日、フレームワーク開発を終了してChainerはメンテナンスフェーズへ移行すること、自社はChainerからFacebookが主導するPyTorchに順次移行することを発表した。

る⁽²³⁾。しかしながら、上述したような「人工知能技術の両輪」ともいえる「データ収集・整理」の部分については、現在のところ、技術開発だけでなく、制度の整備という点でも、「まだ、これから」という点が多く残されている状況と考える。重要なことは、たとえ、自社が、他社の提供する人工知能プラットフォームのユーザ（大企業だけでなく、中小企業も含む⁽²⁴⁾）となるとしても、「競合他社との差別化をどの部分で実現するのか？」また、「人工知能プラットフォームとユーザたる自社との間で、ウィン・ウィンの関係をどうすれば維持できるのか？」ということであろう。

3. データ駆動型ビジネスモデル

3-1. 「データ駆動型ビジネス」とは？

3-1-1. データ駆動型人工知能

「人工知能」については、特に、2015年10月に米DeepMind社が作成した「AlphaGo」が人間のプロ囲碁棋士に勝利して、以降は、「データ駆動⁽²⁵⁾型人工知能」技術の典型であるディープラーニングと呼ばれる手法が注目され、新たに「第三次人工知能ブーム」といえる状況となっている。

このような状況となった大きな原因の1つは、ディープラーニングなどの技術により、大量のデータから対象の計算モデルを構築し、これにより対象の分類、予測、操作ができるようになったものであり、（それまでは、人間が人手で設定していた）学習時に必要となる「特徴量」をコンピュータが“自動的”に抽出することが可能となったことが、大きな要因であるとの指摘がされている。

たとえば、将棋において、機械学習の手法を初めて導入したともいえるソフトウェアの「Bonanza」では、盤面の先手後手の有利不利を判断するための「盤面の評価関数」を、（盤面上で3角形の配置となる）「玉を含む3つの駒の配置パターン」という人間の設定した「特徴量」で評価して、この特徴量に対する評価値（重み）を「過去の棋譜データ」に基づいて機械学習するという構成であった。

これに対して、碁では、盤面に存在するのが「白と黒の石」のみであり、（個々の駒の重要度に差をつけられる）将棋のような特徴量の設定が人間によるのでは難しく、「盤面の評価関数」を作ることが人手では困難であったところ、「AlphaGo」では、畳み込みニューラルネットワークという「人間の視覚認識のモデル」を利用したディープラーニングを利用することで、一定以上の性能の「盤面の評価関数」を形成することが可能となり、さらに、これを利用することで「強化学習」との手法（コンピュータ同士の対戦により学習する）を用いて、ついには、人間のトップレベルの棋士に勝利するに至ったとされている。

ただし、上記のとおり、「特徴量」の設定自体が、「学習用データ」からコンピュータが抽出するとの構成であるために、人間には、その特徴量が示す意味を直ちには理解できず、その結果、判断のアルゴリズムが

(23) 「コモディティ化」という概念自体は、特に、電気・電子分野でのハードウェアの製品・部品等について、使用される場合が多い。ただし、上述のとおり、「データ収集・整理」の部分の整備が、これから、という状況では、これを人工知能技術について当てはめるのは、まだ、時期尚早といえるのではないだろうか。

(24) ここまでの事例は主として、「グローバルな超巨大企業」であるものの、（データ駆動型ビジネスモデルも究極はそうなる予想される）「プラットフォーム型ビジネスモデル」を、大企業だけの特別なものとしてみるべきではないと考える。巨大企業ではなく、中小企業であっても、自身のビジネス領域で、「プラットフォーム」といえるものを構築しているor構築しようとしている例は、多く存在する。たとえば、シードステージの日本のベンチャー企業である株式会社アメグミは、ビジネス向けのスマホのローエンド領域でのプラットフォームの展開を図っている。ビジネス向けのアプリの供給プラットフォームの構築とAndroidベースのOSのアップデートに注力し、H/Wは中国の深センで1端末を原価35USDで製造して、国内および海外への販売ビジネスを目指している。参入障壁は、スイッチングコストと、アプリ供給の仕組みといえるであろう。

(25) 「データ駆動」との用語を、「データに基づいて、〇〇をする」程度の意義で使っている場合が散見されるが、「データに基づく経営判断」などは、「データ駆動型人工知能」という概念が現れる前から実行されていたものとする。この点で、本稿では、「データ駆動」との用語は、それ以上の意義を有する内容を指すものとして使用する。

「ブラックボックス」となってしまう、ということが問題点の1つとも指摘されている⁽²⁶⁾。

このブラックボックス化については、“論理知識型 AI”と“データ駆動型 AI”をハイブリッド化したり⁽²⁷⁾、「説明可能な人工知能 (XAI: eXplainable AI)」に対する研究⁽²⁸⁾も行われ、その判断過程を人間が理解できるようにしようとする試みは多くなされている。ただし、「特微量」の設定自体は人間が行わない(人間による演繹的または経験的な推論で導き出したものでないため、一義的には、事前の人間による設計のみに従って判断が行われるわけではない)という点が、現時点での人工知能の大きな特徴の1つであることには、変わりはないと考える。

3-1-2. 「データ駆動型ビジネス」と言われる例

i) 米国の Deere & Company (ブランド名 John Deere) のビジネスモデル⁽²⁹⁾が「データ駆動型ビジネス」の一例といわれる。John Deere は米国最大の農業機械企業であり、以下のようなビジネスを展開している。すなわち、2015年に大胆なデジタル変革戦略に着手し、自社のビジネスモデルにデータを取り入れ始めた。同社は現在、自社の農業機械の車両に 3G/4G の接続機能を組み込み、米国全体にわたって農機の作業状況に関するデータを収集している。

John Deere のビジョンは、このデータを集約して分析し、第三者に販売することをビジネスとしようとしている。たとえば、以下のようなデータの利用が考えられる。

同社のデータサイエンティストは、米国で作付けされたトウモロコシ、ジャガイモ、砂糖などの量に関するデータをほぼリアルタイムに収集する。そして、特定の自治体/州/地域などの集計条件を指定も可能である。このデータは、より緻密な需要予測を必要としている種子関連企業、化学企業、銀行、政府など農業エコシステムに関わる組織に販売する。

ii) また、日本でも自動車保険として、たとえば、あいおいニッセイ同和損保が、「タフ・つながるクルマの保険」⁽³⁰⁾を販売している。

ここでは、保険料は、「基本保険料」と「運転分保険料」で構成され、「運転分保険料」は、「毎月の走行距離 (1km 単位)」に応じた保険料に、毎月の「運転特性による割引」を適用して算出されるとしている。

iii) さらに、日本の生命保険では、「がん保険」の分野で、これまでの「経験的な分類」を、「データに基づく分析による分類」に置き換えようとする動きがある。

たとえば、従来は、「臓器ごと」に効果が予想される薬が投与されていたのに対して、ゲノム医療では、発がんの原因となる遺伝子変異を解析することで、効果のある薬をピンポイントで探し当てられることを想定して、「がんゲノム医療」⁽³¹⁾が保険収載され、東京海上日動あんしん生命は、がんゲノム医療に対応した保険商品の販売を開始⁽³²⁾している。

あるいは、明治安田生命は、被保険者に定期的に健康診断の結果の提出を求める代わりに、その結果に応

(26) 特に、法律関係者の間では、たとえば、人工知能による「自動運転」が実用化されたとして、事故が起きた際の「責任」をだれが負うか、という際に、この「ブラックボックス」の状態が問題であると指摘されることが多いようである。ただし、そもそも、その前に、「製品レベルの人工知能ソフトウェア」をリリースするにあたって、その「品質評価・品質保証」をどうするのか、ということが、より直近の問題として存在する。これまでは、ソフトウェアの品質保証は、原則として、「適正な手続きで設計処理が実行されていること」を根拠としてきているが、「ブラックボックス」となると、手続きの適正性だけでは、「品質保証」としては不十分という見解もある。

(27) 我妻 広明 (著)「データ駆動型人工知能と論理知識型人工知能の融合による解釈可能な自動運転システムに関する研究」, 「次世代人工知能・ロボット中核技術開発」(人工知能分野) 中間成果発表会 (平成 29 年 3 月 29 日), https://www.airc.aist.go.jp/info_details/docs/170329/1730-Wagatsuma.pdf

(28) 川村 隆浩 (著)「機械学習の説明可能性への取り組み—DARPA XAI プロジェクトを中心に—」
https://www.jst.go.jp/crds/sympo/201906_JSAI/pdf/02.pdf

(29) マーク・アインシュタイン (著)「IoT の未来を担うのは「データ駆動型ビジネスモデル」なのか」, DIAMOND IT&ビジネス <https://diamond.jp/articles/-/146483?page=3>

(30) <https://www.aioinissaydowa.co.jp/personal/product/tough/tsunagaru/>

(31) 「がんゲノム医療、見えてきた課題」, 日経ビジネス <https://business.nikkei.com/atcl/gen/19/00002/102900810/>

(32) <https://www2.tmn-anshin.co.jp/download/775/20190808news.pdf>

じて、保険料を減額する「健康サポート・キャッシュバック」⁽³³⁾などのサービスの提供を開始している。当然ながら、このようにして収集された「健康診断の結果」を、人工知能技術と組み合わせて、新たなサービスの開発および提供を検討することも発表されている。

ただし、たとえば、このようなヘルスケア分野での1つのハードルは、被験者が何らかの疾患になった後に、医療機関で治療を始めるなどの事情が生じたときに、「健康診断」の結果と「診療情報」との間が、連結していない（同一人の健康診断の情報と診療情報とが連続したデータとなっていない）ことなどが課題となっている。これは、「個人情報」としての扱いが最もセンシティブな領域であるが故に、高いハードルがあることは当然としても、「国民全体の医療レベルの維持・向上」という観点からは、ぜひとも解決すべき課題であることは間違いない⁽³⁴⁾。

少なくとも、米国の i) の例は、「データ」を商材とするビジネスではあっても、筆者の理解では、後述するように、「データ駆動型ビジネス」とまでは言えないものとする。

一方で、ii) と iii) を比べた場合には、iii) の「がん保険」の方には、「これまで人間が経験的に用いてきた基準でないものをデータから抽出している」という点で、より「データ駆動」に近いものが現れてきているものとする。ただし、たとえば、「ゲノム診断」の結果と、「抗がん剤の投与」との関係では、医薬品の適応外使用の問題なども検討課題で、現在のところ、まだ研究段階の部分も多い。

3-1-3. 「データ」を「取引対象」にするということと、「データ駆動型ビジネス」の関係

i) 「データ」をビジネス上の「取引対象」にするということ

以下では、人工知能技術との関連から、ビジネス上の取引対象となるデータを「学習用データ」に焦点をあてて考えることにする。そして、状況によっては、(それをデータと呼んでよいかはおくとして)学習用データで学習した「学習済みモデル」自体が、取引対象となる場合もありうるものとする。

a) 「データ」をビジネス上で取引する可能性のある主体は？（「顧客」はだれか？⁽³⁵⁾）

一般には、データの利用者と呼ばれる主体としては、以下のような者が想定される。

i) 一次利用者…利用者自身が、受益者（データ提供者）に直接還元する目的として、及び、自身の事業（たとえば、将来、データ提供者へ製造販売等される可能性のある製品の改良）を目的として収集し「データ」を利用する場合を意味する。

ii) 二次利用者…一次利用の目的外の使用、たとえば、第三者の研究開発目的での使用や、社会的な健康・安全・危機管理、教育などを目的として、一次利用者以外の者が「データ」を利用する場合である。

ただし、これだけの主体を想定したのでは、そもそも、二次利用者自身が、「生データ」から、データ分析を行い、人工知能システムを開発するリソースを有している状況は、（将来はおくとしても）現在のところは、まれであると想定されるので、少なくとも、「データ分析・人工知能システム開発」の業務を、（一次又は）二次利用者から受託する主体を想定することが必要であろう。

また、商材としての「データ」を考えたときには、その特有の事情を考慮する必要があると考える。

データの特徴1：物理的に「データ」を独占する手法は限られ、一旦、流通過程に乗ってしまうと、データ提供者にとって、そのデータがどのように使用されるかをコントロールすることは、厳密には困難である。

(33) https://www.meijiyasuda.co.jp/find/list/beststyle/contract/content/pdf/content_03.pdf

(34) この点の解決策として、後述する「次世代医療基盤法」が制定されたともいえるので、今後は、この法律の下で、現実にどのような運営がされるのか、という段階にある、といえる。「国民の健康・医療」に直結するだけに、その目的を第一義に柔軟な運営と、安全・適正・公平な運営とが両立されることが必要と考える。特に、「健康・医療情報」については、国の健康保険制度、民間の生命保険等を介して、「日本の国富そのもの」とも関連している分野であり、最大限の考慮がなされるべき分野といえる。

(35) ビジネスのマネジメントの分野では、ピーター・ドラッカー氏の、ビジネスに関する「5つの質問」が有名である。その2番目の質問は、「われわれの顧客は誰か？」というものである。そして、ドラッカー氏によれば、企業の目的とは、「顧客を創造すること」とされている。

つまり、仮に、「データ」に対して、特許権のような「絶対的排他権」が認められたとしても、利用者自身が、当該「データ」を使用することを自認する場合を除いて（たとえば、技術標準となっている「データ」を自ら申し出て使用する場合などを除いて実施行為の特定が難しく）、一般には、侵害行為の摘発自体が困難である。この点で、「データ」が、その保持者から直接提供を受ける主体に取引される場所までは契約による規制が想定されるものの、それが第三者にまで流通される状況を考えると、特許権により保護が行えるのは、（データが形態を維持したまま使用される必要があるような）限定的な局面と考えられる。

データの特徴2：そもそも、「情報」は、それを保持する主体のみが、当該「情報」を保持していることに経済上の価値があり、完全に公開されて万人に共有されてしまえば、一般には、その商業的な価値は激減してしまう場合がほとんどである⁽³⁶⁾。

データの特徴3：人工知能の学習用データとして検討した場合、重要なことは、一般には、「学習用データ」として必要な「生データ」を、単一の一次利用者のみが多量に有していると想定される局面は限定的であること（むしろ、個人情報保護の観点からも、限定的であることが望ましい）である。

以上の点から、筆者としては、上述したような「一次利用者」と「二次利用者」との間に、「データ収集／分析者」とも呼ぶべき存在を想定されるべきと考える。これは、以下の理由による。

- i) そもそも、各種データは（特に、有用性の高いと考えられるデータであるほど）単一人・組織のみに帰属していることは想定できず、複数の人・組織からデータを集めることで、初めて「ビッグデータ」と呼べる規模の量が得られる場合が多いと考えられること⁽³⁷⁾。
- ii) 人工知能を学習させる局面などでは、そもそも、生データの内容のすべてを二次利用者が利用できる環境となるとは想定できないこと。つまり、一般には、「生データ」について、加工・調整（典型的には、「外れ値」除去や、属性項目の整理）後の「学習用データ」が、取引の対象となると考えられる。

特に、個人情報に関連するような情報であれば、そもそも、「匿名加工」のために、属性などの一部は、欠落した状態で、二次利用者が人工知能の学習をさせることが必要となると考えられる。しかしながら、人工知能技術が「データ駆動型」であることにより、事前に、必要となる「属性」を予め完全に規定できるとは限らず、当該分野における人工知能技術の展開・発展にしたがって、「属性データ」の追加・交代などが必要となる局面が想定される。そこで、このような「追加・交代」によって生じたニーズに対処する者が必要になるものの、このような業務を、「二次利用者」自身が実行することは、自身がすでに保有しているデータについての「個人」に関わる情報を「二次利用者」自身も保有していることが必要となってしまう、個人情報保護を遵守するとの観点からも、「二次利用者」の使い勝手がよいとはいえない⁽³⁸⁾。

たとえば、当初は、「心臓の疾患」に対する「疾患リスクの予測」を目的として、データが集積され、そ

(36) したがって、現在のところ、「データの取引」の契約のベースとなる考え方は、「秘密保持契約」の考え方がその背景にあることが多いが、「秘密保持」ということと、「データの取引」ひいては「データの流通」ということとは、その性質が大きく異なり、別の観点からの契約の体系が必要と考える。

(37) 複数の人・組織からデータを集める場合（従来であれば、「データベース研究」等と呼ばれる部類の場合）には、後述するような「バイアス」のうちの一部は、統計モデルなどを使用して、統計的に除去する方法があることが知られている。ただし、いわゆる疫学研究などの分野では、従来から指摘されているように、本質的には、「バイアス」を除去するには、適切な「研究デザイン」（本稿のような場合は、「データ収集・分析デザイン」とでも呼ぶべきか）が必要といわれる。野尻 宗子（著）「バイアスと交絡：医療データベースを使った薬剤疫学研究」、日本薬学会「薬学雑誌」乙号 135（6）pp.793-808（2015年）には、以下の記載がある。

“推定における誤差は系統的あるいは偶然誤差に区分される。疫学では、「系統的な偏り（systematic bias）」が「バイアス（bias）」を指していることが多く、バイアスの反対が「妥当性（validity）」である。また、「偶然誤差（random error）」の反対は「精度（precision）」である。「ランダムエラー」はサンプルサイズ増で解決されるが、バイアスはサンプルサイズとは独立に発生する。…Kleinbaumらは、バイアスを選択バイアス、情報バイアス、交絡に分類した。”

(38) たとえば、「個人情報保護法」では、「匿名加工情報」（病歴等の要配慮個人情報については別の観点からの配慮がさらに必要になる）については、もはや個人情報でないため、第三者提供が可能となっているものの、「匿名加工」されているということは、個々のデータが「誰のデータであるか」がわからなくなっているのであるから、事後的に、「匿名加工情報」を有している二次利用者は、「属性データ」の追加・交代を実行できなくなっていることになる。

れに基づく人工知能が生成されたとして、他の研究の成果などにより、当該データの一部が、「認知症の疾患リスクの予測」にも使用できる可能性が判明したとしても、人工知能技術の開発担当組織としては、「認知症」に関するデータの項目については保持していない、というようなことが想定される。

とすれば、特に、生データについては、「個人情報」（そのものとは言えないまでも）に関連する情報が含まれる場合に、「個人情報」の保護の観点からは、「データ収集者」とも呼ぶべき仲介者が間に介在することが必要となる場合が想定される。

たとえば、個人情報として一番厳しい制約が存在すると考えられる「医療情報」について、いわゆる「次世代医療基盤法（医療分野の研究開発に資するための匿名加工医療情報に関する法律 平成29年5月12日公布）」⁽³⁹⁾においては、「認定匿名加工医療情報作成事業者」は、少なくとも、このような「データ収集者」となることが想定されているとも考えられる。

ただし、後述するように、上記の「仲介者」は、「仲介業者（Data Broker）」としての機能と、状況によっては、後述するように、「分析者（Data Analyst）」としての機能も有している必要が想定される⁽⁴⁰⁾。

以上の考察から、データが、ビジネス上の取引対象となるためには、以下の主体について、何らかのインセンティブが必要となると考えられる。

- ・「一次利用者」が、データを外部に提供するインセンティブ（端的には、経済的な対価）
- ・「二次利用者」が、データを自己の業務に利用しやすい環境
- ・「一次利用者」と「二次利用者」との間をつなぐ「データ収集／分析者」のインセンティブ

b) ビジネス上で取引する可能性のある「データ」は？

そこで、まずは、「データ」あるいは「情報」の意義を、改めて、考えてみたい。

データ駆動型人工知能技術の成功により、「データ」の経済的な重要性が著しく大きくなったことについては、上述のとおりである。

そして、このような「データ」は、一般には、「ビッグデータ」と呼ばれるように、一定以上の「量（Volume）」を有していることが前提とされる。

ただし、留意が必要なのは、「ビッグデータ」といった場合、「量」が存在することで、データのばらつきが、相対的に小さくなることが期待されると理解されている場合が多い。しかしながら、人工知能の学習用データについては、第一義的には、「統一された条件・共通する属性」で収集された「性質のそろったデータ」であることが、「よいモデル」を作成するためには、当然ながら、望ましい。「統一されていない条件・共通しない属性」で収集されたデータで、学習を行うこと自体は可能であるとしても、その場合は、「性質のそろったデータ」に比べて、けた違いのデータ量が必要になるか、そもそも、そのようなデータでは、満足なモデルが生成できない可能性すらある。

たとえば、データには、ランダムに分布する「ばらつき」以外に、一定の傾向をもった「バイアス」が存在することが統計学上は知られている。「ばらつき」は、データ量により相対的に低減されうるが、「バイアス」については、そのための統計的手法を用いなければ、除去することは困難である。

この点は、人工知能技術には、「データの量と質が極めて重要」と指摘されるゆえんである。

次に、日本語で「情報」と呼ばれるものを考えるとき、英語では、以下の3階層があるといわれる。

- i) “DATA”: 数字や文字列、特徴や出来事に関わる記述などを収集したものであり、それだけでは、

(39) 『「次世代医療基盤法」とは』、内閣官房健康・医療戦略室、内閣府日本医療研究開発機構・医療情報基盤担当室（2019年4月）

(40) 「情報銀行」については、データを財産とみて、一種の「信託銀行」の機能を果たすものと想定されていると理解する。ただし、「データ」は、動的に変化するものであるし、後述するように、個人情報保護の観点から、直接、学習用データを二次利用者に提供することが困難であるために、むしろ「学習済みモデル」として提供することが必要となる局面も想定される。

どのような意味があるのか分からない状態の素材に当たる「情報」

ii) “INFORMATION”：DATA をなんらかの基準に基づき構造や体系を与え整理した「情報」

iii) “INTELLIGENCE”：INFORMATION を必要性に基づき取捨選択し、内容を分析し、価値判断を与えられた「情報」

状況によっては、DATA と INFORMATION とは、ほとんど同じ意味で使用されている場合もあるが、ここでは、上記のような区別を元に議論することにする。

ここで、当然ながら、経済的に最も高い価値を有しているのは、“INTELLIGENCE”⁽⁴¹⁾である。

そして、上述した通り、「情報」は、完全に公開されて万人に共有されてしまえば、一般には、その商業的な価値は激減してしまうことを考慮すると、ビジネスにおいて、情報が価値を持つ状態としては、少なくとも、以下のような2つの状態がありうる。

ビジネス状況1) “DATA”または、少なくとも“INFORMATION”については、特定のビジネス主体に独占されている状態。この状態であれば、このビジネス主体は、競合相手よりもビジネス上の有利な位置を確保でき、競争に打ち勝つための大きな財産を有していることになる。

ビジネス状況2) “DATA”または“INFORMATION”自体は、広く世間に流布しているものであっても、特定のビジネス主体は、これを分析して“INTELLIGENCE”とする能力において他を圧倒しているという状態。このときも、このビジネス主体は、競合相手よりもビジネス上の有利な位置を確保できることになる。

たとえば、ビジネスにおけるデータ解析システムである「ビジネス・インテリジェンス (BI)」システムにおいては、既存の基幹系の業務システムにおいてデータが収集される「業務DB」からデータをさらに収集して、ある基準にしたがって、構造または体系を与えて整理する処理である「ETL (Extract (抽出), Transform (変換), Load (ロード))」を行った後に、DWH (Data WareHouse) に (技術によっては、仮想的に) 統合して格納され、この統合された後の情報が分析されるのが一般的である (このような ETL のことが、「データプラットフォーム」と呼ばれている場合もある⁽⁴²⁾)。BIにおいては、このDWHに格納される情報が、INFORMATIONであり、このINFORMATIONからさらに、一種の「洞察」を加えて生成されるものが、INTELLIGENCEであって、かつては、「洞察」は、分析者たる人間が担っていた。現在では、この「洞察に対応する処理」が、「人工知能」による処理に置き換わっていきつつあるのが現状である。

ただし、人工知能技術として考えた場合、実は、“DATA”または“INFORMATION”には、その秘密とされるべき程度として、階層構造があることにも注意が必要と考える。たとえば、上述したようなBIを人工知能技術により実現する場合には、AI自体においても、以下のような階層を統合して1つのシステムが作られていると説明されている⁽⁴³⁾。

- a) 企業個別モデル層
- b) 業界別モデル層
- c) ベース AI モデル層

当然ながら、「企業個別モデル層」は、個々の企業内の「学習用データ」に基づいてカスタマイズされるものであるから、「学習用データ」自身も「学習済みモデル」も外部に対しては秘密状態とされることが必要である。これに対して、「ベース AI モデル層」の人工知能は、「ニュースなど公開されたデータ」を「学

(41) CIA (米中央情報局) と呼びならわされている組織の正式名称は、“Central Intelligence Agency”であることにも、これが現れているといえるであろう。

(42) ただし、本稿では、たとえば、対象となるデータが、1社内のデータに限定される場合は、ETLと呼び、複数社にまたがるようなデータが対象となる場合を、「データプラットフォーム」と呼んで、区別することにする。これは、前述したように、「プラットフォーム型ビジネス」を、本稿では、「N × 1 × M」のモデルと考えていることにより、「プラットフォーム」との用語も、そのような場合を指すものとして使用するからである。

(43) 「企業データを価値に変え、ビジネス・プロセスを変える AI」、IBM ホワイトペーパー、IBM Watson | ビジネスのための AI | <https://www.ibm.com/watson/jp-ja/>

習用データ」とすることでも、学習可能なものであり、「業界別モデル層」も、法令、論文、ガイドラインなど、その業界に特化はしているものの、原則的には、「公開データ」を「学習用データ」とすることで、学習可能なものと想定される。

なお、他の「人工知能技術を用いるシステム」であっても、ある程度以上の規模のもの（商業的に製品とされて利用可能なもの）として提供するには、異なるシステムについて共通に使用できる部分は、共用モジュールとして使用して構築されると考えられるので、上記のような階層構造は、いずれの分野においても必要になるものと想定される。

そこで、「市場で取引されるデータ」（需要者が「有償で取得することを欲するデータ」とは、どのようなものであるのか、を想定して制度設計をすることが必要と考える。「経済上の取引の対象となる」または「流通する」ことが想定されるデータとは、以下のようなものになる。

データ類型0）「**ネット等公開情報**」：インターネット上で公開されており、インターネット経由で取得することが可能な（公開された）データ群

データ類型1）「**社会・インフラデータ**」：データとしては、現実世界から取得することが必要なものであるものの、データの「インフラ」として、特定の複数企業間で共有することが想定されるデータ群（たとえば、自動運転のための「地図データ」などが想定される）

データ類型2）「**個人等関連データ**」：特定企業内の個別のデータ、または、個人の履歴等に関する情報（（個人情報保護法にいう「匿名加工」に至らない状態で）個人情報を含む可能性があり得る）のように、通常は、「秘密」とされることが想定されるデータ

「データ類型2のデータ」とは、たとえば、クレジットカード会社が保有している「各ユーザの購買履歴と、当該ユーザの性別、年齢、職業、年収、…などを関連付けたデータ」が想定される。ただし、このようなデータは、特に、履歴データが詳細になればなるほど「個人情報」に近づくことになり、PDSにおいては、プライバシー保護の観点から、個人情報保護法上での「本人同意」の枠組みを拡大するものとして、情報を提供する個人が、自身のデータの扱いをコントロールする仕組みであることが想定されている。

データ類型3）「**技術標準データ**」：（仮に、個人に由来するとしても、原則として個人情報を含まないといえるところまで匿名加工され）「特定の技術分野において、成果物の性能のベンチマークを提供できるようなデータセット」または「特定の技術分野において、成果物の性能が所定の基準を満たすかを判断できる標準を提供できるようなデータセット」

「データ類型0のデータ：ネット等公開情報」については、上述のとおり、すでに改正された著作権法上の手厚い処置（改正著作権法第30条の4、第47条の5）がとられたことにより、多くの場合は、「学習用データ」として利用することの障害は、自身で手間とコストをかけることをいとわなければ、相当に低くなったといえるであろう。逆に言えば、ビジネスとして流通するのは、このようなデータのうちでも、「手間とコスト」がかかるデータなので他社から購入することが経済的という場合になろう。このタイプのデータは、単に相対取引での「商業上の取引の対象」という以上に、広く「流通」するデータと想定される。

「データ類型1のデータ：社会・インフラデータ」については、特定企業のデータであれば、不正競争防止法上の営業秘密（第2条第1項第4号～第10号）として、また、「特定の複数企業間で共有することが想定されるデータ群」については、平成30年に改正された不正競争防止法の「限定提供データ」⁽⁴⁴⁾としての保護（第2条第1項第11号～16号、第2条第7項、第19条第1項第8号）が想定される。

(44) 第2条第7項（定義）では、“この法律において「限定提供データ」とは、業として特定の者に提供する情報として電磁的方法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（秘密として管理されているものを除く。）をいう。”とされる。

3-2. 「データ駆動型ビジネスモデル」は、すでに存在しているか？

以上の検討を基に、「データ駆動型ビジネスモデル」といえるものが、すでに存在しているといえるのか？について、以下、検討をしたい。

(1) 「データ」を取引することそのものが、「データ駆動型ビジネスモデル」とはいえない。

データ駆動型ビジネスを、「データの流通」という観点のみでとらえてしまうと、これは、上述した「ビジネス状況2」を想定していると考えられるが、とすると、初期的には「データを分析できる主体」は、限られたものになることは当然であるとしても、人工知能技術の発展によって、特に、データ類型0のデータでは、他社との差別化が困難となることが想定される。

とすると、データ類型0のデータについては、「データを流通」させることで、利益を得る、というビジネスでは、「データ駆動型ビジネスモデル」といえるほどの広がりを持ちうるのかは、疑問がある。

さらに言えば、「データ駆動」ということの意義は、決して、“DATA”または“INFORMATION”に存在するのではなく、“INTELLIGENCE”にこそ存在するものとする。

つまり、筆者としては、「データ」が取引の対象となっていることは、表面的な事象であり、「データ駆動」とは、それ以上の何か、が実現される状況と考える。

(2) 取引におけるユーザにとっての「価値」は、「(生) データ」に存在するわけではない。

さて、「人工知能技術」にとって、最重要な「データ」は、「学習用データ」である。これは、決して「生データ」ではないのだから、上記の分類でいけば、INFORMATIONに相当することになる。

この意味で、まず、第一義には、人工知能技術において、「取引対象」となるものは、経済的に価値を有するものは、INFORMATIONたる「学習用データ」であることになる。さらに、INTELLIGENCEである「学習済みモデル」も、「転移学習」など、他の用途への応用が可能なものについては、「取引対象」となる可能性がある。そこで、「学習用データ」について考えてみることにすると、「データ」とは呼ばれるものの、これは、上記の分類でINFORMATIONである以上、BIについて言及した“ETLに相当する機能を果たすもの”が必要ということになる。たとえば、以下のような状況が想定される。

- i) 明示的に「個人情報」に相当するもの、または、「匿名処理がされている情報」ではあっても、「匿名加工」（個人情報保護法上は、もはや個人情報ではない）に至らない状態が想定されるものが必要となる状況もあり得るものの、「学習用データ」の段階では、適切に個人情報に対する対処がなされる必要がある。このような対処を実行する主体が存在しなければ、円滑に、データを取引対象とすることができない。
- ii) あるいは、場合によっては、「匿名加工」してしまうと、「学習用データ」として意味をなさなくなる場合も想定される。特に、医療情報などでは、要配慮個人情報（患者の縦断的な診療情報）が、匿名加工されてしまうと、医療上の「洞察」を実施するには、不十分になってしまう場合もあるものと想定される。この場合は、「適切な個人情報に対する対処」に加えて、「学習用データに対する学習処理」までを実行して、委託者に処理結果であって、もはや個人情報を含まない「学習済みモデル」を提供する主体が必要となることも想定される。

以上の意味で、「データ収集／分析者」とでもいえる主体が想定されなければ、そもそも「人工知能技術」にとって必要な「データ取引」を意義ある状態で、実現することは困難である場合があると想定される。

情報銀行の制度は、確かに、「データ取引」を活性化できる可能性を有する制度であるとはいえるものの、個人情報の管理をPDSに依存する方式とすれば、（そのようなデータの提供に進んで同意する主体からのデータのみが収集されるという点で）上述した「情報のバイアス」が存在する可能性が高くなる⁽⁴⁵⁾。

(45) 医学研究などの分野では、「ダイナミックコンセント (Dynamic consent)」が試行されている。被験者と研究者との間で、主に既存資料・情報での研究開始時のインフォームドコンセントを被験者個人の端末経由で実施しようとするものである。場合によっては、被験者の予めの同意を原則として、オプトアウトで研究を実施しようとする案もある。このような方式を、データの商業利用に適用するには、どのような運営が必要とされるかも含めて、検討課題といえる。
国立研究開発法人日本医療研究開発機構理事長 末松 誠 (著)「ダイナミックコンセントについて」第14回ゲノム医療実現推進協議会 (令和元年6月28日(金)) <https://www.kantei.go.jp/jp/singi/kenkouiryou/genome/dai14/siryou4.pdf>

また、データの提供を受ける側からすると、「学習用データ」としての品質を保証する仕組みが、必ずしも明確ではなく、この点に対する対処が、いかに実現されるかが成否を握るといえるであろう。その意味では、「情報銀行」の制度においても、「銀行」自体が、状況により、「分析者」の役割を果たせるような制度設計も検討の価値はあるものとする（そもそも、分析したことがある者でなければ、データの品質を評価する／保証するのは、困難であるとも考える）。

(3) 「データ駆動型」の真の意義は、一義的には、(データ駆動型人工知能が、そうであったように)それまで人間が依拠してきた「分類」ではない「分類」等に基づきビジネスが構築されること、と考える。

以上の観点から、筆者の観点からは、「データ駆動型ビジネスモデル」とは、決して、「データ」を商材として流通させて利益を得る「ビジネスモデル」そのものではないと考える。

『それまで人間が依拠してきた「分類」ではない』ものであるこそ、それまでの延長ではない、新たな「ビジネス」が生じる可能性が生じるものとする。その意味では、上述したような「がん保険」の新たな枠組みは、このような方向性の第1歩とも考えられる⁽⁴⁶⁾。もっとも、「データ」は、特定の企業のみで意義を有する「パイプライン型ビジネスモデル」というより、多くのプロデューサー、ユーザにとって、その価値を共有しうるものであり、「プラットフォーム型ビジネスモデル」に親和性が高いといえるであろう。

そうすると、むしろ、以下のような状況が想定されるのではないだろうか？

第1段階 一定程度、「流通」に至らなくとも)自由な「データ取引」が実現され、これにより、各ビジネス主体が、自己のビジネスに適した「学習用データ」またはこれに基づく(データ駆動型人工知能の)「学習済みモデル」を柔軟に手に入れられるようになることを前提として、

第2段階 上記のようにして得られた「学習用データ」またはこれに基づく「学習済みモデル」と、自身のビジネスのユーザから得られる「データ」をも利用して、それまででない「分類」「予測」を、独自に用いることで、これまででない価値を生む場を提供できる「プラットフォーム型ビジネス」が構築されるという段階を経るのではないかと予想される。

この意味では、「第1段階」の前提が、まだ、不十分である以上、「第2段階」の意味でのビジネスモデルは、現時点では、かなり限定的な意味で、実現されているに過ぎない、というのが現状であろう。

3-3. 「データ駆動型ビジネスモデル」における知的財産の役割

「データ駆動型ビジネスモデル」における知的財産の役割については、以下のようなものが想定される。

まず、「特許」については、その技術を独占するための手段ということが一義的であるため、データに対して、特許権のような排他的な権利を設定することは、データの流通を阻害する、との見解もあり得る⁽⁴⁷⁾。

しかし、すでにプラットフォーム型ビジネスのところでも説明したように、「特許権ひいては知的財産権」は、「自身の領域にビジネスを囲い込むためのもの」というだけの意義づけではなくっており、「オープン&クローズ戦略」において、オープン領域を設けるための手段として使用される場合もあるといえる。

(46) このような変革は、別に、ビジネスに限られるものではないと考える。米国では、受刑者の仮釈放の可否を人工知能が判断する事例もあるとのことである。

(47) M1 グランプリ 2019 では、漫才コンビの「ぺこば」が、全く新しい形の「ポケ・ツッコミ」を披露し、おしくも3位とはなったが、筆者は、その新しさに驚きつつも、大爆笑してしまった(審査員の松本人志氏によれば、「ノリ・ツッコまないポケ」)。一方で、そのネタの中で、「知識は水だ。独占しては、いけない。」と“哲学的”なコメントがあったことに、はっとした。もっとも、ビジネスの世界では、「独占すべき知識」があるのは前提であろうし、そもそも、「ビジネスに関する全知識を公開せよ」というのは、誰も納得しないであろう。一方で、「技術」の世界では、特許制度は、技術を一定期間独占させる代わりに、「技術に関する知識を独占しない(公開する)」ことを第三者に対するメリットとして提供する制度であるのだから、これを「データ」について、どう生かすのかが、特許制度側からデータの保護や流通を見る際には、中心的な課題と考える。一方で、企業が「データの独占」を意図するのであれば、どのような項目・属性を対象としてデータを収集するのを含めて、特許として公開することが妥当であるのかは、知財戦略・経営戦略上の判断が必要となる。

そして、「データ」が「取引の対象」または「流通の対象」となるためには、むしろ、そのために、知的財産権を、いかに利用するか、という観点からの検討も必要と考える。その際には、「コスト」がかかる対象を流通にのせるためには、合理的対価の支払いが保証される必要があることも検討すべきである。

(1) 「データ類型1：社会・インフラデータ」の保護・流通を促進するための知的財産制度について

不正競争防止法の改正により、「限定提供データ」との枠組みができたことにより、そのようなデータの収集に協力した複数社の協力の結果たる「データ」が第三者に不正に利用されることには、一定の歯止めができたといえる。

ただし、今後、このようなデータ収集について、データの拠出を行う側へのインセンティブおよび権利の帰属の調整の観点からは、「データ」に関する標準化、「データの標準化の前提となる学習手法についての一定程度の標準化（たとえば、データ項目の選定や、計測条件の標準化など）」は、必須と予想される。

この意味では、後述するような「データ構造の特許権」「学習用データの生成方法の特許権」「学習方法の特許権」については、この「データ類型1」のデータが、取引対象となるための「データプール」とでも呼ぶ制度の実現のためには、必要となる可能性がある。言い換えると、データを提供した者に、一定のライセンス料が入るなどのインセンティブの設計が必要と考える。

(2) 「データ類型2：個人等関連データ」の保護・流通を促進するための特許制度について

これについては、基本的には、「秘密」にされることに、「情報の価値」が存在する、または、「秘密に保持することがプライバシーの観点から要求される」という点では、特許権の果たす役割は、限定的である可能性はある。逆に言えば、たとえば、「情報銀行」や、PDSなどは、一定の貢献が見込まれる。したがって、このデータ類型については、特許権などが関わる局面は、むしろ、個人情報の取引・流通と、個人の同意・認知などに関する技術そのものの方が、観念しやすい。

もちろん、一方で、「匿名加工情報」となった後の情報処理について、特許権等の考慮が必要な状況もありうるもの、その場合は、「データ類型3：技術標準データ」の扱いに準じるべきと考える。

(3) 「データ類型3：技術標準データ」の保護・流通を促進するための特許制度について

- i) 「生データ」を保有する（個人ではない）一次利用者が、データを「データ収集者」に提供するインセンティブが必要と考える。

現在のところ、「データ権」というような「データ保有者」がデータの提供に対して、一定の対価を受けられる根拠となる権利が存在しない以上、これは、「データ提供者」と「データ受領者」との間での「契約（債権債務関係）」によるもの、とされている。ただし、「データ」に対して、それが自由に流通するような「市場」というものが、現時点では想定できず、結果として、「データの対価」に、「市場原理」が入り込む余地がほぼないこともあり、「データの適正価格」ということの合意形成も容易とは言えない。

古くから財産権としての位置づけを有し、財産権として流通が行われてきた「特許」においてさえ、適正な「市場価格」を想定することが困難である以上、「データ」に対しては、なおさら、困難が想定される。

そのような状況にあって、複数の技術を提供する者が、対価を得る仕組みとしては、特許の場合は、いわゆる「パテントプール」という仕組みがある。これまでの「技術」の発展過程を振り返ってみると、「新しい技術の創出と特許権の取得」→「当初開発を担当した主体以外主体による関連技術の創出と特許権の取得」→「関連主体からなる業界において、特許権をベースとするパテントプールの構築とライセンスアウト」との流れとなったものと考えられる。

そもそも、「パテントプール」が、一定の技術分野で、有効に機能しているのは、a) 1社の技術だけでは、対象となる製品を製造・販売することができないことに、複数の技術提供者側で事前の合意が存在すること、b) 提供する技術に関して、誰の技術であるか、が「特許」という手段により証明されていること、c) 「パテントプール」に収められるライセンス料が製品価格に比して合理的範囲であること、d) 「パテントプール」に収められたライセンス料の分配が、各技術提供者の「特許」に基づいて設定されていること、などの事情

が存在するものと考えられる。

翻って、「データ」に対しても、上述した通り、「1社の保有するデータだけでは、対象となる人工知能の学習のためには不十分である」との状況は生じうると考えられるものの、たとえば、「データプール」という仕組みを想定することが現状では、容易ではないように考えられる。

ここで、上述した「データ類型3：技術標準データ」のようなデータについては、このような「データプール」は、有効な仕組みの候補となり得ると考えられる。そして、現在のところ、データ提供者が対価を受けられる可能性としては、「データ構造に対する特許権」「学習用データの生成方法の特許権」については、データ提供時の一時金だけでなく、継続的な収入をもたらす可能性があり、インセンティブをもたらすための1つの候補となり得ると考える。

ここで、特に、「特定の技術分野において、成果物の性能が所定の基準を満たすかを判断できる標準を提供できるようなデータセット」については、それを制定し、順次、改訂していきける状況を作り出せた企業または企業群が圧倒的に有利な立場となれることが予想される。なぜなら、「データ駆動型人工知能」では、「処理の中核となる部分の判断過程」が（XAIなどの取り組みはあるものの）、ブラックボックスとなることは避けられず、結果として、品質の保証は、最終的には、特定の入力データセットに対する出力値の妥当性ということで、保証することが1つの手段となると想定されるからである。

ii) 公的資金により収集された（技術的な意義を有する）データについて

「データ類型3：技術標準データ」のようなデータについては、私企業が、準備し管理するような性質のものではなく、本来、公的な資金により実施される研究により集積され、公的資金により維持管理されるべき、との見解も存在しうる。ここで、公的な研究機関が収集したデータや、公的な資金で実施された研究（民間企業の参画もあり得る）で収集されたデータについて、「オープンデータ」とすることも現実に、実施されている。

ただし、現在のところ、「公的な資金により実施される研究によりデータが集積される」ところまでは実現されている事例が多く存在するものの、データの「維持管理と公開」まで長期に継続的に実現するのは、それなりの困難がある。これは、公的な資金については、そもそも財源としての限りがあり、また、一定期間の経過後は、さらにより新しい他の研究等に振り向けられるべきものであることをも想定すれば、仮に、「公的な資金により実施される研究により集積されたデータ」であっても、いずれは、民間または準公的な機関（業界団体により設立される法人など）による自立的な運営を可能とするために合理的な額の収益の獲得を前提とした「維持管理と公開」に移行しなければ、長期にわたり、適切なアップデートやユーザへの安定な提供を継続することが困難になると予想されるからである。

とすれば、そもそも、民間等に移管された後でも、当該データの利用に対して、利用料を徴収できるような枠組みが必要となる。言い換えれば、「当該データに対してライセンス料」を徴収できるための「根拠」が必要と考えられる。一方で、データについては、ひとたび提供してしまえば、そのデータが、その後、どのように利用されるのかを追跡することが一般には、困難であり、継続的に収益をあげられる枠組みの構築が容易とは言えない。

iii) 「データ構造」「学習用データの生成方法」「学習済みモデル」の特許権

上述したような意味では、「データの収集」の対象となるデータについて、一次利用者または「データの収集／分析者」などが、「データ構造の特許」「学習用データの生成方法の特許権」などを有している（または利用する権利を有している）ことが、必要となる可能性がある。

なお、「データ構造の特許権」のような「絶対的な排他権」ではなく、権利を弱い方からいえば、「報酬請求権」「行為規制のある権利（FARND宣言など）」「相対的排他権である物権的権利（著作権など）」「絶対的排他権である物権的権利（特許権など）」というような段階が想定されるものの、新たに、「データ権」なるものを想定すべきか、仮に想定すべきとしても、どのような権利の構成とするのが望ましいか、について

は、現在進行中の検討課題であり、将来の（しかしそう遠くない時点での）最も重要なテーマであることは、間違いない。少なくとも、筆者としては、「報酬請求権⁽⁴⁸⁾」の位置づけは必要と考える。

ただし、「報酬請求権」の位置づけの検討に当たっては、以下のような事情も考慮する必要がある。

- a) データ駆動型人工知能である以上、実際に「データを使って人工知能を学習させない限り」所望の結果が得られるかどうかは、自明ではない。

現在のところ、モデル契約などでは、主として検討されているのは、「データの帰属する主体」が、「データ分析者（学習済みモデルの作成者）」に解析を有償で委託し、その結果、一定程度の有用性を有する「学習済みモデル」が生成されたときに、その「学習済みモデル」に対する権利の帰属をどうするか、という点が主となっている場合が多いと理解する。

- b) しかしながら、今後、データ駆動型人工知能が、さらに発展していけば（あるいは、発展していくためには）、「データ分析者（学習済みモデルの作成者）」が、「データの帰属する主体」から、有償にてデータの提供を受け、その結果、一定程度の有用性を有する「学習済みモデル」を自身のために生成する、という場合が想定される。とすると、この場合、まず、「データの帰属する主体」にとって、データを提供するインセンティブの大きな1つは、上述した「報酬請求」ということになると予想される。一方で、結果が出る前の経済的・技術的な意味での「リスク」を負担しているのは、「データ分析者」ということになり、当然ながら、「学習済みモデル」の権利の帰属については、a) の場合とは異なる扱いとすべきである。

このような状況への対処の考え方としては、契約における当事者の交渉次第ということも想定はされるものの、「学習済みモデル」の経済的価値が大きければ大きいと期待されるほど、データ提供時の一時金としての「報酬請求」のみで、当事者間の合意が得られる可能性は、むしろ低くなってしまのおそれがある。

1つの解決方法は、「学習済みモデル」の特許権を、「データの帰属する主体」と「データ分析者」と共同で出願することで、データに関する事業としての実施に関する「報酬請求」を、特許権についてのライセンス料の問題に移し替えることが考えられるであろう⁽⁴⁹⁾。

いずれにしても、「データ類型3：技術標準データ」については、このデータについては広くオープンにして使用を促し、たとえば、「品質保証」などの意味で「認証」が必要な時は、そのための方法の知財等を込みで、それを有償とする（クローズとする）など、オープン&クローズの考え方が必要となると考える。

(4) データ駆動型人工知能技術そのものに対する特許権による保護

上述した通り、「データ駆動型人工知能技術」において、「学習処理を実行するためのソフトウェア」「学習済みモデルの実行環境のためのソフトウェア」は、すでに、技術プラットフォームが構築・提供されているのが現状である。しかも、このような「技術プラットフォーム」自体、少数の巨大企業に独占されつつあるというべきであろう。

このような状況において、技術の展開・発展により産業の振興を図るという観点で、注意しなければなら

(48) 「データ提供」の契約実務上は、これまでの知的財産に関する契約の延長で、「データ」をあたかも「物」（したがって、「所有権」の対象）であるかのような誤認や、「著作物」（したがって、「著作権」の対象）であるかのような誤認から、文言の合意に必要以上の時間を要する可能性があるようである。これまでの「知的財産権」の契約実務とはずれていることから、「データ」の法律上の位置づけを明確化しておくのが望ましいように感じている。

(49) 「学習済みモデル」の特許権を想定したとしても、「データの帰属する主体」自身が、「学習済みモデル」に関する事業を実施しない場合には、これは、まさに、大学と企業との間の産学連携の際に問題となりがちな「不実施補償」に類似する問題を招来する可能性がある。そして、産学連携の場合は、企業側（経済的・技術的な意味での「リスク」を負担している側）では、事業化された後も、継続的にライセンス料を支払わなければならないことを避けるために、一時金での解決を求める傾向があることも実情である。あるいは、「学習済みモデル」の特許権の共有持ち分の譲渡を受けるという解決策もあり得るが、この場合でも、その対価の額の決定には、大きな困難が伴う。この意味では、適正なルールがなければ、そもそも、「データ分析者（学習済みモデルの作成者）」が、「データの帰属する主体」から、有償にてデータの提供を受ける、ということ自体が、促進されず、結局のところ、データが取引されるという環境が整わないおそれがあることにも、注意が必要であろう。1つの解決方法は、「学習済みモデル」を生成した分析者が、さらに第三者にライセンスアウトして、「データの帰属する主体」と「データ分析者」が収入を分配するケースなどであろう。

ないことは、(GAF A に対する非難として時として主張されるような)「データ」が一つの企業に集中することではなく、「データを利用できる環境」が、一つの企業に集中することである。

そのような前提で考えるとき、「学習済みモデル」の権利、たとえば、特許権は、特に、「特定のサービス・応用用途に限定的な技術」だけではなく、むしろ「一定程度の汎用性を有する学習済みモデル」の保護が、「データ提供者」と「データ分析者」との間の権利の調整という観点から、「データの取引」を促進するために、後述するように重要であると考ええる。

また、筆者の見解としては、「学習済みモデル」の技術的範囲については、それを生成するための「学習用モデル⁽⁵⁰⁾の構成」および「人工知能の学習方法」に限定して解釈すべきものとする⁽⁵¹⁾。

とすると、たとえ、「学習済みモデル」の特許権といえども、「学習用モデルの構成」「学習方法」自体も発明の一部を構成するものと考えて、十分な開示を有する明細書・クレームの作成がされるべきと考える。

そして、「学習済みモデル」については、人工知能プラットフォームとは、独立した権利として想定されうるとも考えられるので、「人工知能プラットフォーム」が海外のプラットフォーマーに独占される状況になりつつある現状に鑑みると、その意味でも、日本の企業にとっては重要な権利と想定される。

さらに、そもそも、人工知能技術が、今後、技術的な意義を増加させることを想定するのであれば、日本の企業として、「学習用モデルの構成」および「人工知能の学習方法」について、独自の技術開発を行っていくようなモチベーションとして特許権を認める必要も考えられる。

極論すれば、必要なのは、学習済みモデルを作成するための「学習用データを利用できる環境」および「学習済みモデルを使用できる環境」であって、「人工知能学習用ソフトウェア」そのものではない、という割り切りも可能である。

さらに、本稿の最初に言及したように、今後、重要性を増すと考えられる「データ」として、「データを取得する H / W と技術的に密接な関連性を有するデータ」が想定されることを考慮すると、センサーやデータを取得するためのフロントエンド部分については、独自に、人工知能技術を展開し、それ以降の「汎用的人工知能処理」については、「プラットフォーマー」が提供するものを利用する、ということも、重要な選択肢と考える。

4. 「データ駆動型人工知能」を保護するための知的財産とは？

4-1. 「データ駆動型人工知能」に関連する技術の明細書・クレームの実例

上述の通り、データ駆動型人工知能を現実に使用するにあたっては、「データ」の重要性が高いことも相まって、「学習用データ」の保護が、1つの大きな関心の対象となっている。そのために、「学習用データ」について、「データ構造」の特許クレームについて特許権を取得することが検討されたり、あるいは、「学習用データの生成方法」について特許の取得が検討されたりしているのが現状と理解する。

しかしながら、そもそも、「学習用データ」は、第三者からは、直接的には観測できない状態で使用されることが多いと考えられるために、単純に、これを特許の対象とすれば、自身のビジネスが保護できる、という種類のものとは言えない（侵害の摘発が困難である）点が問題となる。

ただし、ここで、「学習用データ」の「保護」といったときには、通常の知的財産権で保護される対象や、営業秘密として保護される対象とは、少々、事情が違う場合があることが想定される。つまり、ここでいう保護とは、「学習用データ」が“囲い込まれる”ことを保証することではなく、適切な保護を前提として、これが“商業的な意味での取引の対象”あるいは“流通の対象”となり得ることが必要なことである。

(50) 学習に使用するモデルの数学的な構造であって、学習処理でパラメータを決定する前のもの。

(51) 酒井 将行 (著)「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 Patent 第 20 号 (2018 年)、日本弁理士会中央知的財産権所、研究報告第 45 号「特許クレーム解釈と記載要件」、pp.223-257

そこで、たとえば、「学習用データ」について、「データ構造」「学習用データの生成方法」の特許を取得するのを目指す場合は、典型的には、上述した「データ類型3：技術標準データ」のようなデータを対象とする場合と考えられる。

(1) データ構造のクレーム

特許庁は、審査基準に対する事例の追加として、「IoT 関連技術等に関する事例について」との内容を平成 29 年 3 月に公開⁽⁵²⁾している。特に人工知能技術に関連しては以下のものがある。

i) 「データ構造」の発明該当性（特許適格性）に関する具体例

「学習用データ」に直接関わるものではないと考えられるが、「[事例 2 - 13] 音声対話システムの対話シナリオのデータ構造」として、以下の例が挙げられている。

【請求項 1】

クライアント装置とサーバからなる音声対話システムで用いられる対話シナリオのデータ構造であって、

対話シナリオを構成する対話ユニットを識別するユニット ID と、

ユーザへの発話内容及び提示情報を含むメッセージと、

ユーザからの応答に対応する複数の応答候補と、

複数の通信モード情報と、

前記応答候補及び通信モード情報に対応付けられている複数の分岐情報であって、前記応答候補に応じたメッセージ及び前記通信モード情報に応じたデータサイズを有する次の対話ユニットを示す複数の分岐情報と、を含み、

前記クライアント装置が、

(1) 現在の対話ユニットに含まれるメッセージを出力し、

(2) 前記メッセージに対するユーザからの応答を取得し、

(3) 前記ユーザからの応答に基づいて前記応答候補を特定するとともに、前記クライアント装置に設定されている前記通信モード情報を特定し、

(4) 当該特定された応答候補及び通信モード情報に基づいて 1 つの分岐情報を選択し、

(5) 当該選択された分岐情報が示す次の対話ユニットをサーバから受信する処理に用いられる、対話シナリオのデータ構造。

そもそも、特許法上は、「データ構造」が特許権の対象となる根拠は、特許法第二条には、以下の定義規定がある。

「第二条 4 この法律で「プログラム等」とは、プログラム（電子計算機に対する指令であつて、一の結果を得ることができるように組み合わせられたものをいう。以下この項において同じ。）その他電子計算機による処理の用に供する情報であつてプログラムに準ずるものをいう。」

このうち、「その他電子計算機による処理の用に供する情報であつてプログラムに準ずるもの」と規定があり、かつ、「プログラムに準ずる」とある以上、「データ構造」が特許権の対象となりえることとなるものの、「データ構造」とは何であるかという点では、「プログラム」自身については、「電子計算機に対する指令であつて、一の結果を得ることができるように組み合わせられたものをいう。」と定義されていることを考慮することが必須となる。

とすると、「データ構造」とは、「データとしての構造」を単に記載するのみではなく、一般には、「電子

(52) https://www.jpo.go.jp/system/laws/rule/guideline/patent/handbook_shinsa/document/index/app_z.pdf

計算機による処理の用に供」しているときの態様をも、クレームに記載することが必要と解するのが素直であろう。結果として、上記のクレームの例では、前半部分に「データの構造」に関する記載があり、後半部分では、このような「データ構造」の情報を受け取った側（クライアント装置）が、当該データ構造に基づいて、いかなる情報処理を実行するのか、について記載されていることになる。

ii) 「データ構造」の他の記載の例の検討

上述した「IoT 関連技術等に関する事例について」においては、クレームの記載上、「データ構造」が、a) 「コンピュータ（例では、クライアント装置）に使用されている状態」をクレームしているのか、あるいは、b) 「コンピュータの使用が可能な構造であることを記載しているのか」が必ずしも判然としない、という課題があるように思われる。

このような「データ構造」のデータを製造し販売する行為は、前者であれば、このクレームの「間接侵害」に該当し、後者であれば、このクレームの「直接侵害」に該当することになる。

当然ながら、「学習用データ」の保護を第一義に考えるのであれば、「コンピュータに使用される前」に、データ単体として取引の対象たる状態を保護範囲とする必要があり、後者であることが必要と考えられる。

そこで、現実にはデータが流通する際に、当該データ構造のデータの作成や販売が、直接侵害であるとの主張を積極的に可能とすることを意図して、以下のような記載とする可能性も想定されるであろう。

【請求項 1】

クライアント装置とサーバからなる音声対話システムで用いられる対話シナリオのデータ構造であって、前記クライアント装置が、

- (1) 現在の対話ユニットに含まれるメッセージを出力し、
- (2) 前記メッセージに対するユーザからの応答を取得し、
- (3) 前記ユーザからの応答に基づいて前記応答候補を特定するとともに、前記クライアント装置に設定されている前記通信モード情報を特定し、
- (4) 当該特定された応答候補及び通信モード情報に基づいて 1 つの分岐情報を選択し、
- (5) 当該選択された分岐情報が示す次の対話ユニットをサーバから受信する

処理を実行するために使用されるデータ構造であって、

対話シナリオを構成する対話ユニットを識別するユニット ID と、

ユーザへの発話内容及び提示情報を含むメッセージと、

ユーザからの応答に対応する複数の応答候補と、

複数の通信モード情報と、

前記応答候補及び通信モード情報に対応付けられている複数の分岐情報であって、前記応答候補に応じたメッセージ及び前記通信モード情報に応じたデータサイズを有する次の対話ユニットを示す複数の分岐情報と、を備える、対話シナリオのデータ構造。

このようなクレーム形式を検討するのは、クレームのプリアンブルおよびボディ部分の記載の仕方が、権利の保護範囲に影響を与えるからである。すなわち、「データ」ではなく、現実の「物」の発明の事例であり、かつ、システムに使用される「消耗品」に関する裁判例ではあるものの、プリアンブル部分に、「システムの構成を記載した後に、当該システムで使用される消耗品であること」に言及することで、ボディに記載された構成を有する「消耗品」（一体化製品）が、クレームに対する直接侵害となることを認定したうえで、さらに、この「一体化製品」の一部のみを製造・販売する行為が、当該クレームに対する「間接侵害」として認定された事案⁽⁵³⁾が存在する。

したがって、「データ構造」に対しても、上記のようなクレームが認められ、かつ、ボディの構成要件を

満たすデータの製造販売が直接侵害となるのかは、クレームのボディ部分の構成要件の一部しか含まない「データ構造」のデータを製造または販売する行為が、「間接侵害」となり得るのか、という点で、論点となり得ると考えられる。

特許法が、「データ構造」を「物」とみなし、それ自身を特許請求の範囲にクレームとして記載することを認めているとされることからすれば、上記のようなクレームであれば、ボディの構成要件を満たすデータの製造販売が直接侵害となる、と解するのが妥当であり、その結果、上記のようなクレーム形式の方が、データ構造の特許権による保護として、より望ましい、といえる可能性がある。

(2) 学習用データの生成方法のクレーム

学習用データの生成方法に関して、実際に出願された特許のクレームの例として、たとえば、以下のようなものがある。

すなわち、上述した「データ構造のクレーム」は、「学習用データ」を対象とするのであれば、「学習用データが、いかにして学習に使用されるか」という観点から、「学習用データ」を特定しようとするものであるのに対して、「学習用データの生成方法のクレーム」は、「学習用データが、いかにして生成されるか」という観点から、「学習用データ」を特定しようとするものであるといえる。

学習用データの生成方法のクレームの例（特開 2019 - 114168 号公開公報）

(53) 「薬剤分包用ロールペーパー」事件（大阪地判 平成 30 年 12 月 18 日 平成 28 年（ワ）第 6494 号特許権侵害差止等請求事件、知財高判 令和 1 年 6 月 27 日 平成 31 年（ネ）第 10009 号特許権侵害差止等請求控訴事件）、本件で、問題となったのは、以下の特許第 4194737 号の請求項である。

A 非回転に支持された支持軸 (1) の周りに回転自在に中空軸 (1C) を設け、中空軸 (1C) にはモータブレーキ (20) を係合させ、中空軸 (1C) に着脱自在に装着されるロールペーパー (R) のシートを送りローラ (2, 3) で送り出す給紙部と、2つ折りされたシートの中にホッパ (5) から薬剤を投入し、薬剤を投入されたシートを所定間隔で幅方向と両側縁部とを帯状にヒートシールする加熱ローラ (6) を有する分包部とを備え、ロールペーパー (R) の回転角度を検出するために支持軸 (1) に角度センサ (25) を設け、上記中空軸 (1C) と上記支持軸 (1) の固定支持板 (11) 間で上記中空軸 (1C) のずれを検出するずれ検出センサ (26) を設け、分包部へのシート送り経路上でシート送り長さを測定する測長センサ (32) を設け、ロールペーパー (R) を上記中空軸 (1C) に着脱自在に固定してその固定時に両者を一体に回転させる手段をロールペーパー (R) と中空軸 (1C) が接する端に設け、角度センサ (25) 及び測長センサ (32) の信号に基づいてシート張力をロールペーパー径に応じて調整しながら薬剤を分包するようにし、さらに角度センサ (25) の信号とずれ検出センサ (26) の信号との不一致により上記中空軸 (1C) に着脱自在に装着されたロールペーパー (R) と上記中空軸 (1C) とのずれを検出するようにした薬剤分包装置に用いられ、

B 中空芯管 (P) とその上に薬剤分包用シートをロール状に巻いたロールペーパー (R) とから成り、

C ロールペーパー (R) のシートの巻量に応じたシート張力を中空軸 (1C) に付与するために、支持軸 (1) に設けた角度センサ (25) による回転角度の検出信号と測長センサ (32) の検出信号とからシートの巻量が算出可能であって、その角度センサ (25) による検出が可能な位置に磁石 (16) を配置し、

D その磁石 (16) をロールペーパー (R) と共に回転するように配設して成る

E 薬剤分包用ロールペーパー。

特に、構成要件 A は、クレームのプリアンブルとして、「薬剤分包装置」の構成を特定するとともに、「薬剤分包装置に用いられ、」との文言により、クレームのボディには、「薬剤分包用ロールペーパー」の構成要件 B～E が記載されている。被告らが扱っている被告製品は、プラスチック製の筒部にグラシン紙もしくはセロポリ紙からなる薬剤分包用シートを巻き回したものであった。このため、被告製品は、

i) ユーザがこの被告製品を購入し、

ii) この筒部の軸芯中空部分に、原告製の薬剤分包用ロールペーパーの使用済み中空芯管に輪ゴムを巻き、その状態で挿入することにより、

iii) 両者を一体化した「一体化製品」として薬剤分包装置に用いることが、初めて、できるようになる、との構成を有するものであった。

つまり被告らは発明の対象である「薬剤分包用ロールペーパー」の完成形を製造・販売等しているわけではなく、ユーザが一体化製品としているので、本件の被告らの行為について直接侵害は成立せず、間接侵害の成否が問題となった。判決では、「一体化製品」が本件発明の構成要件を全て充足していると認定し、被告製品の製造販売は、間接侵害に相当すると認定した。

【請求項 1】

時系列に並ぶ複数の時区間を含む時間軸と、プラントの運転状態に関する複数の状態項目を含む項目軸との二次元空間において、前記時区間及び前記状態項目の組み合わせにより特定されるセルごとのデータを含むマップデータを生成するマップデータ生成部と、

前記マップデータに対応する前記プラントの状態を示す教師データを生成する教師データ生成部と、前記マップデータと、当該マップデータに対応する前記教師データとを組み合わせ学習用データを蓄積する学習用データ管理部と、

前記学習用データ管理部に蓄積された前記学習用データに基づく機械学習により、前記プラントの状態の推定用モデルを構築するモデル構築部と、

前記マップデータを前記推定用モデルに入力して前記プラントの状態を示す推定データを導出する推定部と、を備えるプラント管理システム。

…

【請求項 16】

時系列に並ぶ複数の時区間を含む時間軸と、プラントの運転状態に関する複数の状態項目を含む項目軸との二次元空間において、前記時区間及び前記状態項目の組み合わせにより特定されるセルごとのデータを含むマップデータを、機械学習の入力側データとして生成することと、

当該マップデータに対応する前記プラントの状態に関する教師データを、機械学習の出力側データとして生成することと、を含む、学習用データの生成方法。

「学習用データの生成方法」のクレームについては、これが「物の生産方法のクレーム」に該当するかが1つの論点となる。「データ構造」と明示的に記載することなく、単に、「学習用データ」と呼ぶのみであり、かつ、「生成方法」との発明の名称であるときには、単なる「データ」は、特許法上では、「物」とは言えない、との結論となる可能性があることに留意が必要であろう。

明細書の開示内容として、「データの構造」を明記するとともに、そのような「学習用データ」が学習処理に使用される態様を、学習処理の中で説明し、「学習用データの生成方法」が、生産方法に相当することに言及が必要であろう。ただし、実際の権利行使の際に、論点となり得るリスクは残る。

ただし、このような「学習用データの生成方法」のクレームは、このような「学習用データの生成方法」自身を含めて「技術標準」とすることを意図する場合などでは、有効なクレームということになる。

(3) 「学習済みモデル」に対するクレーム

ここでは、「学習済みモデル」のクレームを、「データの利用を促進する」との観点から考察することとする。「学習済みモデル」には、筆者の想定としては、大きくは、以下のような2つの類型が存在する。

学習済みモデル類型1) 「学習済みモデル」の構成自体または「学習用モデル」に対する「学習方法」に技術的に新規な特徴があり、このような技術的特徴を請求項に記載することで特許となったもの。

学習済みモデル類型2) 「学習済みモデル」の構成自体または「学習用モデル」に対する「学習方法」自体は、必ずしも進歩性があるとまではいえないものの、そのような「学習済みモデル」を用いた応用用途（アプリケーション）に特徴があり、このような応用用途に関する特徴を請求項に記載することで特許となったもの。

たとえば、「学習済みモデル類型2」に対応するクレームとしては、上述した「IoT 関連技術等に関する事例について」には、「〔事例2-14〕 宿泊施設の評判を分析するための学習済みモデル」として、以下のサンプルクレームが例として挙げられている。

【請求項 1】

宿泊施設の評判に関するテキストデータに基づいて、宿泊施設の評判を定量化した値を出力するよう、コンピュータを機能させるための学習済みモデルであって、

第 1 のニューラルネットワークと、前記第 1 のニューラルネットワークからの出力が入力されるように結合された第 2 のニューラルネットワークとから構成され、

前記第 1 のニューラルネットワークが、少なくとも 1 つの中間層のニューロン数が入力層のニューロン数よりも小さく且つ入力層と出力層のニューロン数が互いに同一であり各入力層への入力値と各入力層に対応する各出力層からの出力値とが等しくなるように重み付け係数が学習された特徴抽出用ニューラルネットワークのうちの入力層から中間層までで構成されたものであり、

前記第 2 のニューラルネットワークの重み付け係数が、前記第 1 のニューラルネットワークの重み付け係数を変更することなく、学習されたものであり、

前記第 1 のニューラルネットワークの入力層に入力された、宿泊施設の評判に関するテキストデータから得られる特定の単語の出現頻度に対し、前記第 1 及び第 2 のニューラルネットワークにおける前記学習済みの重み付け係数に基づく演算を行い、前記第 2 のニューラルネットワークの出力層から宿泊施設の評判を定量化した値を出力するよう、コンピュータを機能させるための学習済みモデル。

このクレームを、「学習済みモデル類型 1」に対応するクレームに書き換えるとすると、たとえば、以下のようになるであろう。

【請求項 1'】

顧客に提供される商品について顧客自身が記載した内容のテキストデータに基づいて、前記商品についての評価を定量化した値を出力するよう、コンピュータを機能させるための学習済みモデルであって、

第 1 のニューラルネットワークと、前記第 1 のニューラルネットワークからの出力が入力されるように結合された第 2 のニューラルネットワークとから構成され、

前記第 1 のニューラルネットワークが、少なくとも 1 つの中間層のニューロン数が入力層のニューロン数よりも小さく且つ入力層と出力層のニューロン数が互いに同一であり各入力層への入力値と各入力層に対応する各出力層からの出力値とが等しくなるように重み付け係数が学習された特徴抽出用ニューラルネットワークのうちの入力層から中間層までで構成されたものであり、

前記第 2 のニューラルネットワークの重み付け係数が、前記第 1 のニューラルネットワークの重み付け係数を変更することなく、学習されたものであり、

前記第 1 のニューラルネットワークの入力層に入力された、前記テキストデータから得られる特定の単語の出現頻度に対し、前記第 1 及び第 2 のニューラルネットワークにおける前記学習済みの重み付け係数に基づく演算を行い、前記第 2 のニューラルネットワークの出力層から前記評価を定量化した値を出力するよう、コンピュータを機能させるための学習済みモデル。

ここでは、たとえば、明細書中で、「商品」とは、有形で物理的な「製品」だけでなく、無形で活動を通じて提供される「サービス」も含む」というような定義がされているものとする。また、上記のような請求項でも、学習済みモデルの構成と学習方法の記載（上記の例では、いわゆる、「自己符号化器（オートエンコーダ）」を使用していることの記載）とで、新規性・進歩性の要件は満たされているものとする⁽⁵⁴⁾。

ここで、請求項 1 のような、下位概念の「学習済みモデル」についての特許が必要となる局面としては、たとえば、上述したような「学習済みモデル」の特許権を、「データの帰属する主体」と「データ分析者」

と共同で出願する』場合が、想定される。

一方で、請求項1'のような上位概念のクレームは、データ分析者側のインセンティブとなる可能性がある。つまり、データ分析者の立場からすれば、特定の具体的なサービスに限定されることなく、自身が「学習済みモデル」に対する特許権を取得していれば、「特定の具体的なサービス」に関する「データの帰属する主体」との間に、対等の関係で、データ提供に関する契約を締結または取引をすることが可能となる可能性が高まる。

上述した通り、少なくとも、第一段階では、多くの業務としての経験を有する「データ分析者」が生まれなければ、「データ駆動型ビジネス」自体の発展は、そもそも、おぼつかないと考えられるので、このような状況は、「データ分析者」が人工知能技術の開発を行う上でのインセンティブを知的財産の観点から考慮する必要があるとも考える。

一方で、「特定の具体的なサービス」に関する「データの帰属する主体」の側からすれば、少なくとも、請求項1'のような自身のビジネスに直接関わる学習済みモデル（請求項1'の特許の利用発明の関係になると考える）についての特許を取得できれば、あるいは、請求項1'の特許についての第三者の実施に関して、一定の制限を設けることを、「データ分析者」との間に合意することで、自身のビジネスを守ることは可能となる。

ただし、「鶏と卵」の関係ではあるものの、「データ分析者」にとって、自身が独自に、請求項1'のような学習済みモデルに対する研究開発を実施するために、独自ルートで、少なくとも類似する分野のデータを取得できるようにしておくことも必要であろう。この点で、むしろ、「学習用データ」が準備され、公開されている環境を作り出す必要がある。日本として注力したい技術分野については、状況によっては、「学習用データ」の準備までは、公的な資金により実施することも必要となろう。

4-2. 「データ駆動型人工知能」についての特許庁ハンドブックの記載と検討課題

平成31年1月30日において、特許庁審査第一部調整課審査基準室より「AI関連技術に関する事例の追加について」との資料が発行されている⁽⁵⁵⁾。

以上、説明してきたとおり、「学習済みモデル」(=「学習用モデルの構成」+「人工知能の学習方法」)、「データ構造」「学習用データの生成方法」などが、特許権としての検討が必要なものとして説明したが、学習済みモデルのところでも言及したように、そもそも、「人工知能の学習方法」自体も、特許要件を考慮するときには、検討対象となる。

そこで、以下では、「AI関連技術に関する事例の追加について」で記載のある事例を基に、特許要件として考慮すべき点について、さらに検討する。

(54) もっとも、米国では、「特許適格性」の要件を満たすための1つの要件として、“practical application”の記載を求められることがあるので、最低限、この程度の限定は必要となろう。もっとも、その場合に、「学習済みモデル」そのものをクレームの対象とすることはできないので、米国では、米国実務に対応した検討が別途必要になる。

(55) https://www.jpo.go.jp/system/laws/rule/guideline/patent/document/ai_jirei/jirei_tsuika.pdf
https://www.jpo.go.jp/system/laws/rule/guideline/patent/document/ai_jirei/jirei.pdf
 筆者としては、日本の特許庁が、技術上また法律上も、まだ必ずしも評価が定まっているとは言えない状況であっても、積極的に「ガイドライン」「参考事例」を公表されていることに、敬意を表するものである。このような公表資料があるからこそ、そこに議論が生まれるからである。そもそも、特許制度を主管する特許庁が、その方向性に対する指針を提供したとして、それに対して、自身のビジネスに対して、将来、生じるであろう課題や紛争に備えて、どのような知的財産権の取得を考えるのかという戦略の立案は、出願人の責任においてなすべきことであるし、それを適切にサポートすることを意図して行動するのが、弁理士の仕事であろう。したがって、以下の議論は、あくまで、「データ駆動型人工知能」の明細書・クレームをいかに記載すべきか?という課題に対する「建設的な問題提起と議論」を意図したものであり、かつ、筆者の勤務先の見解を代表しているものでもないことをお断りしておく。

4-3. 「データ駆動型人工知能」の明細書・クレームをいかに記載すべきか？

データ駆動型人工知能については、以下のような検討項目があると考えられる。

- (1) 「実施可能」「サポートあり」を、特に、「機械」の実行する処理」として容易に判断できない。
- (2) どのような入力データ（特徴）に基づいて「モデル」をつくるのか、どのようにつくるのかで、大幅に効果が異なる。ところが、そもそも、「対象の全てのデータ」を計測・取得することはできない以上、必然的に、選択は常に行われるものであり、「(入力データの) 選択」に基づいたとき、進歩性をどう考えるかが問題となる。

(1) 実施可能要件・サポート要件の充足のために

まず、これまで説明してきたような「データ駆動型人工知能」において、「実施可能要件、サポート要件」を満たすために、明細書・クレームにどのような要件が要求されるのかを考察する。

「AI 関連技術に関する事例の追加について」の中で、本来は、「進歩性あり」との事例として提示されたものを用いて、「実施可能要件、サポート要件」について検討してみたい。もちろん、特許庁自身、あくまで「進歩性」の事例であるので、「実施可能要件、サポート要件」の充足については、対象外との断りがある中での例であるので、本来の趣旨とは、ずれているものの、「実施可能要件、サポート要件」を考える上では、むしろ、典型的な状況が生まれていると考えるので、あえて、検討対象とさせていただきます。

すなわち、筆者としては、「データ駆動型人工知能技術」について、特許明細書化を行う際には、『学習に使用している「データ」からして、請求項に記載した発明について、実施可能要件・サポート要件を満足しているかについて、一応の合理的な疑いが存在する場合』が存在すると考える。

それは、以下のような事情による：

- a) データ駆動型人工知能では、判断処理を実施する部分が「ブラックボックス」である。
- b) ある「データ」に基づいて学習した「学習済みモデル」が、クレームされたような一般的な場合にまで適用可能であるかが、具体例によっては、サポートされていないという場合が頻出する可能性がある。
- c) 一例としては、物理現象ではなく、人間の心理的な状態が関与するような
 - ・特定の環境で、特定の人間の間での「インターアクション」についての観測情報に基づいて学習した「学習済みモデル」が、
 - ・他の環境で、他の人間の間での「インターアクション」に対する予測・分類にまで、一般化できるかが不明である場合 などが想定される。

ここで、「進歩性：事例 36 認知症レベル推定装置」において、「請求項 1：学習に用いる教師データに対する前処理により進歩性が肯定されるもの」との事例が挙げられている。

【請求項 1】

回答者と質問者の会話に係る音声情報を取得する音声情報取得手段と、
前記音声情報の音声分析を行って、前記質問者の発話区間と、前記回答者の発話区間とを特定する音声分析手段と、

前記質問者の発話区間及び前記回答者の発話区間の音声情報を音声認識によりそれぞれテキスト化して文字列を出力する音声認識手段と、

前記質問者の発話区間の音声認識結果から、質問者の質問種別を特定する質問内容特定手段と学習済みのニューラルネットワークに対して、前記質問者の質問種別と、該質問種別に対応する前記回答者の発話区間の文字列とを関連付けて入力し、前記回答者の認知症レベルを計算する認知症レベル計算手段と、を備え、

前記ニューラルネットワークは、前記回答者の発話区間の文字列が対応する前記質問者の質問種別に関連付けて入力された際に、推定認知症レベルを出力するように、教師データを用いた機械学習処理が施された、認知症レベル推定装置。

「進歩性」の判断としては、「前記質問者の発話区間の音声認識結果から、質問者の質問種別を特定する質問内容特定手段と学習済みのニューラルネットワークに対して、前記質問者の質問種別と、該質問種別に対応する前記回答者の発話区間の文字列とを関連付けて入力」という学習処理における、いわゆる「前処理」の部分が、従来技術にはないことをもって、以下の通り、進歩性が肯定されるとされている。

“(相違点)

請求項 1 に係る発明は、…前記質問者の発話区間と、前記回答者の発話区間とを特定すると共に、特定された前記質問者の発話区間及び前記回答者の発話区間の音声情報を音声認識によりテキスト化して文字列を得、…ニューラルネットワークは、前記質問者の質問種別と、該質問種別に対応する前記回答者の発話区間の文字列とを関連付けて入力し、認知症レベルを出力するように機械学習処理が施されるのに対して、

引用発明 1 のニューラルネットワークでは、質問者及び回答者の発話区間の区別なく、音声認識によりテキスト化された文字列をそのまま入力し、認知症レベルを出力するように機械学習処理が施される点。

…

上記相違点 1 について検討する。

…認知症レベルの評価手法として回答者と質問者の会話に係る音声情報のテキスト化された文字列に対して、質問者の質問種別を特定し、当該質問種別に対応する回答者の回答内容とを関連付けて評価に用いるという具体的な手法を開示する先行技術は発見されておらず、そのような評価手法は、出願時の技術常識でもない。

したがって、引用発明 1 のニューラルネットワークに…学習させるに当たり、質問者の質問種別を特定し、当該質問種別に対応する回答者の回答内容とを関連付けて教師データとして用い学習をさせることは、当業者が容易に想到し得ないことである。…

さらに、請求項 1 に係る発明では、質問者の質問種別を特定し、当該質問種別の質問に対応する回答者の回答（文字列）を関連付けることによって、ニューラルネットワークは、教師データから熟練した専門医の知見を効果的に学習することができるので、精度の高い認知症レベルの推定を実現することができるという、顕著な効果が得られる。”

すなわち、クレーム上の構成として従来技術にない「前処理」を採用することにより、「精度の高い認知症レベルの推定を実現することができるという、顕著な効果が得られる」ことが進歩性の根拠とされていることになる。

ただし、上記のような構成の記載のみで、「実施可能要件」「サポート要件」の充足がいえるかについては、筆者としては、以下のような点に問題があると考ええる。

上述した事例 36 では、1) 請求項には、「回答者」と「質問者」に何の限定もなく、かつ、ニューラルネットの学習に使用される「質問種別」「文字列」「教師データ」にも、何の限定もない。そこで、厳密に言えば、

・病院 H1 において、医師 D11（または臨床心理士 C11）により学習用データが取得されたとして、他の医師 D12（または臨床心理士 C12）が質問者となっても同様の推定が可能なのか？

・病院 H2 の医師 D21（または臨床心理士 C21）が質問者となっても、同様の推定が可能なのか？

が、データにより示されないと、クレームに対するサポートがあるとは、いえない可能性がある⁽⁵⁶⁾。

また、後述するように、「AI 関連技術に関する事例の追加について」の中では、(1) 学習用データに含ま

(56) ある環境において成立する「相関関係」が、別の環境では、成立せず、前者の環境で作成した学習済みモデルが後者の環境では、全く機能しないということは、通常、むしろ、よくある事象である。

特に、たとえば、事象 A と事象 B との間の「相関関係」に、いわゆる「交絡」がある場合は、このような状況となることは、統計学の世界でも、認識されている事実である。

たとえば、清水 昌平（著）「統計的因果探索」（機械学習プロフェッショナルシリーズ）、講談社（2017/5/25）を参照。

れる情報（入力データ・特徴量の項目）の選び方が、当業者にとって想定しがたいものであれば、それが進歩性の主張の根拠となる、との言及がある。

一方で、サポート要件・実施可能要件を満たすためには、入力と出力の間に「そのような相関関係があることが技術常識等から（論理的に）推認されうる」ことが、一つの要件とされている。

ただし、そもそも、論理的に推認できるなら、そのような相関に対して、人工知能技術を用いることで、「予測」「分類」等ができることでは、「進歩性」は主張できない。

つまり、明細書作成者にとって、本当に重要なのは、「新たに発見したモデルへの入力データの項目とモデルからの（予測・分類結果の）出力との間に相関があることが技術常識からは推認されない」場合に、そこから「予測」「分類」等ができるモデルが生成できたときに、

- ・「実施可能要件」「サポート要件」を充足することを、どう考えるか、ということと、
- ・「新しい入力データの項目を選んだこと」が、進歩性の主張の根拠に本当にできるのか、ということとを両立させること、である。

筆者としては、これらは、いずれも、厳密に言えば、これも「データ」（実験）により証明される必要があると考える。また、応用用途によっては、学習用データの相関関係についてだけではなく、独立した検証用データセット（テストデータ）に対する予測精度等の開示も必要と考える。

したがって、「実施可能要件・サポート要件」の充足のためには、明細書中で、少なくとも、以下のような記載がされていることが必要と考える。

a) 学習用データ（訓練データ）の範囲で、「新たな入力データの項目」と「出力」（教師データ）との間に、モデルにより「説明可能な関係」（入力から出力が導かれる関係）があることを、実験データとして示すことが必要である。ただし、上述した「過学習」や「人間の心理の影響」などのことも考慮すると、これのみでは、この新たな知見が「技術的な意味で“発明”」といえるかは不明確である。

b) そこで、対象や分野に応じて、クレームされた範囲に対応し、学習済みモデルが、学習用データに含まれない「テストデータ」（検証用データセット）に対して、一定以上の性能を示すことを実験データで示すことも必要である、とすべきであろう。

(2) 進歩性の要件について

2-1) 学習に用いる教師データの追加に、顕著な効果が認められるため、進歩性が肯定される事例 (事例 34 請求項 2「水力発電量推定システム」)

この事例では、以下のような請求項が例として挙げられている。

【請求項 1】

情報処理装置によりニューラルネットワークを実現するダム水力発電量推定システムであって、入力層と出力層とを備え、前記入力層の入力データを基準時刻より過去の時刻から当該基準時刻までの所定期間の上流域の降水量、上流河川の流量及びダムへの流入量とし、前記出力層の出力データを前記基準時刻より未来の水力発電量とするニューラルネットワークと、前記入力データ及び前記出力データの実績値を教師データとして前記ニューラルネットワークを学習させる機械学習部と、前記機械学習部にて学習させたニューラルネットワークに現在時刻を基準時刻として前記入力データを入力し、現在時刻が基準時刻である出力データに基づいて未来の水力発電量の推定値を求める推定部と、により構成されたことを特徴とする水力発電量推定システム。

【請求項 2】

請求項 1 に係る水力発電量推定システムであって、前記入力層の入力データに、さらに、前記基準時刻より過去の時刻から当該基準時刻までの所定期間の上流域の気温を含むこと、を特徴とする水力発電量推定システム。

このような請求項の事例において、請求項1は、進歩性なしとされるものの、請求項2については、進歩性ありとの判断となるという記載がされている。この点についての解説は以下の通りである。

“(拒絶理由がないことの説明)

請求項2に係る発明と引用発明1とを対比すると、両者は以下の点でも相違する。

請求項2に係る発明は、入力層の入力データに、基準時刻より過去の時刻から当該基準時刻までの所定期間の上流域の気温を含むのに対して、引用発明1ではそのような構成になっていない点。

上記相違点について検討する。

請求項2に係る発明は、水力発電量の推定に上流域の気温を用いているが、水力発電量の推定に上流域の気温を用いることを開示する先行技術は発見されておらず、両者の間に相関関係があることは、出願時の技術常識でもない。

一般に、機械学習においては相関関係が明らかでないデータを入力データに加えるとノイズが生じる可能性があるところ、本願の請求項2に係る発明では、入力データに、基準時刻より過去の時刻から当該基準時刻までの所定期間の上流域の気温を用いることにより、春のシーズンにおいて「雪解け水」による流入量増加に対応した高精度の水力発電量を推定することが可能である。この効果は、引用発明1からは予測困難な、顕著な効果であるといえる。よって、水力発電量の推定における入力データに、基準時刻より過去の時刻から当該基準時刻までの所定期間の上流域の気温を含めるという事項は、引用発明1に周知技術を適用する際に行い得る設計変更ということとはできない。”(下線は筆者)

2-2) 事例34における進歩性の主張について

すなわち、上記のような事例は、発明の「構成」については想到容易であっても、「顕著な効果」の存在をもって、進歩性の存在をも認めようとの立場⁽⁵⁷⁾とも考えられる。そこで、「データ駆動型人工知能」についての「顕著な効果」とは何であるか、を検討することが必要になる。

このような議論をするにあたり、技術上の重要なポイントとして、モデルの構造の枠組み自体は、従前のものと本質的な相違がなくとも、たとえば、ニューラルネットワークについては、入力値に対する出力値を教師信号として内部パラメータを学習させることで、任意の関数を表現できることが知られていること(ニューラルネットワークの「普遍性定理(万能近似定理とも)」)がある。したがって、言い換えれば、「学習用データ」の範囲に限るのであれば、ニューラルネットワークを使用して、いかような「入力データ」についても、一定以上の精度で「出力データ(教師データ)」を出力できるモデルを生成できることは、数学的に証明された定理として認められていることには、留意が必要と考える。

(57)「医薬品」の事例であって、技術分野を全く異にする事案ではあるが、「進歩性判断における顕著な効果の位置付け」に関する最高裁の判決も出ており、人工知能分野においては、これをどのように考えるかは、1つの重要な論点であろう。最三小判 令和元年8月27日(平成30年(行ヒ)第69号) 審決取消訴訟事件

『本最高裁判決は、従来から、進歩性の判断において、構成が容易想到であるとしてもその構成が奏する効果が公知技術(引用発明)が奏する効果と異質であるかまたは同質であるが顕著である場合には、構成の容易想到性とは別に効果の顕著性を評価することで進歩性があるとする独立要件説と、構成の非容易想到性こそが最終的に評価されるべき規範的要件(評価的要件)としての進歩性であるとし、進歩性を肯定する方向での事実(評価根拠事実)とこれを否定する方向での事実(評価障害事実)とに分類し、顕著な効果は進歩性を肯定する方向での事実と位置づけられるとする二次的考慮説とがあるとされてきた』(*2)(下線筆者)ところ、本判決は、「顕著な効果」を『条文上の根拠が不明確な顕著な効果の位置付けについて、いわゆる独立要件説ないしそれに近い考え方を採用したもの』(*1)との評価がある一方で、『この判決は、「化合物の医薬用途に係る発明」であることと、「当該化合物を発明に係る医薬用途に適用することが容易想到であること」を前提としたうえで、発明の効果の顕著性は、当該発明の構成が奏するものとして当業者が予測した範囲を超えているかとの観点から十分に検討されるべきことを確認し、…他の化合物が複数存在することのみをもって、「効果の顕著性」を否定したものを咎めたものであって、その射程は、きわめて狭く、独立要件説と二次的考慮説のいずれが採用されるものであるかについて触れるものではない』(*2)との見解も存在することに注意が必要と考える。

*1) 飯島 歩(著)「進歩性判断における予測できない顕著な効果の位置付けに関するドキシピン誘導体含有局所的眼科用処方物事件最高裁判決について」<https://innoventier.com/archives/2019/09/9159>

*2) 高林 龍(著)「最高裁判決『進歩性判断における顕著な効果の位置づけ』、日本評論社、2019年12月25日発行、pp.24-32

このような「普遍性定理（万能近似定理）」は、仮に、隠れ層が1層しかないニューラルネットワークであっても（その隠れ層内のノードの数さえ多くなれば）成り立つことが知られている⁽⁵⁸⁾。すなわち、いわゆる「ディープラーニング」である必要は必ずしもないことになるが、層数が多い（深い層）の場合は、近似の程度や、入力データとして、より複雑な構成を有する対象に対してもモデル化が可能であり、かつ、同様に一定以上の精度の実現が可能なが知られている^{(59), (60)}。

つまり、一般には、各層の中でのノード数が一定以上であり、層数が多い（ディープな）ニューラルネットワークをモデルとして採用すると、このモデルは、いかなる関数関係（入力データと出力データの関係）であっても、一定程度以上の精度で近似することが可能なモデルであるといえることになる。

全く同様のことが言えるかは自明ではないものの、一般論としては、ニューラルネットワークでない場合でも、パラメータの数が多い（自由度が高い）モデルであれば、複雑な関数関係も表現しうることは、容易に想像がつく。

ただし、その代わりに、「学習用データ」の範囲で、複雑な（自由度の高い）モデルを使って、出力が正解となる精度を上げることに注力すると、一般には、そのモデルは「過学習」や「過適応」と呼ばれる状態となり、「学習用データ」以外のデータ（「テストデータ」とも呼ばれる）に対しては、全く性能を発揮できなくなることがあることが知られている。

この観点から、そもそも、「学習用データ」の範囲で、いくら予測性能が高いモデルができたとしても、それだけでは、実用に耐える「学習済みモデル」といえない、といえる。言い換えれば、このような「学習用データ」の範囲についての性能を超える「効果」が提示されない限り、「進歩性の根拠たる」顕著な効果とはいえないことになる。

このような前提の下に、上記の事例 34 の進歩性の存否について検討してみる。

a) 「基準時刻より過去の時刻から当該基準時刻までの所定期間の上流域の気温」の技術的な意義については、「AI 関連技術に関する事例の追加について」では、「雪解け水」による流入量増加に対応した高精度の水力発電量を推定することが可能との記載となっている。ただし、「上流域の気温」は、単純に、何時間前の気温が何度であれば、水位がどうなる、という関係性ではないとも想定される⁽⁶¹⁾。

i) 請求項 1 の「所定期間」は比較的短期間（入力データと出力データの相関関係が明確）であるのに対して、請求項 2 の「所定期間」は、比較的長期間の必要があると考えられる。

つまり、請求項には、「所定期間の上流域の気温」としか限定事項がないことからして、人工知能側で、特徴量抽出等を実行することが必要とするのが妥当ではないだろうか。そもそも、「積雪期、融雪期、梅雨期、夏季、秋季…」があることからして、単純に「上流域の気温」だけで流量が決まるという関係性は想定しがたい。

ii) 単純なモデルを前提に、「上流域の降水量、上流河川の流量及びダムへの流入量」と「上流域の気温」というような異質な（次元の異なる）ものを、同格で入力とするモデルで、うまくいくのかに、疑問がある。少なくとも、「前処理」や「モデル構造」自体、「学習用データ」「学習方法」に何らかの工夫を導入する必要があるのではないかと考えられる。とすると、請求項で、これらの構成について、何らかの言及が必要ではないかと考える。

2-3) モデルへの入力の項目に新たな項目を追加することによる進歩性の主張について

上述した事例 34 のような事例に関連する見解として、さらに以下のようなものも、ありうる。

(58) G. Cybenko, Approximation by Superpositions of a Sigmoidal Function, Mathematics of control, signal and systems, vol. 2, no. 4, 1989

(59) M. Telgarsky (著), "Benefits of depth in neural network", arXiv:1602.04485, 2016

(60) S. Liang and R. Srikant (著), "Why deep neural networks for function approximation?," arXiv: 1610.04161, 2016

(61) <http://www.hrr.mlit.go.jp/shinano/ohkouzu/bousaic/090430yusetsu.htm>. "信濃川と利根川の月別の水量" が開示される。また、<http://river.ceri.go.jp/contents/archive/uploads/h8-79-88.pdf> には、北海道：鶴川（穂別橋）の融雪期の水位、気温（穂別アメダス）の変化（1996年4月～5月）が開示されている。

- i) 入力と出力のみを特定して、入力から出力を計算する過程を具体的には限定しないクレームの記載とした場合、権利範囲は最も広い。
- ii) 人工知能モデルへの入力・出力として何を選ぶかにポイントがある場合があり、具体的な学習手法には特段特徴がない場合でも、入力・出力をどうするか、何を学習するのかを簡潔に記載すれば、明細書・クレームとしては十分である。

「データ駆動型人工知能」の場合、判断や分類をする「学習済みモデル」の構成については、人間にとっては、「ブラックボックス」である以上、上記のような主張がなされるのも、一理あるようにも思われる。

ただし、「発明のポイントは、人工知能モデルへの入力・出力として何を選ぶかという部分のみ」との主張がされる場合に、進歩性ありとして、独占排他権たる特許権を付与してよいのかについては、以下に説明するように、少々、検討が必要であると考え⁽⁶²⁾。

a) 「特微量選択」を人工知能が実施するときに、「学習時の入力データ（または出力データ）の選択のみを特徴とできるような発明」との主張はあり得るか？

学習処理において、学習用モデルへの入力として何を選択するか？（特微量として何を選択するか？）は、機械学習においても重要な検討事項の一つである。

そして、このような技術は、「特微量選択」などと呼ばれて、すでに、「フィルター法」「Wrapper 法」「Embedded 法」など、いくつもの手法が検討されてきている⁽⁶³⁾。このような処理は、人間が、そのような特微量の意義を認識できる場合も、後述する汎化性能の向上などの意味で、実施される場合もある。

ただし、特微量選択を、ソフトウェアの処理として実行するのであれば、そもそも、明細書には、このような特微量選択のためのソフトウェアの処理の開示が必要であるとともに、クレームには、この「特微量選択のための処理」の記載が必要となるというべきである。

(62) もちろん、筆者は、人工知能研究の専門家でもなく、また、統計学の専門家でもないもので、あくまで、「頭の体操」という観点で、事例を紹介するものである。ただし、筆者としては、「(データ駆動型)人工知能の判断プロセス」が「ブラックボックス」であるということが強調されるあまり、少なくとも、区別して考えるべき、以下の3つの事例の間で、すべてが一律に議論されているように感じている：

- 1) 入力と出力との間に「狭義の相関関係（入力 x と出力 y との間の線形関係）」があることがわかっているときに、機械学習により「人工知能のモデル」が得られる場合の事例（この場合は、特微量の選択・組み合わせ方や重みづけの仕方をどうするか、という点で、ブラックボックスといえる場合があるかもしれないが、人間には、判断のプロセスは、本質的に、理解できるはずである）
- 2) 入力と出力との間に、明示的には「狭義の相関関係」があるとはいえないものの、機械学習により「人工知能のモデル」が得られる場合であって、「何を特微量に選択するか、という点で、ブラックボックス」であったとしても、モデルができた後には、人間が、入力と出力との関係を分析して判断プロセスを理解可能である事例。
- 3) 入力と出力との間に、明示的には「狭義の相関関係」があることはわかっておらず、機械学習により「人工知能のモデル」が得られた場合に、モデルができた後に、人間には、入力と出力との関係を分析しても、何故、そのような入力と出力の関係が生じるのかが理解不能である事例。

このうち、明細書の記載要件についての事例やコメントでは、1) の場合を取り上げていることが多いのに対して、進歩性についての議論は、外見上、3) の事例を想定しているように見える場合が多い。ただし、本来、明細書作成者が出くわす事例としては、1) または 2) の場合がほとんどであろう。

(63) 人工知能のモデルへの入力に対する「特徴（量）選択」の手法としては、事前に入力と出力との「相関」を見る（たとえば、ランク付けする）という手法（フィルタ法などと呼ばれる）や、特微量を1つずつ追加して、あるいは1つずつ削除して、最適な予測性能を発揮するモデル入力を探索するもの（Wrapper 法と呼ばれる）もあるものの、人工知能自体が、学習の過程の中で自動的に特微量を選択する場合（Embedded 法などと呼ばれる場合もある）も多い。（データ駆動型人工知能ではむしろ特微量の抽出および選択が学習過程でなされることも特徴の1つ）。

特許庁の「AI 関連技術に関する事例の追加について」の記載において「相関関係」とは、「狭義の意味での相関関係（入力 x と出力 y との間に線形関係がある）」だけでなく、「広義の意味での相関関係（線形関係があるかどうかは自明ではないが、入力 x から出力 y が導かれる関係）」があることをも指そうとされているのではないかと推認される。なお、「特微量選択」の手法を開示したものとしては、例えば “Isabelle Guyon, Andr’e Elisseeff, “An Introduction to Variable and Feature Selection” Journal of Machine Learning Research 3 (2003) 1157-1182 Submitted 11/02; Published 3/03” の他、多数の報告がある。

特に、“Embedded 法”の手法には、汎化性能の向上が得られる手法の1つとして知られている「スパースモデリング」が含まれ、これは、現在の人工知能技術の研究開発における重要なテーマの1つとなっている。

したがって、この場合は、「学習時の入力データ（または出力データ）の選択のみを特徴とできるような発明」というものが想定されがたい。

b) 学習モデルへの入力の項目に、新たな一項目を追加することのみで、進歩性を主張できるような事例は、本当に存在しうるか？

このb)のような例は、もっとも極端であるとともに、問題の本質を反映すると考えられる。

出力データは、予測・分類の目的となる「目的変数」(出力)なのであるから、「学習」を実行するには、確定していると考えられるので、基本的には、「入力データの項目」(「特徴量：説明変数」を形成する)の選択に、特徴が生じる可能性があるというのが一般的であろう。そこで、最も極端な問いとしては、「学習用モデルへの入力の項目に、新たな一項目を追加すること」により、当業者では予測できないような「顕著な効果」(予測性能の向上等)が生じる例が存在するか、というものになる。

上記の問いに対しては、一例ではあるものの、後述するような「統計学上でよく知られた事例」によれば、「答えは(外見上)イエスの可能性がある」ということになる(このような事例は存在しうる)。

ただし、「ブラックボックス」をブラックボックスのまま、議論しても、残念ながら、人間が理解できない議論になってしまうため、以下のような条件のもとで検討してみることにする。

i) モデルへの入力データと出力データとの間の関係は、人間が理解できる事例を取り上げる。

ii) その入力データにより機械学習を実行して、学習済みモデルを生成し、一定程度以上の性能が現実実現されるのかは、その技術的な詳細には、ここでは立ち入らず、可能性の検討にとどめる。

さて、以下では、あくまで、上述したような事態が発生しうるのかを、人間にもわかる事例として検討してみるものの、筆者の理解では、問題の本質自体は、変わらないと考える。

このような事例としては、1951年にE. H. シンプソンによって記述された「シンプソンのパラドックス」^{(64), (65)} というものが有名であり、これは、母集団での相関と、母集団を分割した集団での相関は、異なっている場合があるというパラドックスである。つまり集団を2つ(または複数)に分けた場合にある仮説が成立しても、集団全体では正反対の仮説が成立することがある、というものである。

たとえば、図4.1のような調査結果のデータが得られたとする。これは、複数の被験者について、週ごとの運動量とコレステロール値について調査した結果を示すものとする。

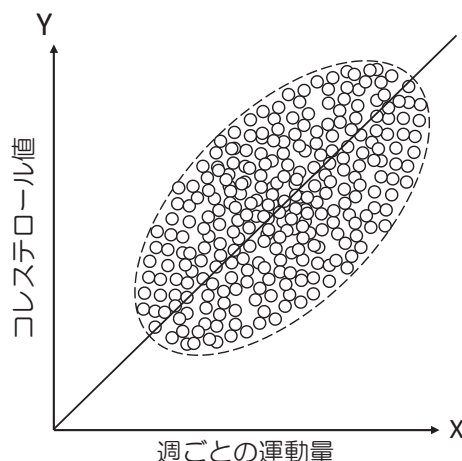


図4.1 運動量とコレステロール値についての調査結果⁽⁶⁶⁾

(64) Simpson, Edward H (著) “The interpretation of interaction in contingency tables.” Journal of the Royal Statistical Society. Series B (Methodological) (1951) : pp.238-241.

(65) Blyth, Colin R (著) “On Simpson’s paradox and the sure-thing principle.” Journal of the American Statistical Association 67.338 (1972) : pp.364-366.

(66) Judea Pearl (著), Madelyn Glymour (著), Nicholas P. Jewell (著), 落海 浩 (訳) 「入門統計的因果推論」, 朝倉書店, (2019/8/25), pp.5 図1. 2を基に筆者作成

この図4. 1を見る限りにおいて、「運動をすればするほど、コレステロール値が高くなる」（コレステロール値を下げるには運動をしない方がよい）という一般的な常識に反するような結果となっている。そこで、この相関関係をモデル化しようとする者があるかは、疑問ではあるが、「仮想的な事例」ということでご容赦願いたい。

そして、従来技術として、図4. 1の状況において、機械学習を実施して、「週ごとの運動量」をモデルへの入力とし、「コレステロール値」を出力とするようにニューラルネットワークで「学習済みモデル」を生成するというものが存在するとする。この場合、「学習用データ」の範囲で、精度は高くはないものの、一定程度の予測性能を有するモデルを作成すること自体は可能であるといえる。

これに対して、本件発明では、この図4. 1を年齢により層別化すると、以下の図4. 2のような状況となっていることを前提と想定する。これは、シン普森のパラドックスの1つの例である。

すなわち、各年齢層で分けてみれば、「運動をすればするほど、コレステロール値が低くなる」という、先ほどの集団全体でみたときとは、全く逆の相関関係が現れる場合があることになる。

このような例は、統計学上は、他にも数多く存在しうることが知られている⁽⁶⁷⁾。

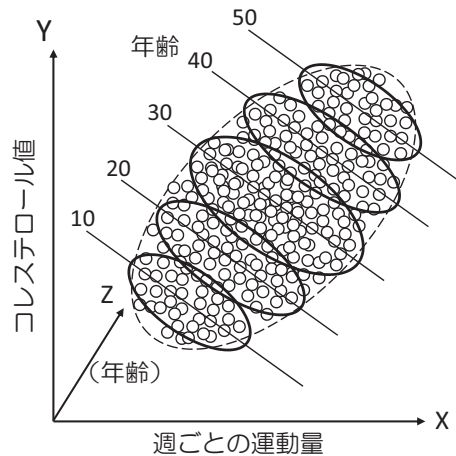


図4. 2 運動量とコレステロール値についての調査結果。年齢により層別⁽⁶⁸⁾

そして、 $(X, Z) = (\text{運動量}, \text{年齢})$ を2次元の説明変数として、 $Y = (\text{コレステロール値})$ を目的変数とする3次元でのデータの分布を考えれば、図4. 1は、あくまで、このような3次元の分布を、 $X - Y$ 面に垂直な方向から見た分布ということになる。

そこで、入力を (X, Z) 、出力を Y とするモデルを考えるとすると、図4. 2のように年齢で分割した集団（年齢で層別化した、全体集合に対する部分集合）に属するデータでは、「説明変数」と「目的変数」の関係では、全体としてみたときよりも、ばらつきが小さくなっていることからして、「学習用データ」の範囲内で、（図4. 1のデータのみからは想定できないほど）顕著に予測精度の高くなったモデルが生成されるのみならず、「学習用データ」の範囲外のデータについても、予測性能の良好な「学習済みモデル」が生成できる可能性があることも、推測できる。

さて、このような場合に、果たして、「モデルへの入力として、年齢を加えたこと」のみをもって、常に、**進歩性の根拠**とできるといえるであろうか？

「学習用モデルへの入力の項目に、新たな一項目を追加すること」で、著しい予測性能の向上が得られる

(67) たとえば、男性の群、女性の群の各々については、「薬効あり」との結果となるような薬剤候補について、男性と女性とを区別しない全体の集合では、「薬効なし」との結果となってしまうような例が存在することも知られている。

(68) (再掲) Judea Pearl (著), Madelyn Glymour (著), Nicholas P. Jewell (著), 落海 浩 (訳)「入門統計的因果推論」, 朝倉書店, (2019/8/25), pp.5 図1. 1を基に筆者作成

ことは新たな特徴である、との主張は可能であろう。

ただし、見方によっては、このようなモデルができたことは、「人工知能技術に対する“技術的な”貢献」があるというよりも、従来技術における、単なる「データ解析の不備」を発見しただけ、と言えなくもないようにも思われる⁽⁶⁹⁾。

しかも、請求項の記載が、単に、「モデルそのもの」に向けられているような場合は、上述した例からしても、「自然法則の利用」というよりも、「自然法則自体の発見」に限りなく近似する場合も想定され、「特許適格性⁽⁷⁰⁾」の観点からも論点となりうるであろう。

この状況で、単に、モデルの構造について、一切の言及をすることなく、「機械学習」との限定事項と、「入力と出力のデータを何にするか」との限定事項のみが含まれるクレーム（以下、「入出力限定クレーム」と呼ぶことにする）では、筆者としては、進歩性を認めてよいのかは、疑問なしとしない。

一方で、もちろん、「学習用データ」の範囲を超えて、他の環境で取得されたデータ（テストデータ）に対しても、正しい予測結果を出力できる（汎化性能がある）モデルが得られたとすると、確かに、それは、一定程度の「技術的な貢献」を達成しているといえる場合もあるであろう。

ただし、このような「技術的な貢献」が達成されることが、どのような構成の「学習用モデル」を用いて、どのような「学習処理」により得られるのかを明細書に具体的に記載することは必須であろう。そして、この場合に、仮に、「入出力」と「機械学習」との限定しかない「入出力限定クレーム」を特許請求の範囲に記載したのみで、「入出力の設定それ自体が新規で進歩性の根拠となる部分である」という主張をするのであれば、（モデルの構成や学習方法で予測精度は一般に異なるのであるから）課題解決の手段が請求項に反映されておらず、多くの場合は、「サポート要件違反」ということになるのではないだろうか？

また、仮に、「入出力限定クレーム」のまま、特許として登録されたとしても、権利範囲の解釈においては、結局のところ、そのような「学習済みモデル」は、「何を達成するのか？」という記載のみで、「学習済みモデル」を特定している一種の「機能的クレーム」といえるのだから、明細書に開示の「具体的な構成」を中心として、当業者が実施可能な範囲に限定して解釈されるというのが、妥当であろう。

もちろん、データ駆動型人工知能において、データの前処理や、モデルの構成そのもの、学習方法などについて、発明者が何らかの新たな技術上の工夫を導入したことにより、初めて、予測性能が高いモデルが生成できた、という事実に対応する記載がクレーム中にあったり、そのようにして生成されたモデルを新たな用途に使用する（利用する）との構成がクレーム中に存在するのであれば、その点をもって、進歩性の根拠とできる可能性があることについては、異論はない。

また、たとえば、「モデルの構成」やそれに対応する「学習方法」に新たな技術的な構成を創案した場合の「顕著な効果」としては、「モデルとしての予測性能」のようなものに限らず、「学習に要する計算負荷の低減（学習に要する時間の低減）」なども、より現実的なものとして想定される。

c) 完全に「データ駆動型」で特徴量の抽出が行われる場合はどうか？

この場合は、そもそも、「データ駆動型」では、何を特徴量とするのかは、「データの中から、人工知能ソフトウェアが自動的に抽出する」のであるから、「学習時の入力データ（または出力データ）の選択のみ」を特徴とできるか、という問い自体が意義があるものとは思われない。

(69) 「シンプソンのパラドックス」は、統計学上の「交絡」（目的変数と説明変数の双方に相関のある別の変数がある）の一種として説明される場合もある。また、計量経済学の世界などでは、経済上の施策の効果を図るにあたり、「介入」に対する「効果」の定量化のために、「交絡因子」の影響を除くための方法として、「回帰分析法」と「傾向スコア法」が使用されることが多い。このとき、「回帰分析法」では、「共変量」（目的変数と説明変数の双方と相関がある量）を、説明変数に追加することで分析するのは、周知な分析方法である。たとえば、以下を参照。

安井 翔太（著）「効果検証入門～正しい比較のための因果推論 / 計量経済学の基礎」技術評論社（2020/1/18）

(70) 筆者は、この点では、米国において、現在、「特許適格性」の判断の中に、クレーム中の“practical application”への言及の存否が取り上げられていることは、あながち、間違っているとはいえないとも考える。

当然ながら、明細書には、「モデルの構成」やそれに対応する「学習方法」についての実施可能要件を満たすような開示が求められ、請求項にも、「サポート要件」の観点から、このような構成・処理に対して、一切の言及をしない、ということは想定できないであろう。

したがって、上述した a) ~ c) の事例からみて、筆者としては、「請求項の記載において、発明のポイントは、モデルへの入力・出力として何を選ぶかという部分のみであって、単に機械学習との限定しかない」発明については、特許を付与することに、抑制的であるべきと考える。そして、仮に、そのポイントのみが進歩性の根拠であると主張するのであれば、そのモデルにより奏される「効果の顕著性」について、これまで説明してきたような事情があることを超えるような説得力のある説明が必要と考える。

現状の特許制度においては、明細書の開示要件として「学習用データ等」そのものを第三者に対して公開することは要求されていない。とすると、「データを公開することもなく」、「機械学習の技術的な構成を公開することもない」特許明細書・クレームに対して、独占的な権利を付与することは、「技術の進歩に何ら貢献することのないもの」に権利を付与することになりかねない。そこで、学習用モデルへの入力の項目に新たな項目を追加することによる進歩性の主張については、以下のように考える。

a) クレームに、先行技術との相違として、単に選択された「新たな入力データの項目」が記載されるだけでは多くは不十分ではないか？

選択を人間がしたなら、「単なる設計事項」との相違を主張できるような根拠の開示が必要と考える。

b) 基本的には、「新たな入力データの項目」のみではなく、それを選択した技術的意義が把握できる程度に、「新しい入力データの項目」に対して、

i) 入力データとしての観測量から特徴量の抽出・選択までに、いかなる技術的処理がされるのか（前処理／特徴量抽出・選択） および／または

ii) いかなる「学習モデル」を採用し、いかなる「学習処理」が実施されるか、

の記載が必要となると考える。

3) 「前処理を実行する H / W, 前処理を実行するプログラム」

上述のとおり、「学習用データの生成方法」のクレームは、その保護できる範囲が、想定に比べて、限定的である可能性がある。そこで、むしろ、「前処理を実行する H / W, 前処理を実行するプログラム」の特許クレームの対象とするという戦略も考えられる。この場合は、「前処理を実行する H / W, 前処理を実行するプログラム」を、クローズとする戦略もあり得る。

5. まとめ

(1) 「超スマート社会 (Society 5.0)」において、中核的に重要となると考えられる「(データ駆動型) 人工知能技術」に関連する特許の役割の 1 つは、「データ」が「商業的な取引の対象」となることに、特許権（または知的財産権）が寄与する、という形で制度やビジネスが構築されることであろう。具体的には、「データ構造」、「学習用データの生成方法」、「学習済みモデル」などの特許権が、技術を標準的に使用できる環境を作るために、データ拠出側のインセンティブとして働くような制度の構築が必要であろう。

(2) まずは「データ」が「商業的な取引の対象」となることで、より大きな「データ流通の市場」などが想定されるものとする。この点では、上記のような特許権などの知的財産権が、「データ収集／分析者」とも呼ぶべき存在を市場の中で形成できることも 1 つの方法論として検討されるべきではないだろうか？

(3) 「データ駆動型人工知能」により、「データ駆動型ビジネスモデル」とも呼ぶべき存在が生まれるのは、上記のようなデータの取引の活性化を背景に、「データ分析者（人工知能を使う者）」の数と質が向上することが必要と考える。この意味では、「特許制度」は、むしろ、「新しいモデル構造」や「新しい学習方法」の開発などを奨励する方向で運用されるべきと考える。

「データ駆動型人工知能」とはいっても、「ブラックボックス」のままの権利が発生することは、この意味

でも、むしろ、抑制的であるべきであろう。筆者としては、研究・技術の世界で、いわゆる「シンギュラリティ⁽⁷¹⁾」が到来する前に、特許の世界で、人間の方から、いち早く「人工知能」に「手を上げて」しまう必要はない (There is no reason to throw up our hands now.) と考える。

(71) 「シンギュラリティ (技術的特異点)」とは、AIなどの技術が、自ら人間より賢い知能を生み出す事が可能になる時点を指す言葉とされる。