

新たな情報財検討委員会（第3回）

論点：「コミュニティ利用」

国立研究開発法人 産業技術総合研究所

情報・人間工学領域長 関口智嗣

博士（情報理工学） 技術士（情報工学部門）

2016年12月19日

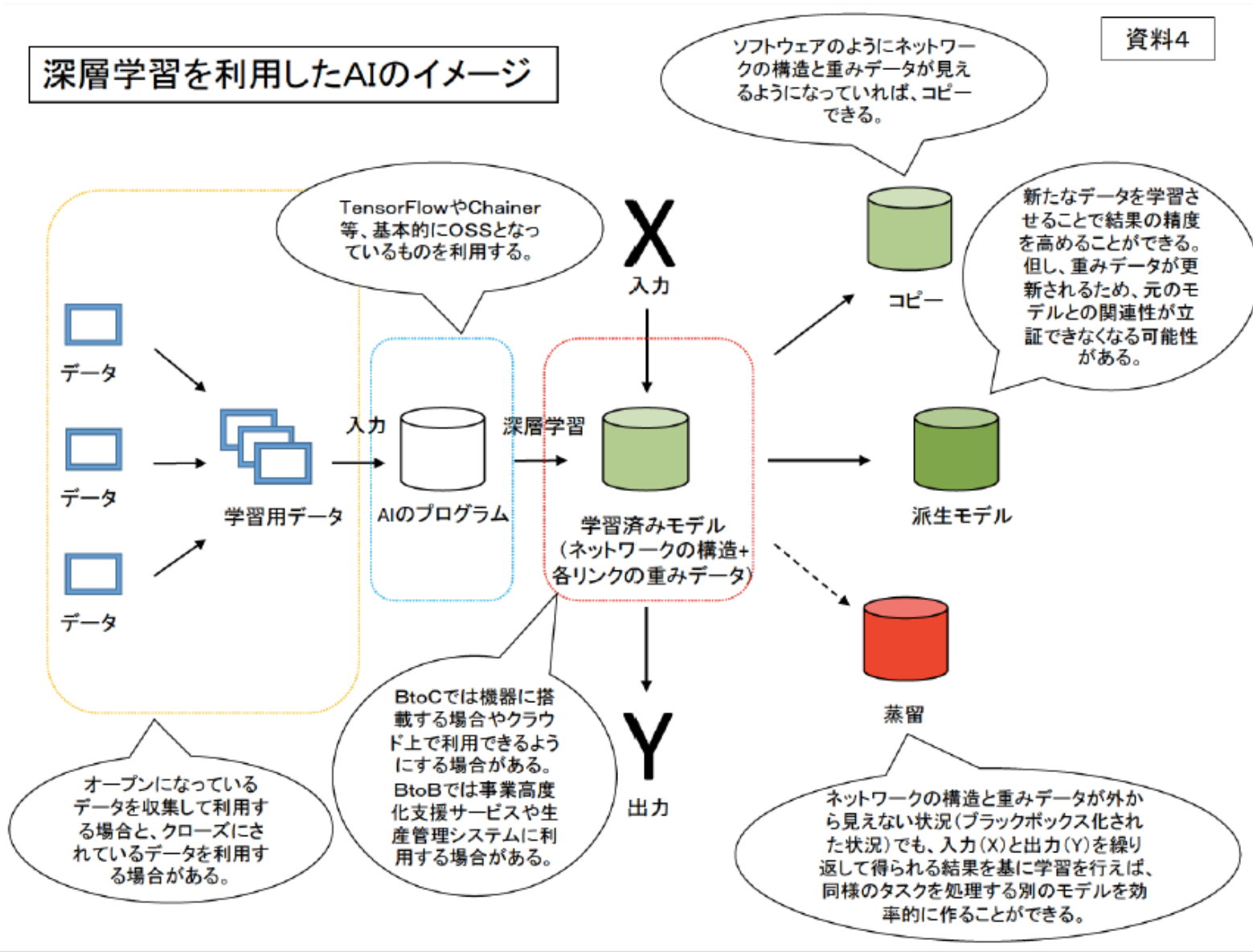
Disclaimer: 個人的な見解です。

本報告の論点：
AI学習をコミュニティ〈協調領域〉で
実施する場合の「情報財の論点」を整理

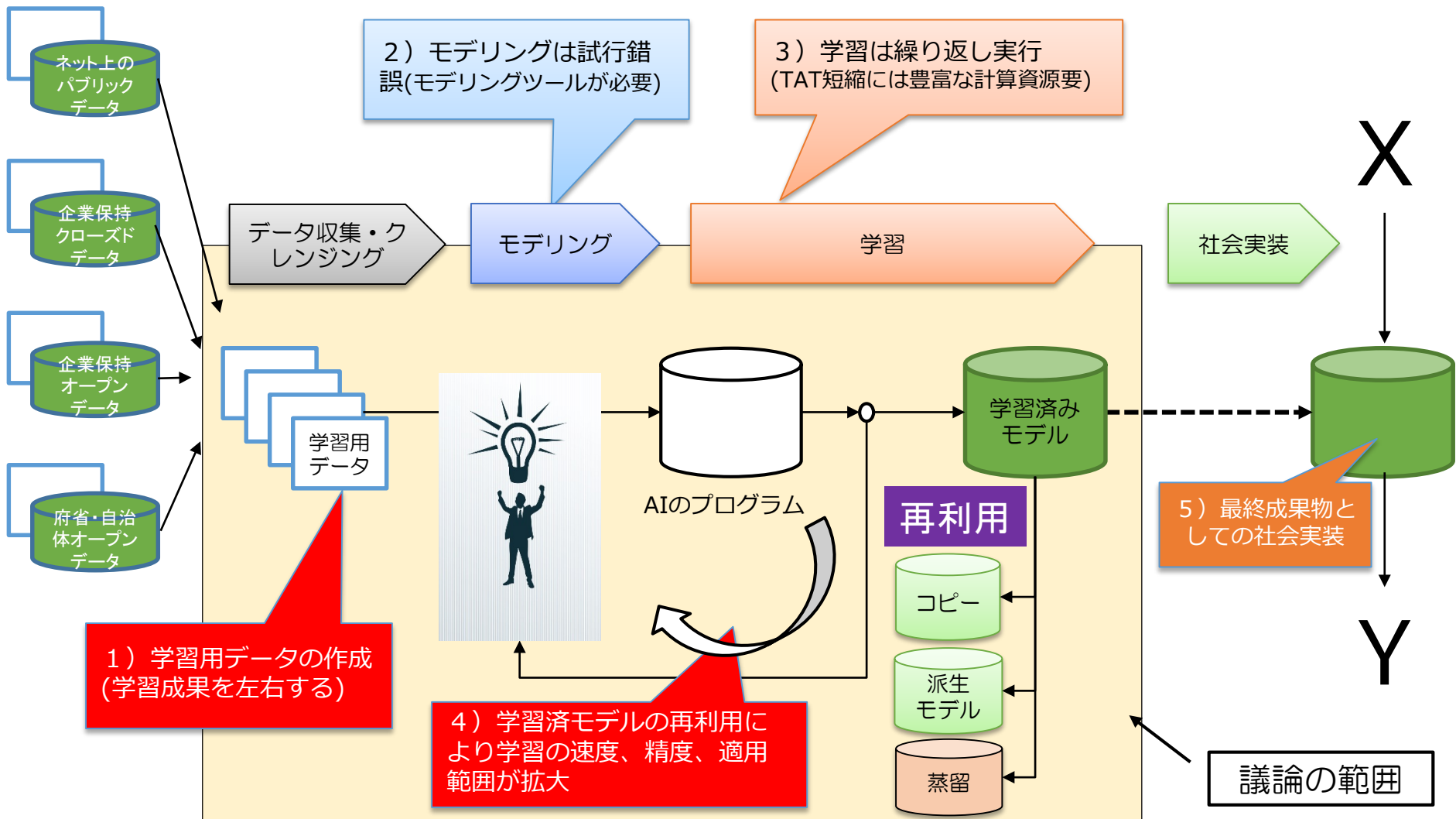
(参考) 新たな情報財検討委員会 (第2回) より

深層学習を利用したAIのイメージ

資料4



コミュニティによりAI学習効率を向上して、社会実装を加速



単独で実施するよりも、**コミュニティ**で実施した方がAI学習効率は向上する「**知の循環**」⇒社会実装の加速

AI学習形態の選択

オープン・クローズ戦略 ①～③について検討

クローズ



オープン

AI学習の技術調達と権利	学習用データセット		学習済モデル	
	作成	利用	作成	権利
インソース (独自調達)	自社内 (47条の7)		自社	
アウトソース (外部調達)	自社／外注先 (47条の7-適用可能)		外注先	自社 or 外注先 or 共有 (契約で規定)
クロスソース (相互調達)	自社／提携先 (47条の7-適用を期待)		自社／ 提携先	自社 or 提携先 or 共有 (契約で規定)
コモンソース (共同調達)	作成者(コミュニ ティメンバ)	①コミュニ ティ	②作成に関 わった者 (共同作成)	③作成に関 わった者 (規約で規定)
オープンソース (公開調達)	Creative Commons適用		Creative Commons適用 (稀にGPL)	

本報告では、「**コモンソース**」における論点を整理する

AIの社会実装が期待されている

例えば、製造業

社会環境

- 少子高齢化の進展
 - ② 労働人口減少
 - ② 技術継承問題
- 日本の製造業の国際競争力低下
 - ② 大量生産の限界
 - ② 各企業で閉じた生産システム
- 環境負荷の低減
- 進展する第4次産業革命

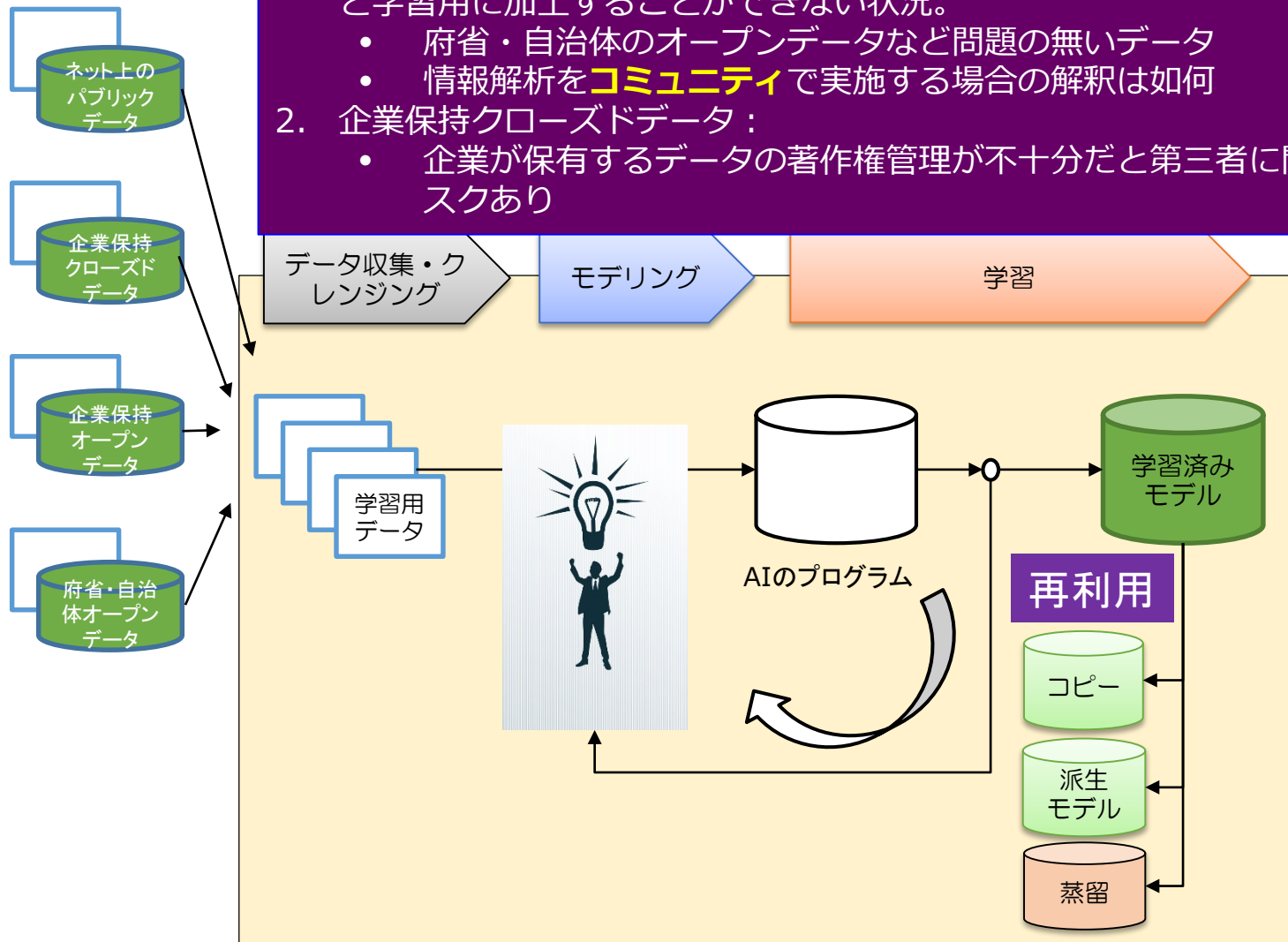
AIへの期待

- 人、ロボット、AIが協調する社会の実現
- 生産性の飛躍的な向上
- 柔軟で少量生産可能な生産システムの実現
- 産業の活性化、国際競争力の向上

しかし日本では、「データと研究者、技術者が分断されているため、大学などで開発された技術と、社会における現実的な問題とがうまく繋がっていない」*
⇒ コミュニティが「橋渡し」を加速する

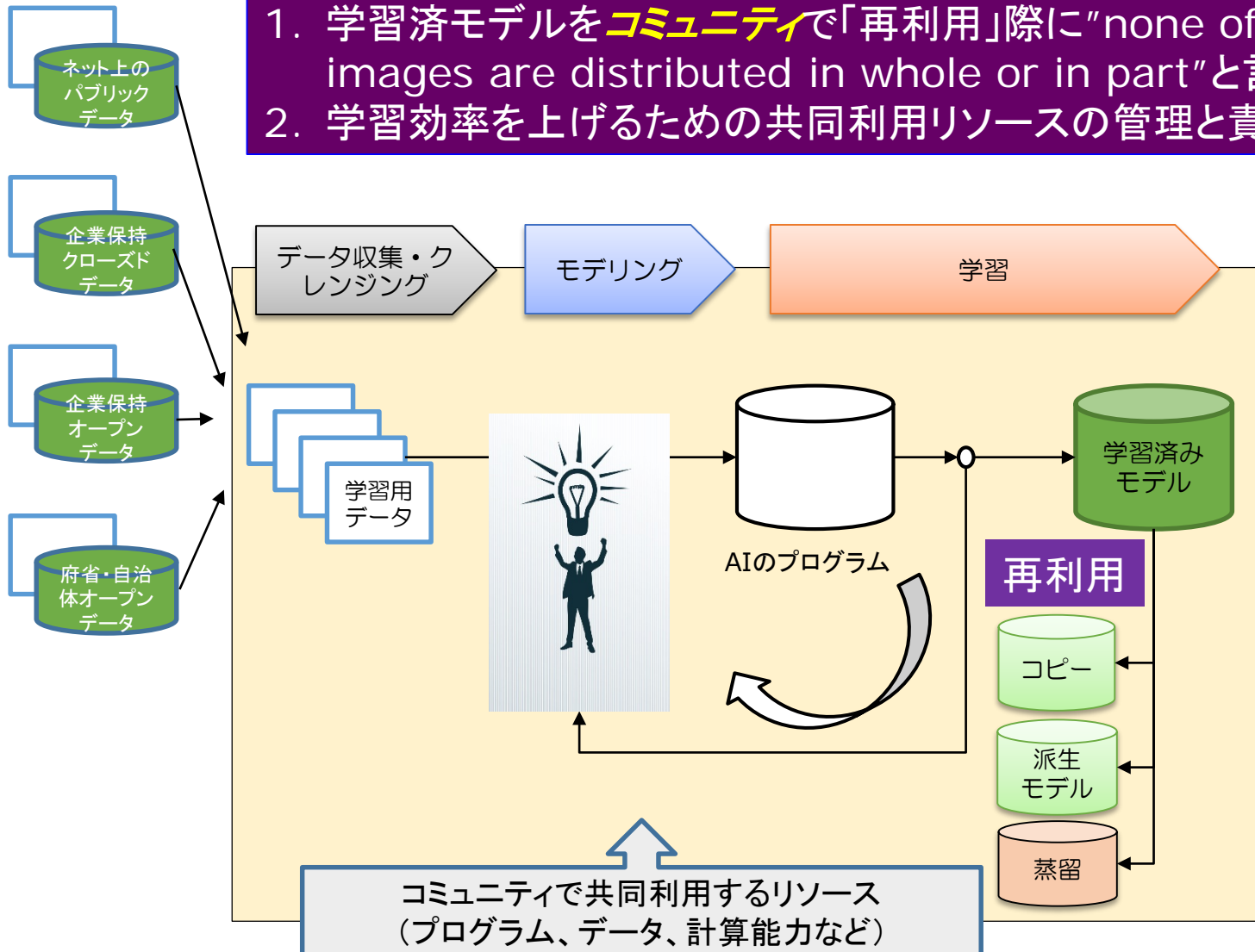
論点①：作成した学習用データセットの共同利用 元データ(コンテンツ)へのアクセス・著作権

1. (著47条7の適用範囲を超える) コミュニティでは権利がクリアにならないと学習用に加工することができない状況。
 - 府省・自治体のオープンデータなど問題の無いデータ
 - 情報解析を**コミュニティ**で実施する場合の解釈は如何
2. 企業保持クローズドデータ：
 - 企業が保有するデータの著作権管理が不十分だと第三者に開示するリスクあり



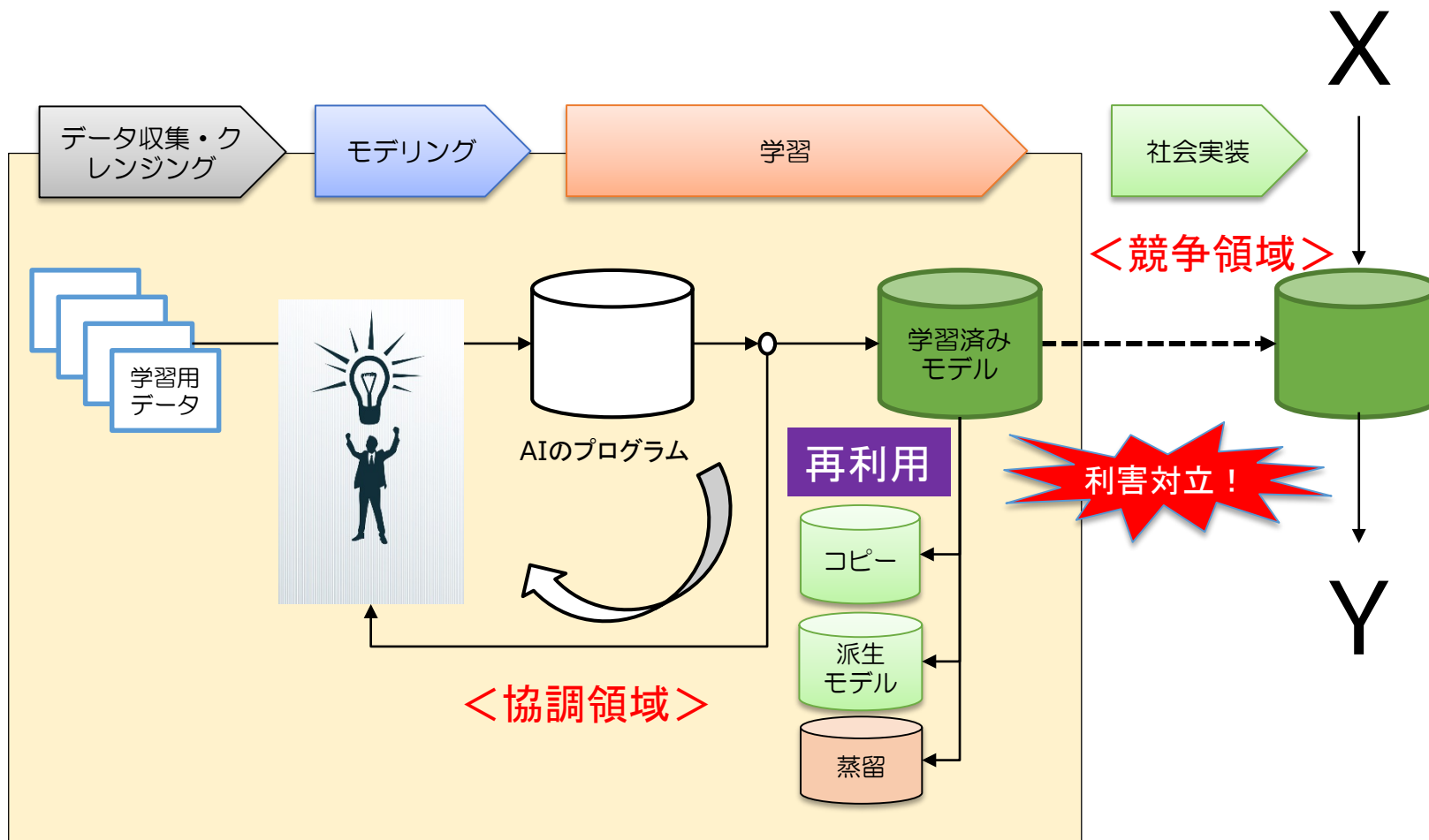
論点②：学習済モデルの再利用 コミュニティで「再利用」して、学習効率を上げたい

1. 学習済モデルを**コミュニティ**で「再利用」際に“none of the original images are distributed in whole or in part”と言えるか
2. 学習効率を上げるための共同利用リソースの管理と責任の制限



論点③：学習済モデルの権利

再利用<協調領域> vs. 社会実装<競争領域>



1. 再利用: 著作権がなければ再利用は促進される。
2. 社会実装: 企業としては学習済モデルの権利を確保したい。

コミュニティにおける情報財の整理

- 1) **学習用データセット**：著作権47条の7（情報解析のための複製等）、30条の4（技術の開発又は実用化のための試験の用に供するための利用）
- 2) **学習済みモデル**：著作権或いは、特許で保護されるものか？ 著作権で保護されるには、「創作性」が1つの要件。特許で保護されるには、「プログラム等」として認められることが1つの要件。
学習済みモデルは、その他に、「営業秘密」として不正競争防止法で保護する場合と、民法上の「不法行為法」で保護される（例えば、その作成に多大な労力を要した場合）場合が考えられる。
- 3) Caffe/Chainerなどの**フレームワーク**は、プログラム（著作権が適用される）であり、オープンソースの各種ライセンスが適用されている(MIT, Apache, BSD, LGPLなど)。
- 4) **学習済みモデルが「プログラム等」と認められれば**、これらのライセンスにより、共同利用を図ることができる。
- 5) また、学習済みモデルが「著作権を有する」データ（コンテンツ）を元に作成された場合、データの著作権者が学習済みモデルに対して権利を主張する可能性があるため、共同利用にリスクが発生する。これに対して、Caffeは「自己責任で使え」と開き直っている。

AI社会実装を加速するための論点

協調領域と競争領域のバランスの取れた議論が必要

AI学習効率＜協調領域＞

- ① 著作権の制約なく研究目的で、「学習用データセット」を作成／共同利用したい。
(47条の7を拡張し、データセットをコミュニティで利用できるか?)
- ② 「学習済モデル」を研究目的で、自由に再利用したい。
- 多種多様な生データ、学習済モデルを利用できる「環境」が欲しい。

企業収益＜競争領域＞

- ③ 学習済モデルを社会実装して、「収益」を上げたい。
- AI社会実装によって、グローバル「競争」に勝ちたい。
- 学習済モデルは「権利化」したい。



協調領域をコミュニティベースで推進するため、これを支援する環境と育成

まとめ

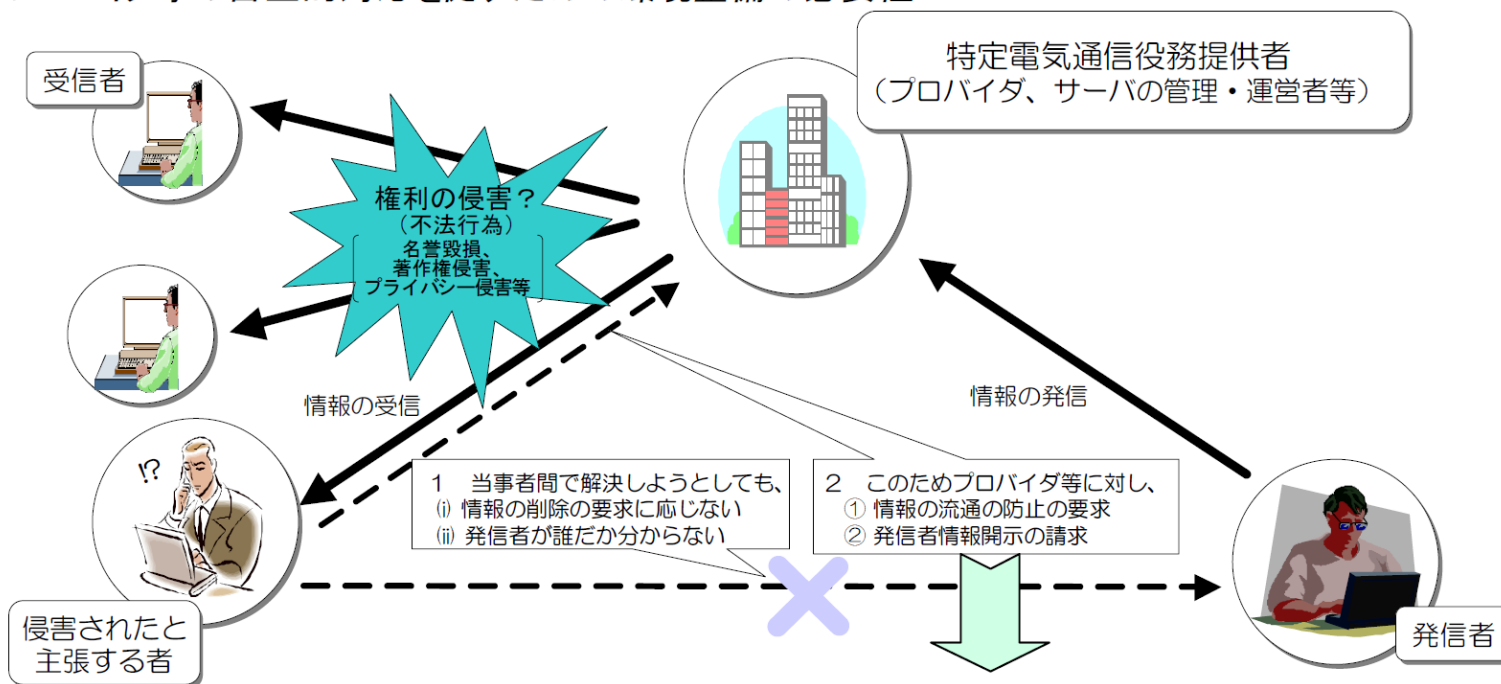
AI社会実装を拡大、加速するには、「協調領域」(研究開発)と「競争領域」(社会実装)のバランスの取れた議論が必要である。

- 本報告では、AI学習をコミュニティで実施する場合〈協調領域〉の新しい情報財の論点を整理した。
 - ① 学習用データセットの共同利用
 - ② 学習済モデルの再利用
 - ③ 学習済モデルの権利
- 「競争領域」(社会実装)については、別の視点からの議論が必要

(参考) プロバイダ責任制限法

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
(プロバイダ責任制限法)

プロバイダ等の自主的対応を促すための環境整備の必要性



- (a) 情報の違法性の判断が困難等自主対応による措置の責任が不明確な場合がある
- (b) 民事事件ではほとんど発信者情報の開示はされず、被害者救済が困難なことがある

➡ プロバイダ等による自主的対応を促し、その実効性を高める環境整備の必要