



AI の知財に関する法律上の諸論点

執筆者: 福岡 真之介

本ニュースレターでは、AI を開発・利用するにあたっての法的問題を概観する。2017 年も年末となり、1 年を振り返ると AI ブームの年であったように感じるが、今年の流行語大賞には一切なっていないようであり、AI に関わる者としては若干残念に感じるところである。本稿では、AI の知財に関して、どのような法律問題点があるかを述べた上で、AI の知財をどのように保護するのか、AI を開発・利用する際にはどのような法律問題に気を付けなければならないかについて概括的な解説を試みたい。

1 AI の開発・利用過程

AI の開発・利用過程は典型的には以下のようなものが考えられる。

- ① ユーザが、開発者にデータを提供
- ↓
- ② 開発者が、AI を開発(既に関済済みの場合もある)
- ↓
- ③ 開発者が、受領したデータについて、自らが開発した AI のプログラムに適合するように学習用データセットを作成
- ↓
- ④ 開発者が、学習用データセットを用いて AI のプログラムに学習させる
- ↓
- ⑤ 学習済みモデルが完成、開発者がユーザに納品(あるいはサービス提供)
- ↓
- ⑥ ユーザ又は開発者が、追加データから作成した新学習用データセットを学習済みモデルに投入して、追加学習させる

本ニュースレターは法的助言を目的とするものではなく、個別の案件については当該案件の個別の状況に応じ、日本法または現地法弁護士の適切な助言を求めて頂く必要があります。また、本稿に記載の見解は執筆担当者の個人的見解であり、当事務所または当事務所のクライアントの見解ではありません。

本ニュースレターに関する一般的なお問い合わせは、下記までご連絡ください。

西村あさひ法律事務所 広報室 (Tel: 03-6250-6201 E-mail: newsletter@jurists.co.jp)

このような一連の流れは、①学習用データセット作成段階、②学習済みモデル作成段階、③学習用モデル利用段階に分けることができる。

AI を知財として保護する主要な方法として、著作権法、特許法、不正競争防止法が考えられるが、本ニューズレターでは著作権について論じる。

著作権法で保護される著作物とは、①思想又は感情を②創作的に③表現したものであって、④文芸、学術、美術又は音楽の範囲に属するものである(著作権法 2 条 1 項 1 号)。AI 特有の問題として、この①思想又は感情とは、人間の思想・感情を意味することから、AI 作成物が著作物といえるのがクローズアップされることになる。また、単なる事実としてのデータは、思想又は感情を表現したものではないので、著作物にはならないが、AI 時代ではデータの重要性が増していることから、データの権利保護と利用が問題となる。

2 学習用データセット作成段階の法律問題点

AI を学習させるためには、学習用データセットが必要となるが、学習用データセットを作成するためには、元となる生データが必要である。生データとしては、例えば、ウェブサイトや SNS のデータや、顧客の購買履歴や工場の稼働データなどユーザが保有するデータなどがある。

そこで、学習用データセットの作成時に、生データとしてウェブサイトにアップロードされている画像・文章データなど他人の著作物を利用する場合に、他人の知的財産権を侵害することにならないかが問題となる。

また、学習用データセットは、生データを収集して、選別し、加工することによって作成される。また、教師あり機械学習をするには、生データに正解データ(ラベル)をつけていかなければならないが、これには膨大な労力を要する。加えて、生データの選択・取得方法や、生データから学習用データセットに加工するためのノウハウも重要であり、このノウハウは AI に関する深い理解や多くの試行錯誤と経験によって形成されるものである。そのため、学習用データセットの知財の保護が問題となる。

他人の著作物を無断で利用することは原則として著作権侵害となるので、他人の著作物を含んだデータに基づいて学習用データセットを作成することは原則として複製権・翻案権などを侵害することになる。もっとも、著作権法 47 条の 7 の規定により、コンピュータの情報解析を目的とする場合に、必要と認められる限度で、記録媒体への記録と翻案(改変)をすることが例外として認められている。

なお、著作権法 47 条の 7 により認められているのは、記録媒体への記録と翻案に限られており、公衆への譲渡やウェブへのアップロードは認められておらず、そのような行為は譲渡権・公衆送信権の侵害となる。したがって、学習用データセットを作成して一般に販売するビジネスは、そのデータセットが他人の著作物を含んだデータを含む場合、現行法の下では、ビジネスとして成立することに制約がある。

学習用データセットは、データベース著作物として認められれば著作権により保護される。データベース著作物として認められるためには、①情報を検索できるように体系的に構成されていること、②情報の選択や体系的な構成によって創作性を有することが必要である(著作権法 12 条の 2 第 1 項)。これらの要件の中で、「創作性」の要件を満たすか否かが問題になることが多い。汎用性の高いデータベースであればあるほど、AI に学習させるための使い勝手はよくなるが、情報の選択や体系的な構成は一般的なものを採用することになるので、「創作性」が認められにくくなり、データベース著作物として認められなくなってしまうリスクがある。

3 学習済みモデル作成段階の法律問題点

機械学習では、プログラムを構築し、これに学習用データセットを与えて、入力データと出力データの誤差を最小化するようにパラメータの調整が行われる。このような調整を繰り返し計算することで生成されるのが学習済みモデルである。AI を学習させるためには、膨大な労力・費用・マシンパワーやノウハウを要することから、それらに対する投資も必要である。そのため、学習済みモデルの権利者が誰なのか、また、その権利をどのように保護するのが問題となる。

学習済みモデルは、プログラム部分と学習済みパラメータなどの数値データ部分で構成されている。AI のプログラム部分が汎用的プログラムや OSS の場合には、プログラム部分では大きな差がつかないため、学習済みパラメータこそが AI の競争力の源泉といえる。しかし、この学習済みパラメータは、AI のプログラムが自動的に生成するデータであることから、著作権が生じるのが問題となる。

学習済みモデルは、「プログラム＋学習済みパラメータ」であるといえるが、プログラム部分については、著作権法はプログラム著作物を保護しているところ、大規模であったり複雑なプログラムであればプログラム著作物として認められる可能性が高い。他方、学習済みパラメータについては、AI のプログラムが自動的に作成しているものであることから、著作物の要件である「思想または感情」の表現といえるのか、また「創作性」があるのが問題となる。この点は、著作権法で規定されていないことはもちろん、現時点では裁判例もないことから何ら確立した判断基準がない。この点、学習済みパラメータは、AI のプログラムが自動的に作成していることから、「思想または感情」の表現ではない、あるいは創作性がないとして、著作物として認められない可能性がある。もっとも、学習済みパラメータは、人間の AI の設計思想が反映されたニューラルネットワークの構造や学習によって生成されたものであり、そこに人間の創作的寄与が認められるとして、「思想または感情」の創作的表現であると解する余地はある。

また、学習済みモデル全体を、プログラムと学習済みパラメータとに分けて考えずに、一つのプログラムとして考えたとしても、著作物の一部を侵害した場合に著作権侵害になるのは、その部分が創作性を有している場合であるとされており、結局、上記と同様に創作性の有無の問題に帰着する。

このように、現状、学習済みモデルの学習済みパラメータ部分については著作権が認められるかは不透明である。そのため、AI の法的な保護としては、契約で学習済みモデルの権利関係を定めることや、営業秘密化することで不正競争防止法による保護を図ることが現時点における現実的な対応といえよう。

4 学習済みモデル利用段階の法律問題点

学習済みモデルが完成すると、これを実用に供することになる。しかし、学習済みモデルが一度完成すればそれで終わりではない。さらに新しいデータを使って学習させ、より高い精度が出るように学習済みパラメータを調整し、学習済みモデルを進化させていくことが想定される。そのため、元の学習済みモデルの権利者やデータの提供者が、進化後の学習済みモデル(派生モデル)に対してどのような権利を有するかが問題となる。また、元の学習済みモデルの権利者以外の者が、学習済みパラメータを改変して、派生モデルを作成することが適法かが問題となる。

派生モデルが、元の学習済みモデルの学習済みパラメータ部分だけを改変したものである場合には、元の学習済みモデルのプログラムの著作権者は、派生モデルのプログラムの著作権を有していることになる。

学習済みパラメータ部分については、その部分に著作権がないのであれば、誰もが利用し、改変することが可能となる。この場合、元の学習済みモデルの権利者であっても、学習済みパラメータ部分について権利を主張できないことになる。

他方、学習済みパラメータ部分に著作権がある場合には、原則として無許諾で利用・改変できない。この場合、元の学習済みモデルの権利者は、学習済みパラメータ部分について二次的著作物としての権利を主張できる可能性がある。もっとも、学習済みパラメータがプログラムの著作物である場合には、プログラムの著作物の複製物の所有者は、自らコンピュータで利用するために必要と認められる限度で複製・翻案することができる(著作権法 47 条の 3 第 1 項)。

5 最後に

以上のとおり、AI を開発・利用するにあたっての法律問題は多岐にわたる。我が国の AI の開発・利用を促進するためには、これらの法律問題の対処法についての理解を広め、また解決すべき法律問題について立法などにより解決するなどして、権利者の権利が適切に守られ、他方で、利用者も安心して利用できるような法制度・法律実務を構築していくことが重要であるとする。

なお、本稿で取り上げた AI の法律問題の詳細については、2018 年のニューズレターで解説していく予定である。

以上



ふくおか しんのすけ
福岡 真之介

西村あさひ法律事務所 パートナー弁護士

s_fukuoka@jurists.co.jp

96 年東京大学法学部第一類卒業。98 年弁護士資格取得。01 年西村あさひ法律事務所に所属。06 年デューク大学ロースクール卒業(LL.M.)、06-07 年シュルティ・ロス・ゼイベル法律事務所(米国)勤務、07-08 年ブレイク・ドーンソン法律事務所(オーストラリア)勤務。著書は、『IoT・AI の法律と戦略』(商事法務・2017)等多数。

西村あさひ法律事務所では、M&A・金融・事業再生・危機管理・ビジネスタックスロー・アジア・中国・中南米・資源/エネルギー等のテーマで弁護士等が時宜にかなったトピックを解説したニューズレターを執筆し、随時発行しております。

バックナンバーは<<https://www.jurists.co.jp/ja/newsletters>>に掲載しておりますので、併せてご覧下さい。

(当事務所の連絡先) 東京都千代田区大手町 1-1-2 大手門タワー 〒100-8124

Tel: 03-6250-6200 (代) Fax: 03-6250-7200

E-mail: info@jurists.co.jp URL: <https://www.jurists.co.jp>

© Nishimura & Asahi 2017