

機械学習と知財・契約



2019.11.22

STORIA法律事務所
弁護士 柿沼太一

【自己紹介】

- ▼ 2000年4月に弁護士登録
- ▼ 2015年3月に神戸三宮にSTORIA法律事務所設立
- ▼ AI、IT、知的財産、ベンチャーを主として取り扱う
- ▼ 2016年10月からAIに関して積極的な情報発信を始め、現在自動車系、医療系、工場系、WEB系など多様なAI企業からの相談、顧問契約を締結
- ▼ 2017年12月東京事務所開設
- ▼ 経産省の「AI・データ契約ガイドライン」検討委員会委員
(~2018.3)

機械学習と知財・契約（総論）

■ 収集→開発→展開の流れ

収集

- ・ 収集・利用するデータに知財や法的規制がかかっているものが多い
- ・ 特に個人情報と著作権法

知識・情報が重要

開発

- ・ 価値があるにもかかわらず、センサデータなどそもそも知財権では守れないものがある。
- ・ アルゴリズム・ソフトウェアなど、制度的に特許権で守るのが難しいものがある。
- ・ 成果物が横展開可能なので全員横展開したいという希望を持っている

何が知財権で守れるのかを知っておくことが重要

* それ以外にも精度保証・瑕疵担保や責任の問題あり

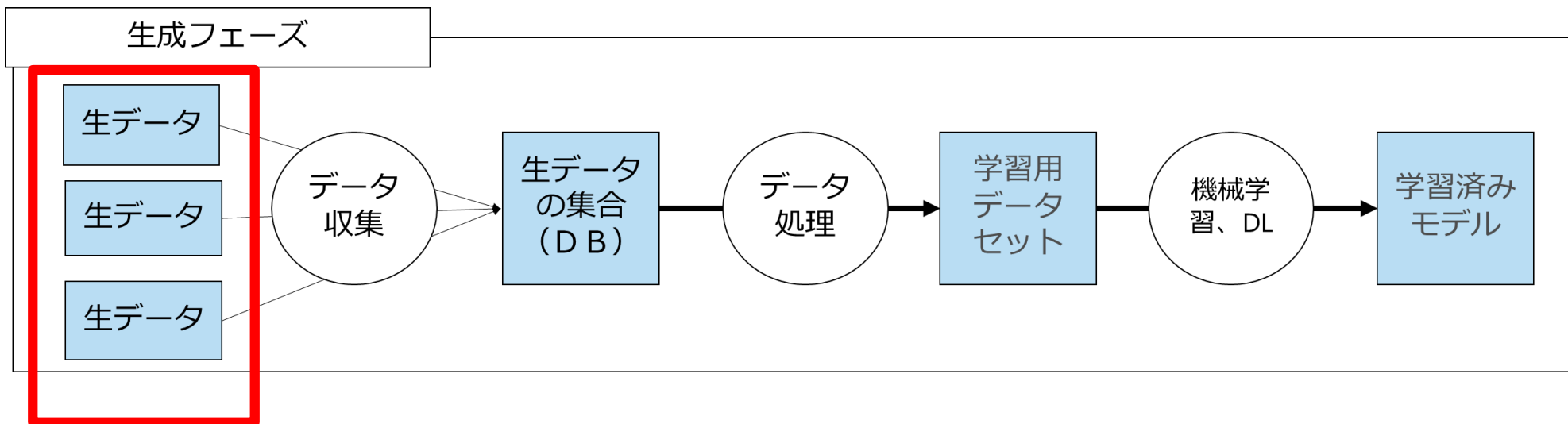
展開

- ・ 成果物や派生物が思わぬ価値を生むことがある。
- ・ 「作って終わり」「作業して終わり」にしない方が双方メリットあり。

開発契約
+利用契約が重要

機械学習と知財・契約（各論） ～データ収集～

1 データ収集の問題点



「生データ」と言っても様々な種類のデータがあり、規制が異なっている
→ 「データの種類×データの取得方法」で整理するとわかりやすい

1 データ収集の問題点

1 データの種類

- ① 個人情報
例：医療AI
- ② 著作物
例：コンテンツ生成AI
- ③ 肖像権
例：顔認識AI
- ④ 事実を示すデータ
例：投資AI
- ⑤ センサーデータ
例：工場用ロボットAI



2 データの取得方法

- ① 自ら生成
- ② 第三者からデータ利用契約以外の契約に基づき収集
- ③ 第三者からデータ利用契約に基づき収集
- ④ インターネット上で収集



3 規制

- a 法律による規制
- b 契約による規制
- c a+b
- d 規制なし

1 データ収集の問題点

特によく問題となるのは

個人情報と著作権

1 データ収集の問題点

【具体例】

1 患者さんの医療データについて医療機関から提供を受けて疾病診断用モデルを開発する

- 医療データ：個人情報（要配慮個人情報）
- 個人データの第三者提供：本人の同意が必要

→これから新たにデータ採取をするのであれば（前向き研究）本人同意取得可能だが、過去のデータを利用する場合（後ろ向き研究）本人同意取得不可

→どうすればよいのか（委託スキーム、匿名加工スキーム、次世代医療基盤法等）

1 データ収集の問題点

【具体例】

2 AIモデル内蔵カメラを店頭提供し、撮影された顔写真をリアルタイムで解析して店員に情報提供する

- ・ 顔写真：個人情報
- ・ 個人データの第三者提供：委託スキームで整理すれば本人同意不要。ただし、適切な形で利用目的の公表等が必要。
- ・ ベンダが複数の会社から提供を受けた顔写真を利用して、モデルを生成することについては、委託の範囲を超えるため、本人同意が必要となる可能性が高い。

1 データ収集の問題点

【具体例】

3 自動翻訳エンジン生成のため、ネット上の自然言語データをクロールし、当該データを用いて学習用データセットやモデルを生成・販売する

自然言語データ：著作物

著作物：原則として著作権者の同意なく複製等は不可能

→著作権法30条の4：機械学習モデル生成等「情報解析」のためであれば原則として著作権者の同意なく複製等が可能

→利用規約違反でなければ適法

→日本独自の規定：機械学習パラダイス

1 データ収集の問題点

【まとめ】

- この領域は知識・情報が重要
- 「データの種類」×「データの取得方法」で分析する
- 特に気を付けなければならないのは個人情報と著作権

機械学習と知財・契約（各論）
～開発～

【よくある質問】

ベンダとして、事業会社との共同プロジェクトを立ち上げる際の、知財・法務に対する最初の二ギリ方（契約書への文言の入れ方）が事業会社側、ベンダ側とも曖昧なケースが多く、実際に高性能な学習済みモデルや中間成果物、ノウハウが発生した際にモメそうな不安があります。

（１） 通常システム開発と学習済みモデル開発の違い

- ① 通常システム開発と異なり、複数の材料、中間成果物、成果物が存在する
- ② 材料、中間成果物、成果物が高い価値を持ち、ユーザ・ベンダ共に独占/再利用したいという需要が存在する

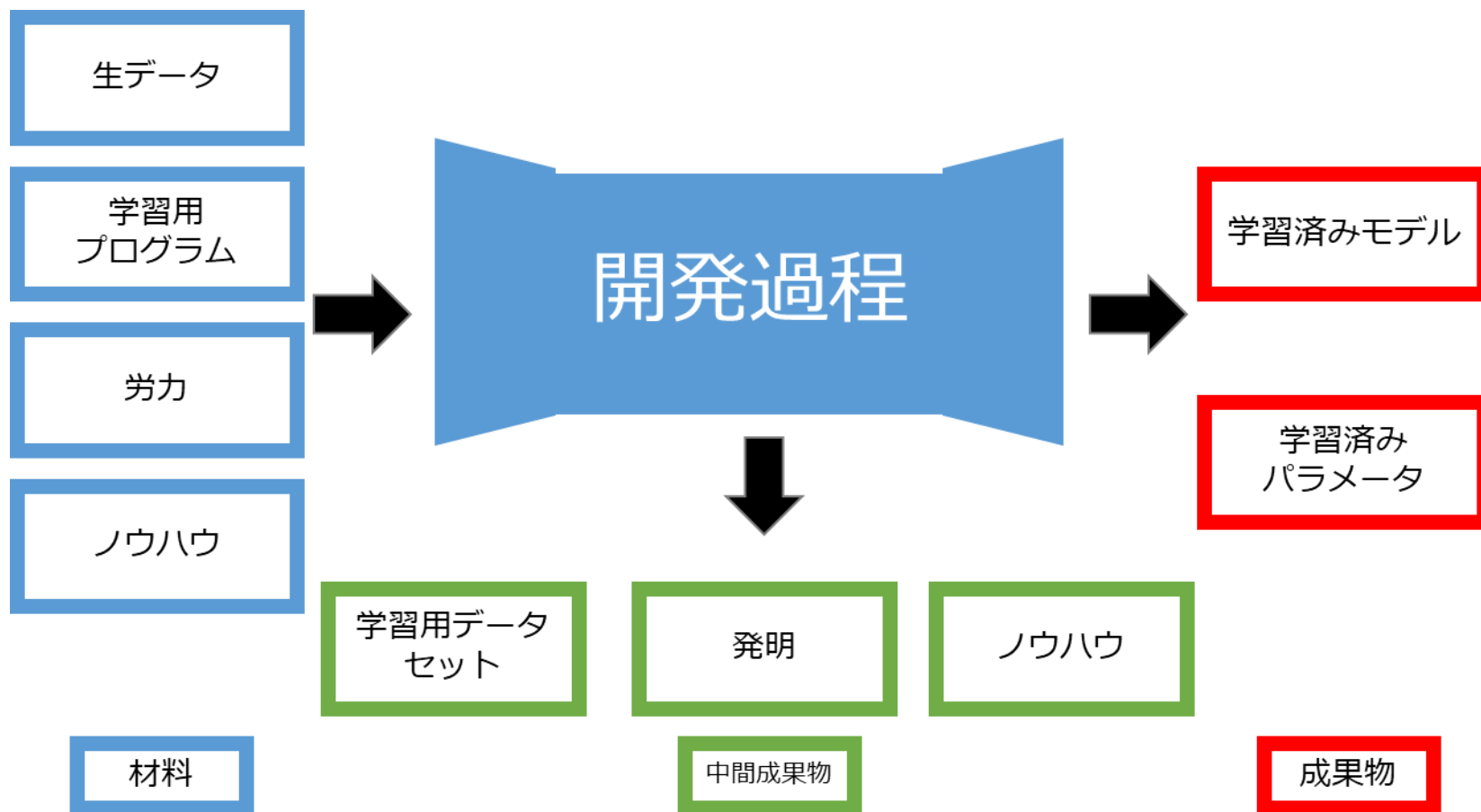
① 通常のシステム開発と異なり、複数の材料、中間成果物、成果物が存在する

【通常のシステム開発の場合】



① 通常のシステム開発と異なり、複数の材料、中間成果物、成果物が存在する

【学習済みモデル開発の場合】



② 材料、中間成果物、成果物が高い価値を持ち、ユーザ・ベンダ共に独占/再利用したいという需要が存在する

▼ ユーザ

学習用データセット、学習済みモデルは自社のノウハウや機密が詰まった生データを用いて生成されたものであり、開発に際して委託料も支払っていることから、自社で独占したい。

▼ ベンダ

生データを用いて学習用データセットを生成する過程、訓練過程いずれにおいても自社の高度のノウハウ及び多大な労力を用いていること、学習用データセットや学習済みモデルは再利用が可能であることから、この開発案件以外にも横展開したい。

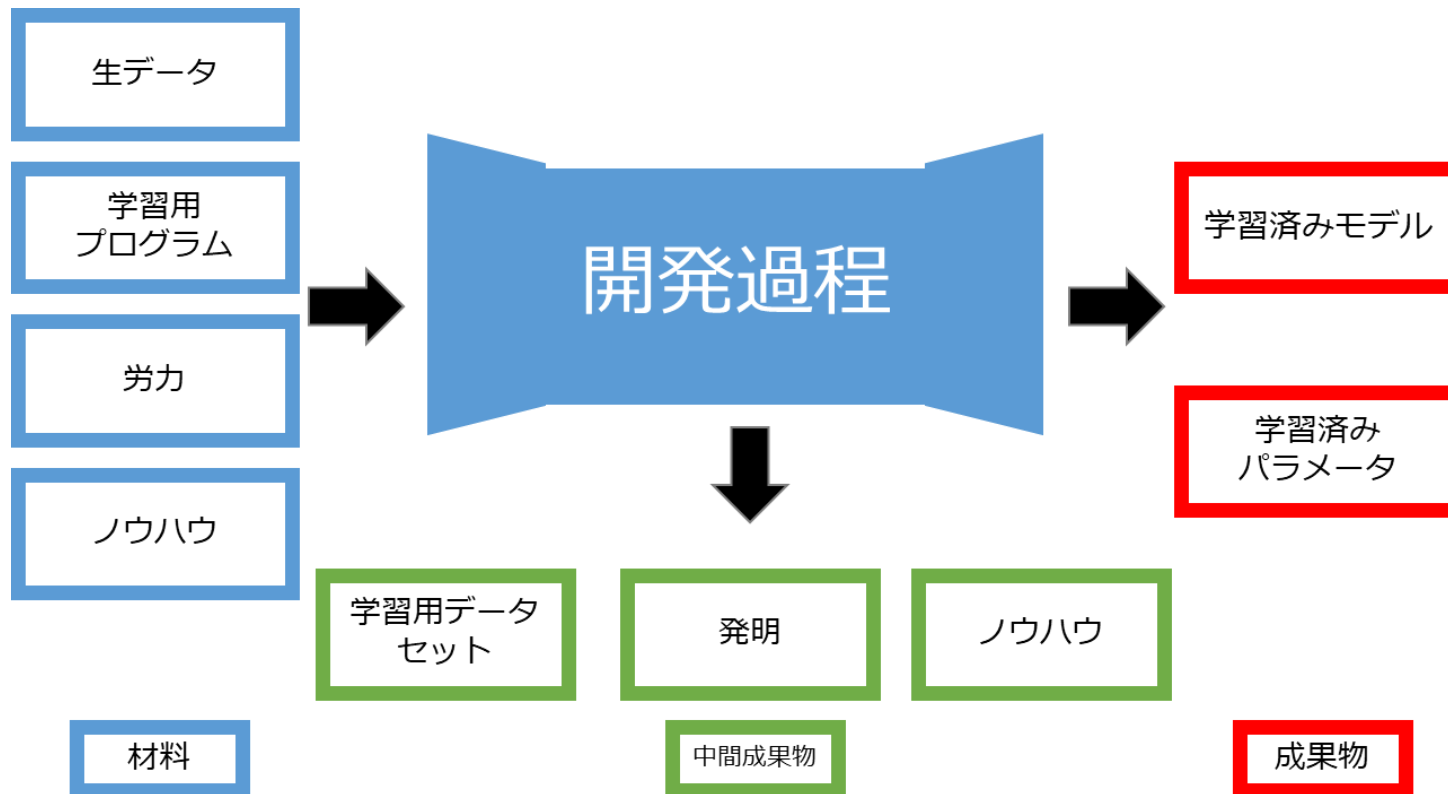
→この2つの相反する希望を調整する枠組が必要

（２） どうしたらよいのか

① 材料・中間成果物・成果物について、何が知的財産権の対象となるのか・ならないのかを知っておく

② ①についてデフォルトルール（＝法律上のルール）として誰がどのような権利を持っているかを知っておく

- ① 材料・中間成果物・成果物について、何が知的財産権の対象となるのか・ならないのかを知っておく
- ② 知財法の規定（デフォルトルールとして誰がどのような権利を持っているか）を知っておく



- ① 材料・中間成果物・成果物について、何が知的財産権の対象となるのか・ならないのかを知っておく
- ② 知財法の規定（デフォルトルールとして誰がどのような権利を持っているか）を知っておく
 - 1 生データ
 - 2 学習用データセット
 - 3 学習用プログラム
 - 4 学習済みモデル
 - 5 学習済みパラメータ
 - 6 ノウハウ

機械学習と知財・契約（各論）

	特許法	著作権法	不正競争防止法
生データ	×	△ (著作物性がある データのみ)	○ (要件を満たす場合。 以下同様)
学習用データセット	×	△ データベースの著作 物に該当する場合	○
学習用プログラム	○ アルゴリズム部分	○ コード部分	○
学習済みモデル (推論プログラム+学習 済みパラメータ)	○ アルゴリズム部分	○ コード部分	○
学習済みパラメータ	×	×	○
ノウハウ	○ 「発明」の要件を満 たす場合	×	○

機械学習と知財・契約（各論）
～展開～

■ 「展開」の重要性

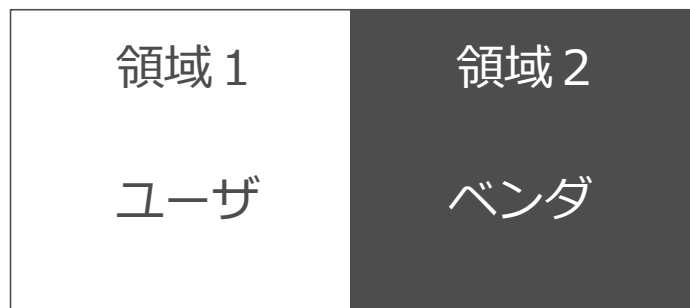
- ・ 成果物や派生物が思わぬ価値を生むことがある。
- ・ 「作って終わり」「作業して終わり」にしない方が双方メリットあり。

→完成した成果物を、契約当事者双方がどのように使えるかを「開発+利用契約」において定めることが非常に重要となる。

→ただし、いきなり「開発+利用契約」を検討・交渉すべきではない。順番は、「双方における成果物の展開方法の検討・交渉」→「開発・利用契約の交渉」となる。

→展開方法についての検討・交渉がないまま開発・利用契約の内容について交渉をしても意味がない（が、結構そのようなケースはある）

■ 機械学習モデル・データビジネスにおける展開パターン



① 事業領域ごとに棲み分けるパターン



② 全事業領域において双方が自由に事業展開するパターン



③ 一方が自社のみ利用し、他方は全領域において利用するパターン



④ 一方当事者のみが事業展開するパターン（買い切り・売り切り）

【典型的な暗礁乗り上げパターン】

■ ユーザ

学習用データセットや学習済みモデルは、うちのノウハウや機密が詰まった生データを用いて生成されたものですし、開発に際して委託料も支払っています。うちに権利がありますよね？

■ ベンダ

生データだけでは学習済みモデルは生成できません。高性能なモデルができるのは、データの前処理やモデルの訓練過程いずれにおいてもうちの高度のノウハウと多大な労力あってこそです。うちに権利がありますよね？

- 「どちらが権利を持っているか」（権利の帰属）に双方がこだわっている限り永久に溝は埋まらない。

- 交渉に多大な労力と時間がかかり結局競争力を失う。

→「権利の帰属」とは別の視点が必要

→本来、ユーザとベンダのビジネス構造は異なるのであって、双方のニーズを同時に満たしうる契約条件は当事者双方が思うよりもっと多いはず。

→対象物の「権利帰属」ではなく「利用条件」を調整することで、双方が合理的なビジネス展開ができるような契約内容にする。

対象物の「権利帰属」ではなく
「利用条件」を調整することで、双
方が合理的なビジネス展開ができる
ような契約内容にする。

→極端な話、自社が学習済みモデルに関する権利を保有していなくても、「モデルの第三者提供を含め、何の制限もなくモデルを自由に利用できる」という利用条件を設定できれば、実質的にはモデルの権利を保有していることとほとんど同じ。

■ 開発契約+利用契約の内容について

「展開パターン」についての双方の検討・交渉を行ったうえで、開発契約+利用契約の内容についての交渉に入ることになる。

【視点】

「誰が権利・知財を持っているか」（権利帰属）

+

「誰がその権利・知財をどのように利用できるか」（利用条件）
の視点が必要

→開発契約において権利帰属、利用契約において利用条件を決めるというイメージを持っておけばよい。

■ 具体例

領域 1	領域 2
ユーザ	ベンダ

① 事業領域ごとに棲み分けるパターン

▼ 権利帰属

ユーザに全部権利帰属

▼ 利用条件

ユーザはベンダに対して領域 2 についてのみ完全独占的利用許諾（領域 2 についてはユーザも展開不可）

■ 具体例



② 全事業領域において双方が自由に事業展開するパターン

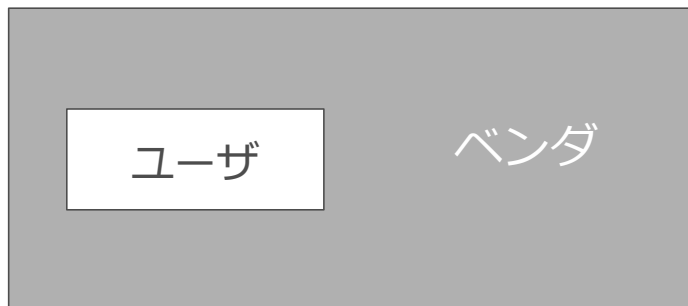
▼ 権利帰属

ユーザ・ベンダ共有

▼ 利用条件

ユーザ・ベンダ双方が相手の同意及び利用料の支払いなく、自己実施・利用、第三者に対するライセンスが可能

■ 具体例



③ 一方が自社のみ利用し、
他方は全領域において利用す
るパターン

▼ 権利帰属

ベンダに全部権利帰属

▼ 利用条件

ベンダはユーザの自己実施・利用に限って非独占的に無償で
利用許諾。

■ 具体例



④ 一方当事者のみが事業展開するパターン（買い切り・売り切り）

▼ 権利帰属

ユーザに全部権利帰属

▼ 利用条件

ベンダは一切利用不可

【まとめ】

- ・ データ収集フェーズにおいては、各種知財・規制に関する知識・情報が重要
- ・ 開発フェーズにおいては、開発における材料・中間成果物・成果物についてどのような知財権が発生するか、あるいはしないかを知っておく。
- ・ 展開フェーズにおいては、展開方法を検討したうえで、開発契約（権利帰属）+利用契約（利用条件）をセットで検討する必要がある。



STORIA

STORIA法律事務所について

弁護士法人STORIA（ストーリーア）

<https://storialaw.jp/>

【神戸】

〒650-0031

神戸市中央区東町123-1 貿易ビル8階

STORIA法律事務所

【東京】

〒100-0004

東京都千代田区大手町1丁目6-1

大手町ビル6階（Inspired.Lab内）

STORIA法律事務所東京オフィス

弁護士 柿 沼 太 一

e-mail kakinuma@storialaw.jp