

# データ流通促進の取り組み



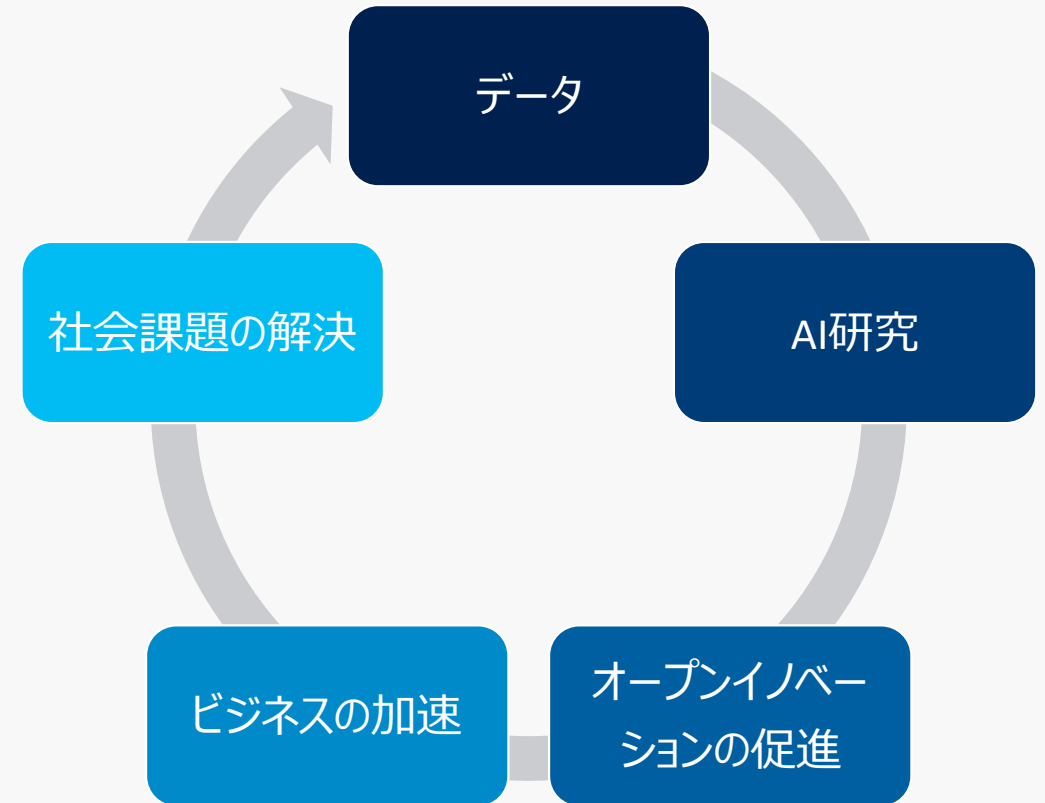
**AI Data Consortium**

AIデータ活用コンソーシアム

# 当初の課題意識： AIとデータを取り巻く課題

細分化されたデータ提供者、異なるライセンスの考え方、個人情報、計算リソースとの連携など、AI研究、オープンイノベーション、そしてソリューション化（商用化）には解決すべき多くの課題が存在しています。

- データ提供、データ活用における契約プロセスの標準化とデータ共有基盤の構築
- 個人情報・プライバシーに配慮したデータ活用モデルの実現
- 多種多様なデータのストア、データの商取引の実現により、持続可能なデータ活用基盤を実現



**円滑データの流通の実現を通して社会課題の解決を促進**

# 直面する様々な課題



知的財産



製造物責任



倫理・公平性



変化する価値



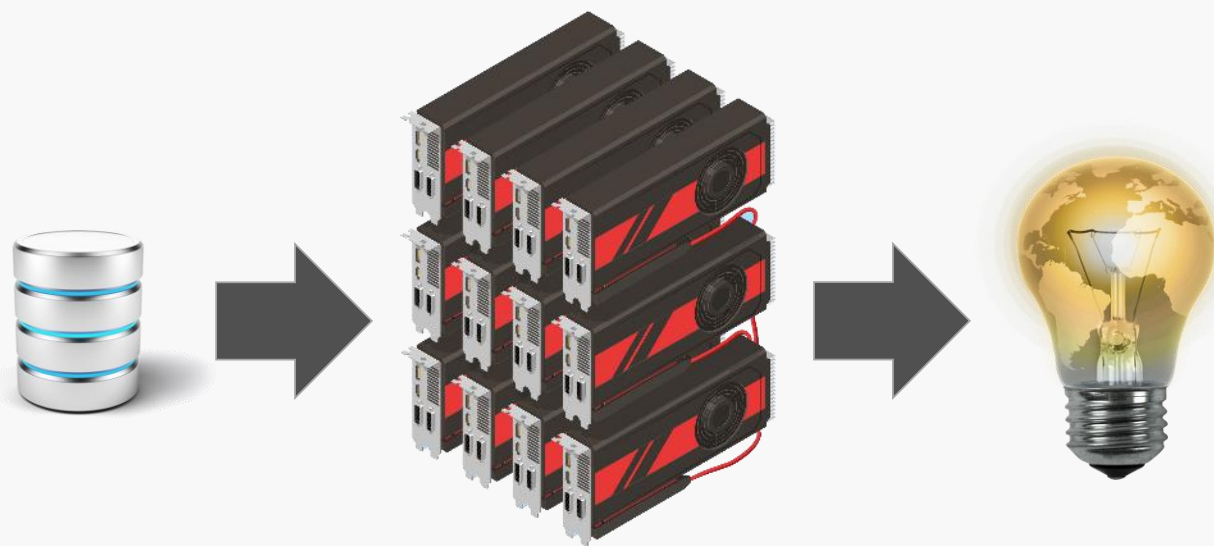
統合データ活用

# データが持つ異なる性質

- 従来のデータ活用は、分析、可視化の完了と共に役目を終え、データが副次的な経済的価値を持つ事は稀である。
- AI・機械学習では学習データが形を変え永続し、新たな価値を創出する。技術特許同様の商流に応じた契約内容、価格の交渉と締結が必要。



使い捨てられるデータ



永続するデータ

# ソフトウェア・プログラムと製造物責任

**（製造物責任法第二条）** この法律において「製造物」とは、製造又は加工された動産をいう。

## 消費者庁「逐条解説 製造物責任法」

ソフトウェア自体については、無体物であり、製造物責任の対象とはしていない。ただし、**ソフトウェアを組み込んだ製造物については、本法の対象と解される場合がありうる。**ソフトウェアの不具合が原因でソフトウェアを組み込んだ製造物による事故が発生した場合、**ソフトウェアの不具合が当該製造物自体の欠陥と解されることがあり、**この場合、その欠陥と損害との間に**因果関係**が認められるときには、当該製造物の製造業者に本法に基づく**損害賠償責任**が生ずる

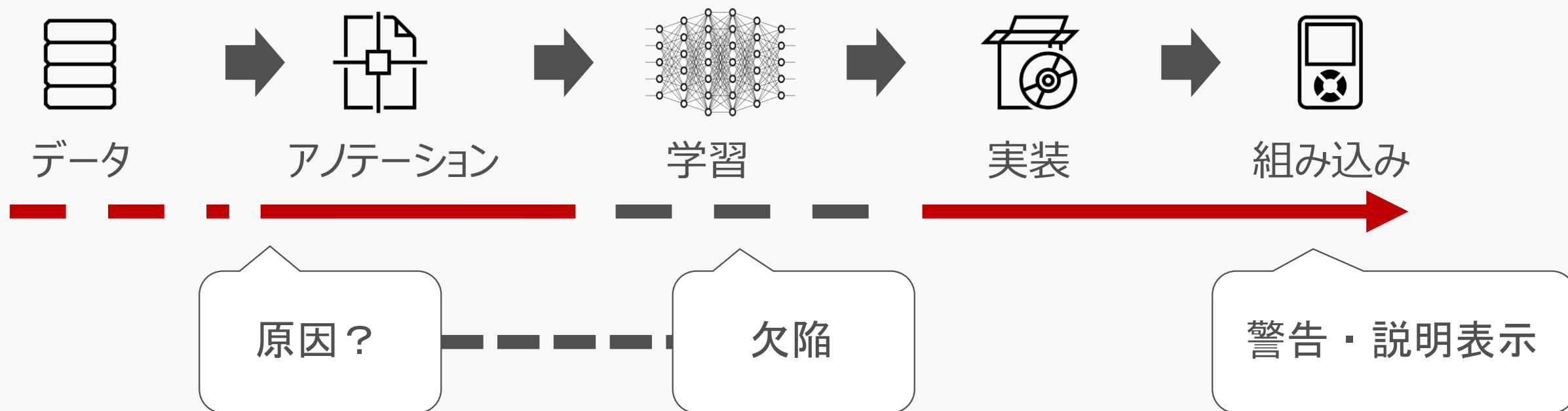
# 明確な製造物と責任

## ソフトウェア開発と責任



# 想定される責任範囲

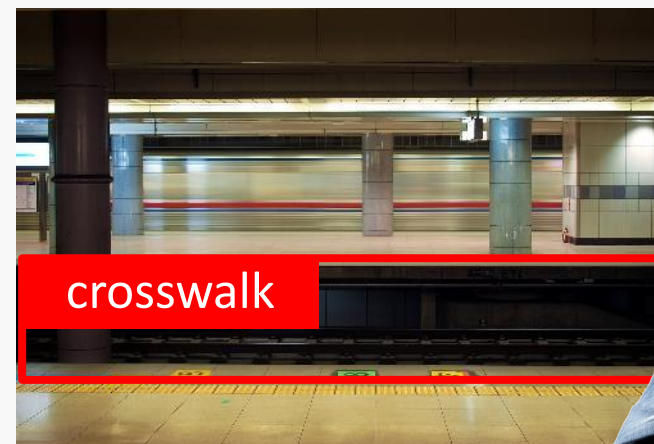
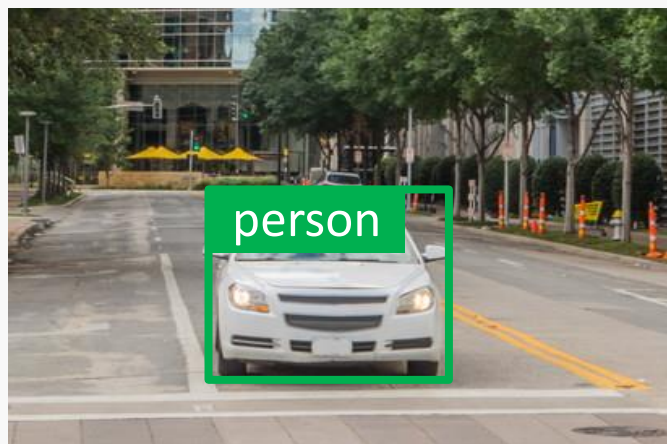
## AIを用いたシステムと責任





# 課題となるデータの確からしさとAI品質の関係

- 目的、アルゴリズムによって異なる様々なアノテーションの品質と基準
- 排除が難しい「誤り」と「悪意」によるアノテーション



重要性を増す作業者の**身元確認**と**保証**





# AIの用途 x 保証 x データ

## レベル 1



AIを使用することによる影響が想定可能、且つ使用者に限定され許容可能である。

## レベル 2



AIを使用することによる影響が想定可能であり、補償もしくは回復可能である。

## レベル 3



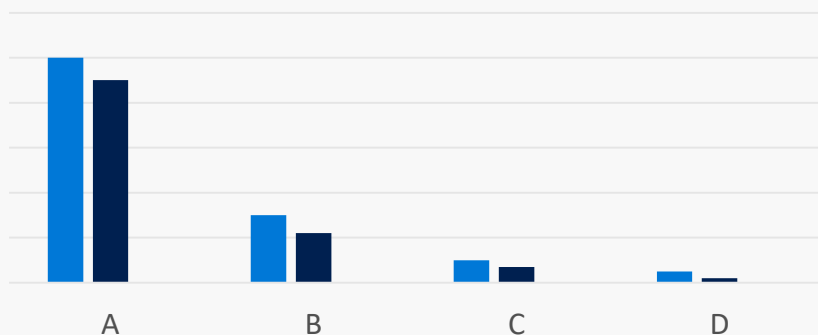
AIを使用することにより身体、社会権に影響を及ぼす可能性がある、若しくは補償、回復が困難である。

用途により異なるデータに求める保証レベル

# 重要性を増すAIの品質とデータの偏り（バイアス）の問題

データの偏りがAIの品質の偏りとして現れる  
AIを適用する対象問題によっては、学習データのバランス、透明性は倫理上の問題に直結する。

## データ量と品質



分布、ドメイン、品質保証とそのレベルより異なるデータの価値

## The New York Times *Facial Recognition Is Accurate, if You're a White Guy*

By STEVE LOHR FEB. 9, 2018



Gender was misidentified in up to 1 percent of lighter-skinned males in a set of 385 photos.

Gender was misidentified in up to 12 percent of darker-skinned males in a set of 318 photos.

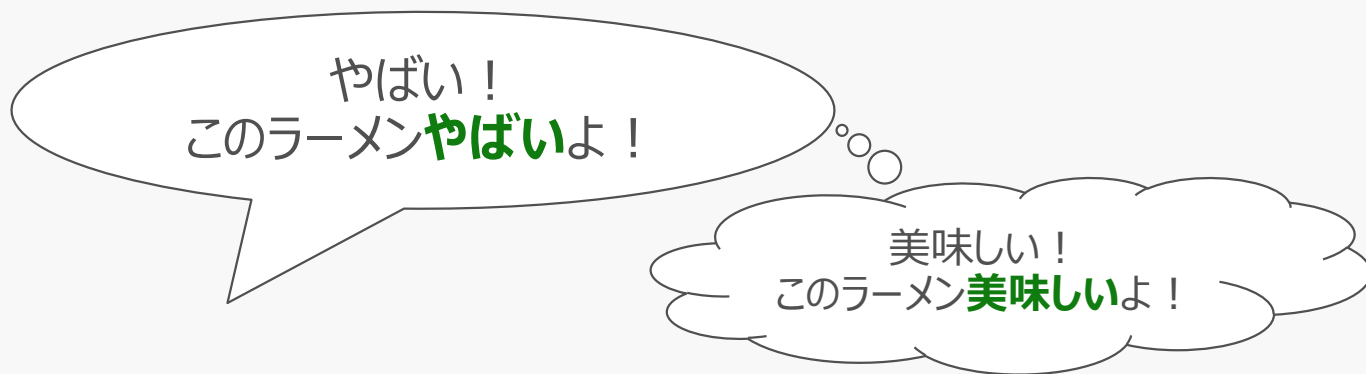


Gender was misidentified in up to 7 percent of lighter-skinned females in a set of 296 photos.

Gender was misidentified in 35 percent of darker-skinned females in a set of 271 photos.

# 変化するAI品質とデータの価値

- 時間と共に増える、変化する言葉
- 世代により意味を異にする言葉



時間と共に**変化する**データの価値

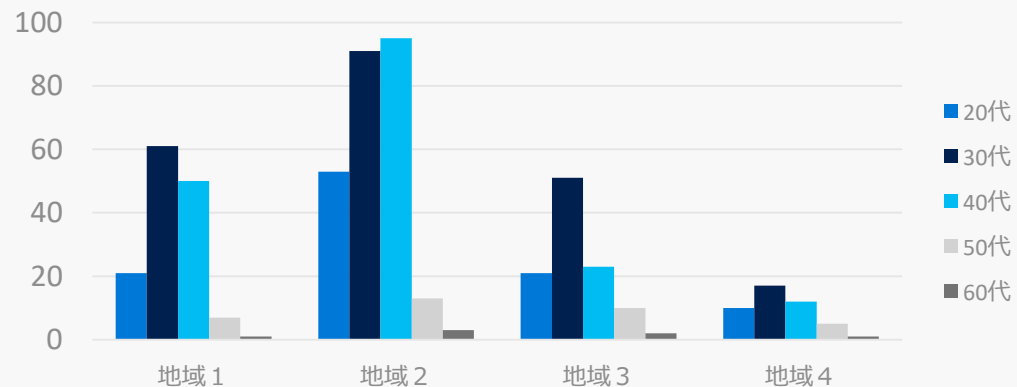
順位	2018年	2019年
1	マジ卍	り
2	すこ	パリピ
3	彼ピッピ	夕ピる
4	バズる	とりま
5	ワンチャン	あね
6	パリピ	それな
7	ズッ友	フロリダ
8	草	おけまる
9	ぷちよへんざ	乙
10	好きピ	草



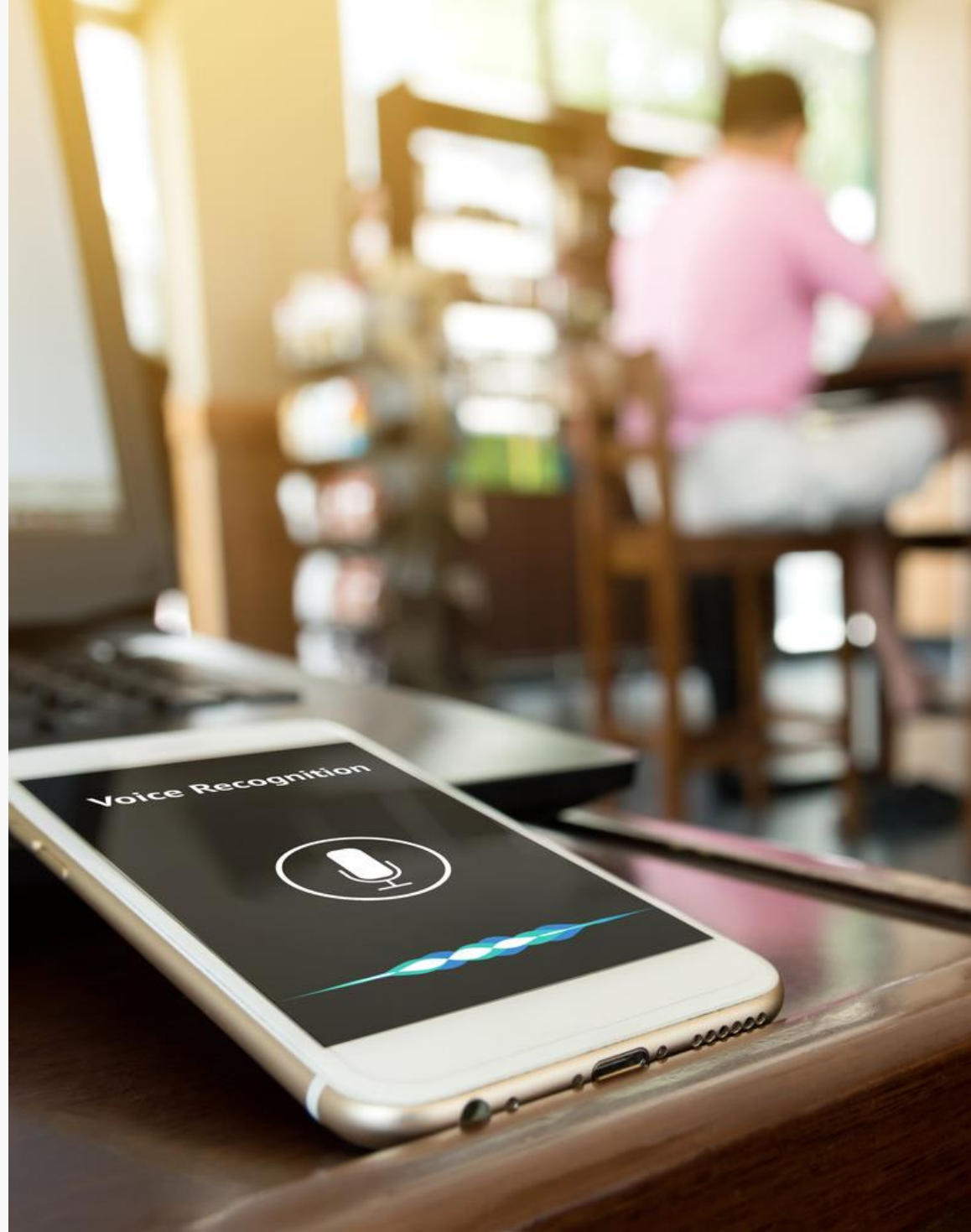
# 様々なデータの偏り

- 例) 地域差、世代差の大きい音声、会話データ
  - 標準語、20代~40代
  - 地方、高齢者
- 少子高齢化、Society 5.0で重要性を増す高齢者、障がい者をサポートするAIの実現と活用
- **相反する限られたデータ**

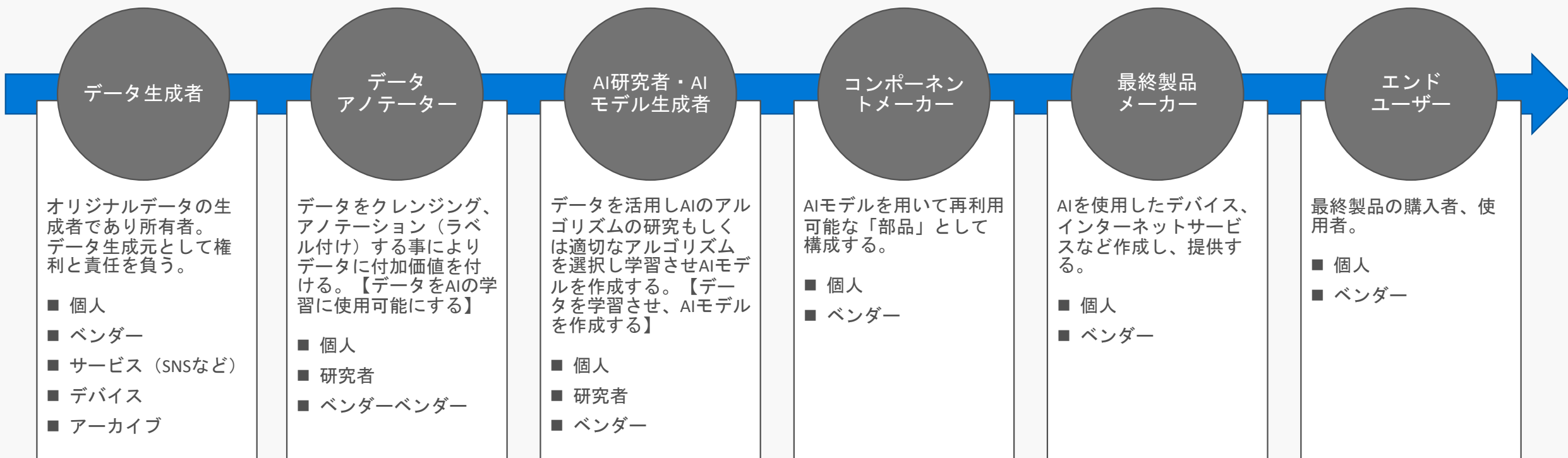
例：音声データ分布



**ニーズ**により異なるデータの価値

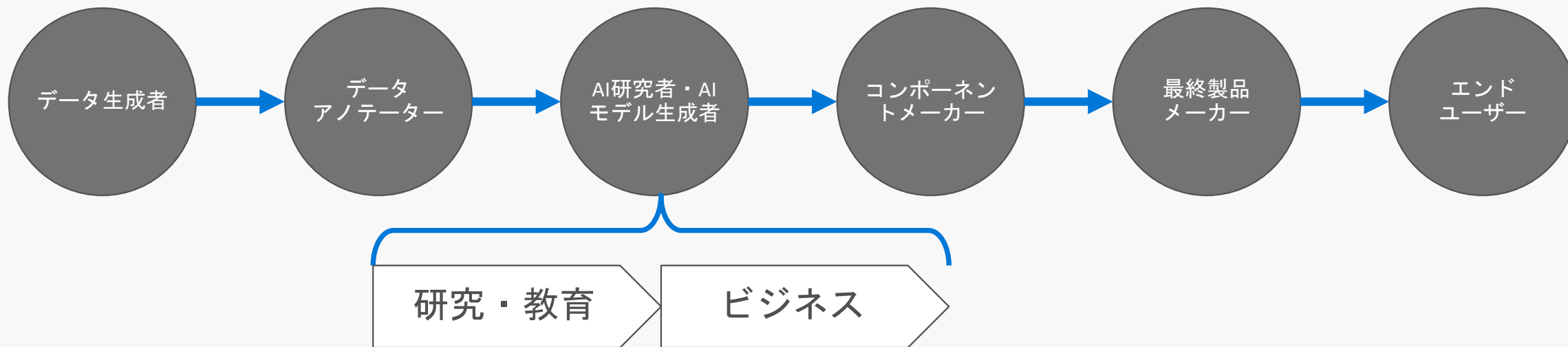


# AIにおけるデータの役割

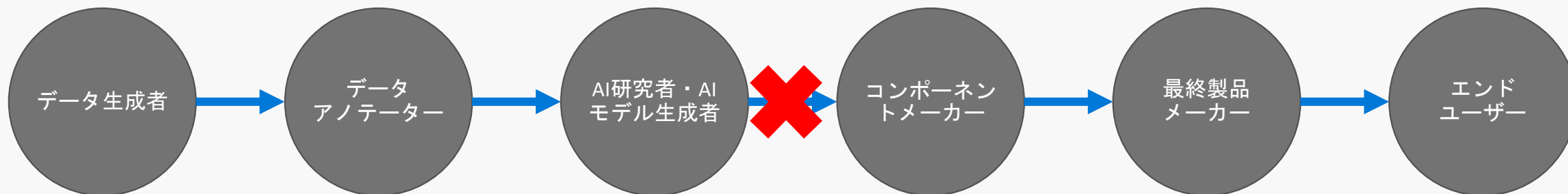


# 想定される商流（例）

## 活用フェーズによる変化する使用条件



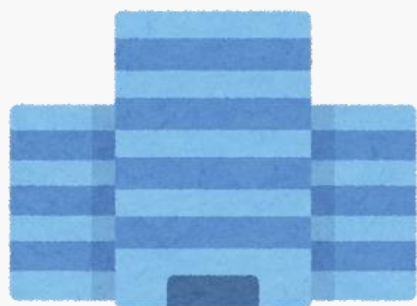
AIモデルの品質を向上できない場合、製品化を断念する場合がある。





# データ提供者の気付き

- データが使い捨てではなく、形を変え永続し、価値を創出し続ける事への気付き
- 従来とは全く異なる契約モデルが必要であることへの気付き



データ所有者

あれ？  
我が社のデータには  
もっと高い価値が  
ある？



① 従来のデータ提供、販売を想定

② データの用途（AI）を説明

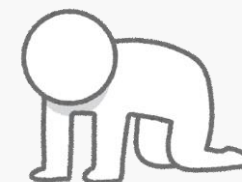
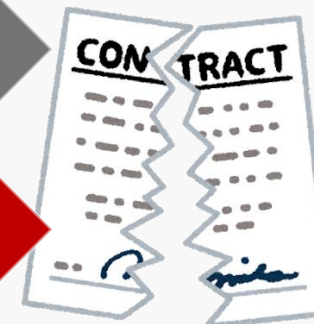
③ データが使い捨てではなく、付加価値を持つ事を把握

④ 適切な契約モデルの欠如、知財契約知識の不足

⑤ データを提供・販売しないことを決定

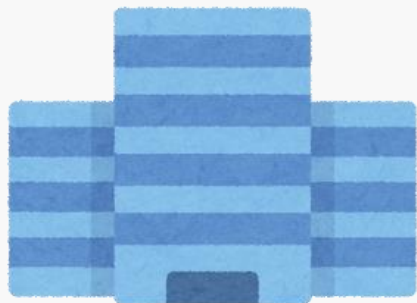


AI研究者  
スタートアップ



# データ利用者、AI開発者の気付き

- 製造物責任への対応として、学習用データの品質、トレーサビリティが必要であることへの気付き
- 取得済み、アノテーション済みデータの来歴情報が必要であることへの気付き



データ所有者

① AIをシステムに組み込む事を前提に、データの保証レベルを確認

② データ取得者、アノテーション作業者が特定できない  
(説明責任を果たせない)

③ 製造物責任およびその所在を担保できず、当該データの使用を断念



AI、システム  
ベンダー



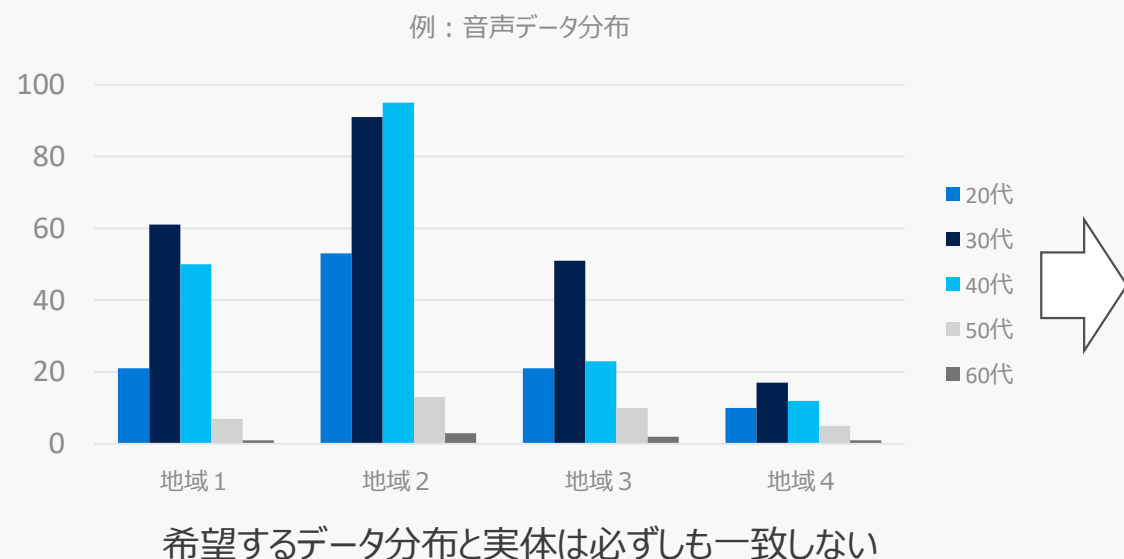
取得元、取得方法を説明、保証できるデータが見つからない

# AI用のデータ流通に求められる複雑な要件

## 学習用データ取り引における主な要素

- データの取得元・取得方法の情報（来歴）
- データの分布（偏り、ドメイン）の情報
- データの品質（評価方法、評価結果）
- データの鮮度（時間の経過と共に価値が変動するデータ）
- データの利用目的、フェーズに応じた価格・契約体系（教育・研究・評価・商用）
- 動的データ、静的データ
- （保証可能範囲）

様々な条件により契約、価格  
およびデータ流通に求められる  
仕様は異なる。



## 求められるデータホルダー、データ種別ごとの柔軟なライセンス形態

- 同種のデータであっても、情報ドメインによって取得、生成コストは異なる
- 難しいモデル生成で求められる倫理的配慮とデータのバランス

(例)

### 選択的データアクセス

要とするドメインのデータへの  
選択的アクセスを可能にする  
ことで、データ利用者のニーズ  
に応える。

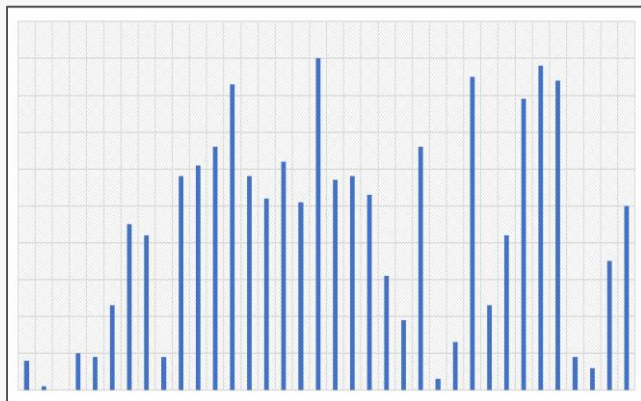
### アクセスボリューム

可変的なデータボリュームのサ  
ポート。データボリューム（サイ  
ズ、数など）による提供、取  
得を可能に。

### データの鮮度・更新

陳腐化するデータ、更新され  
るデータへの対応（更新イベ  
ント、データ同期ニーズへの対  
応。例：交通情報など）

# これまでと大きく異なるデータ取引要件



## データドメイン

データ分布とドメインの選択



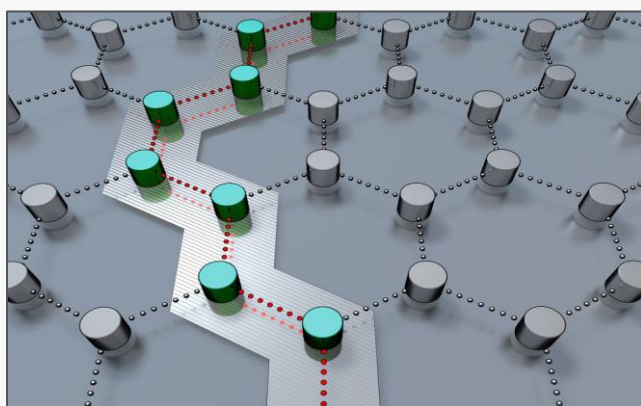
## 知財

権利関係の整理



## 商流

商流により異なる価格と契約



## 来歴

説明可能性とトレーサビリティ



## 品質

データ品質と保証



## 契約

来歴、品質、知財など複雑に絡み合う契約内容

# コンソーシアムの取り組み





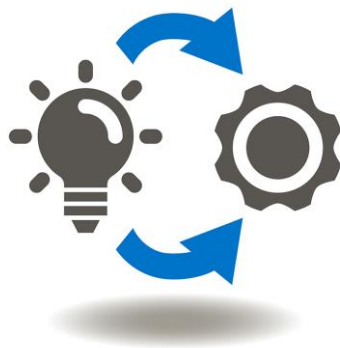
# AI Data Consortium

AIデータ活用コンソーシアム



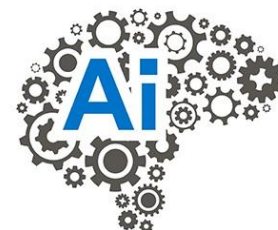
## 知的財産・契約

形を変えデータが永続するAIでは従来と異なる取引、契約の考え方が必要です。スマートコントラクトを用いた柔軟なデータ取引の実現を目指します。



## AI研究

コンソーシアムが取り組むデータ収集・流通を通して、国内の大学・研究機関等と連携し社会課題の解決に資するAI研究を加速していきます。



## データ収集・活用

領域を問わず様々なデータをコンソーシアムが提供するデータ基盤を通じて提供するとともに、AIによる社会課題の解決に貢献します。



## データ基盤

様々な契約条件、基準、セキュリティ要件に対応した柔軟なデータ流通基盤を提供し、重要性を増すデータ統制に対応します。

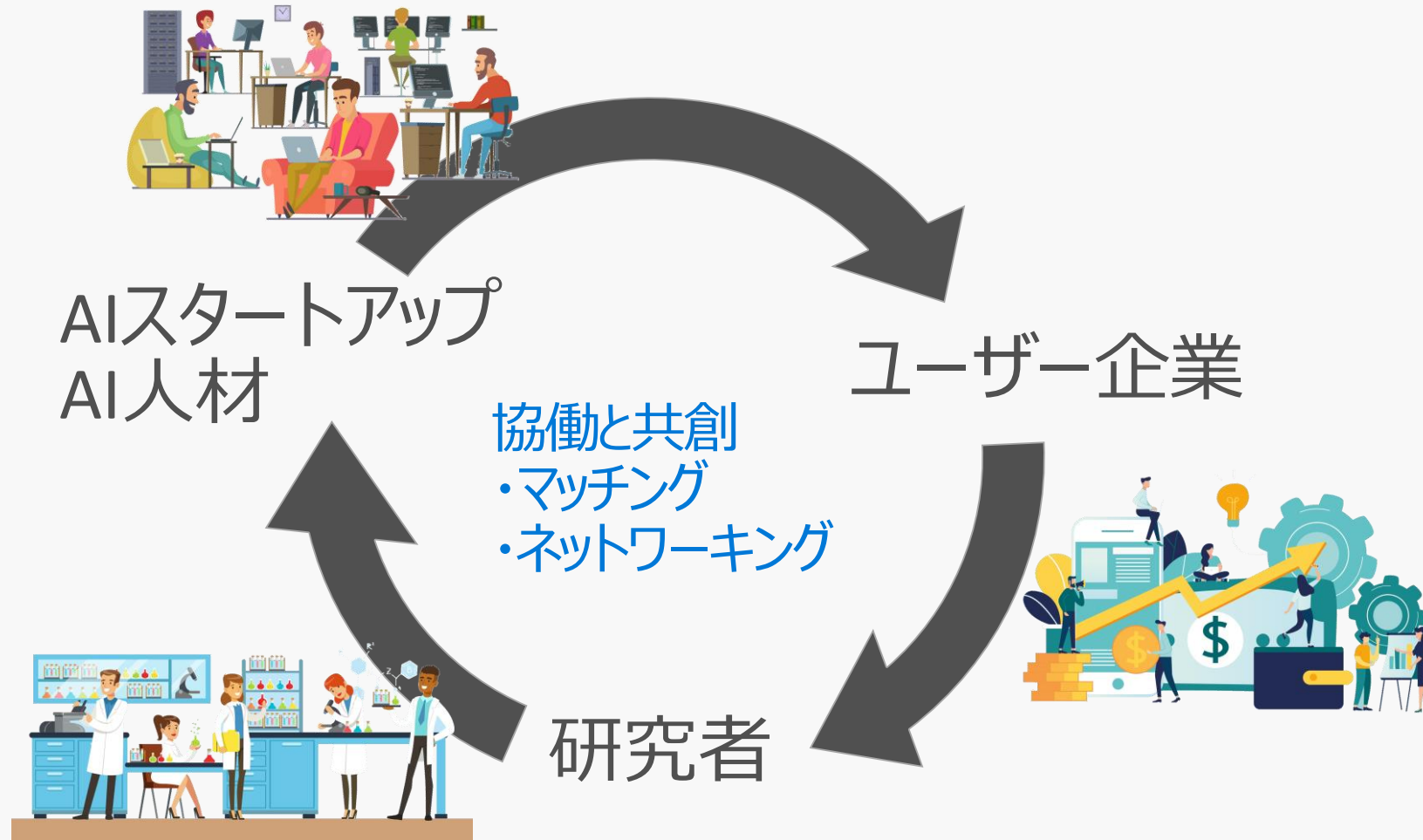
<https://aidata.or.jp>



# 目指す姿とチャレンジ①： データ利用・提供に係る契約手続きを媒介し、手続きを簡素化



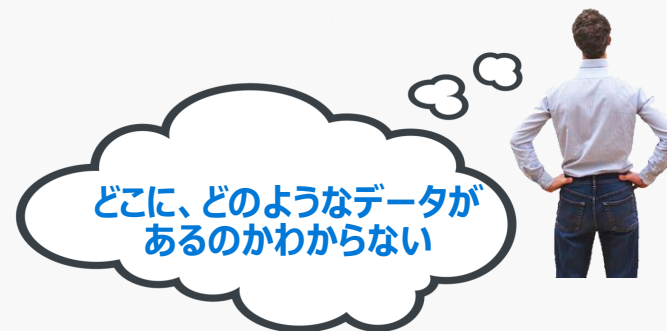
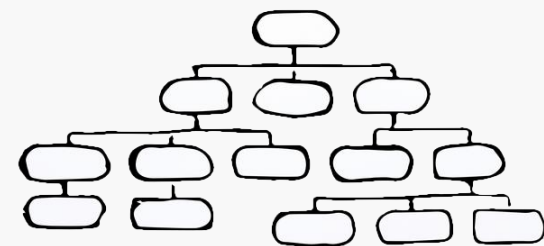
# 目指す姿とチャレンジ②： 会員間連携による課題解決（マッチングとネットワーキング）



# 目指す姿とチャレンジ③： 組織内データ基盤構築の実現

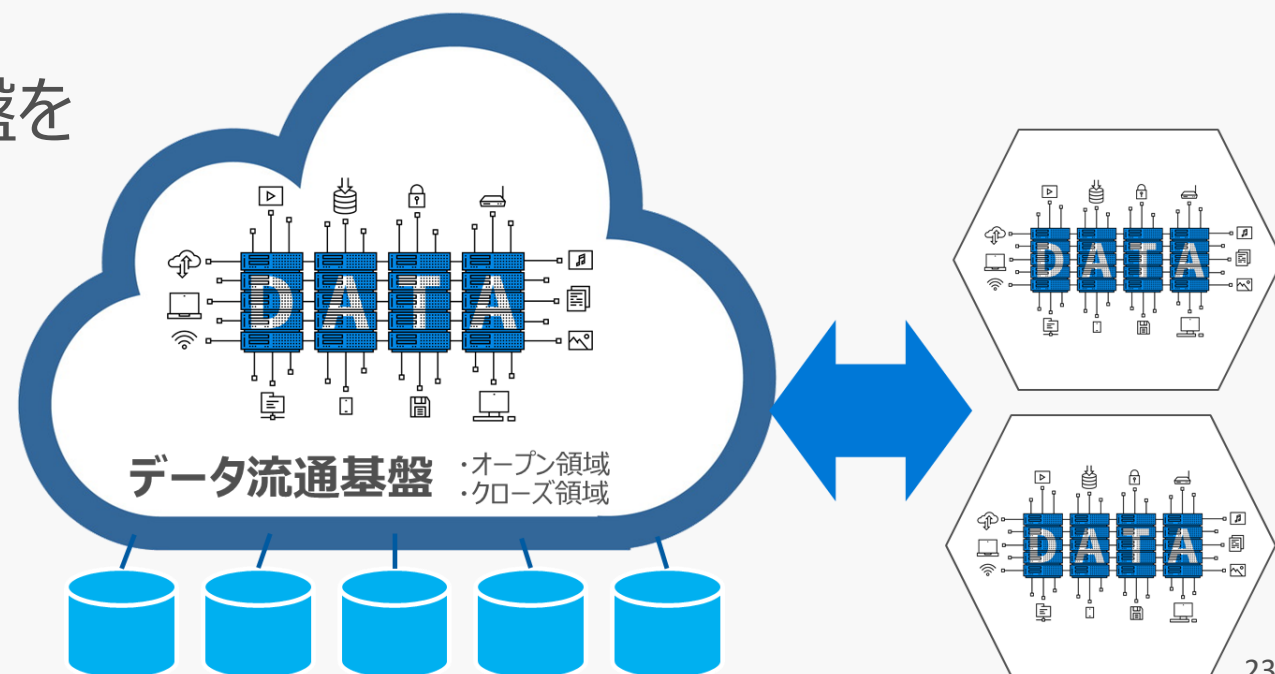
## • 重要性を増す組織内データ流通、共有基盤の構築

- データの可視化（組織内データカタログ）
- 煩雑な手続き（権利・知財・契約・セキュリティ）
- 多様なデータへの対応  
（構造化・非構造化データ、ストリーミングデータ、他）

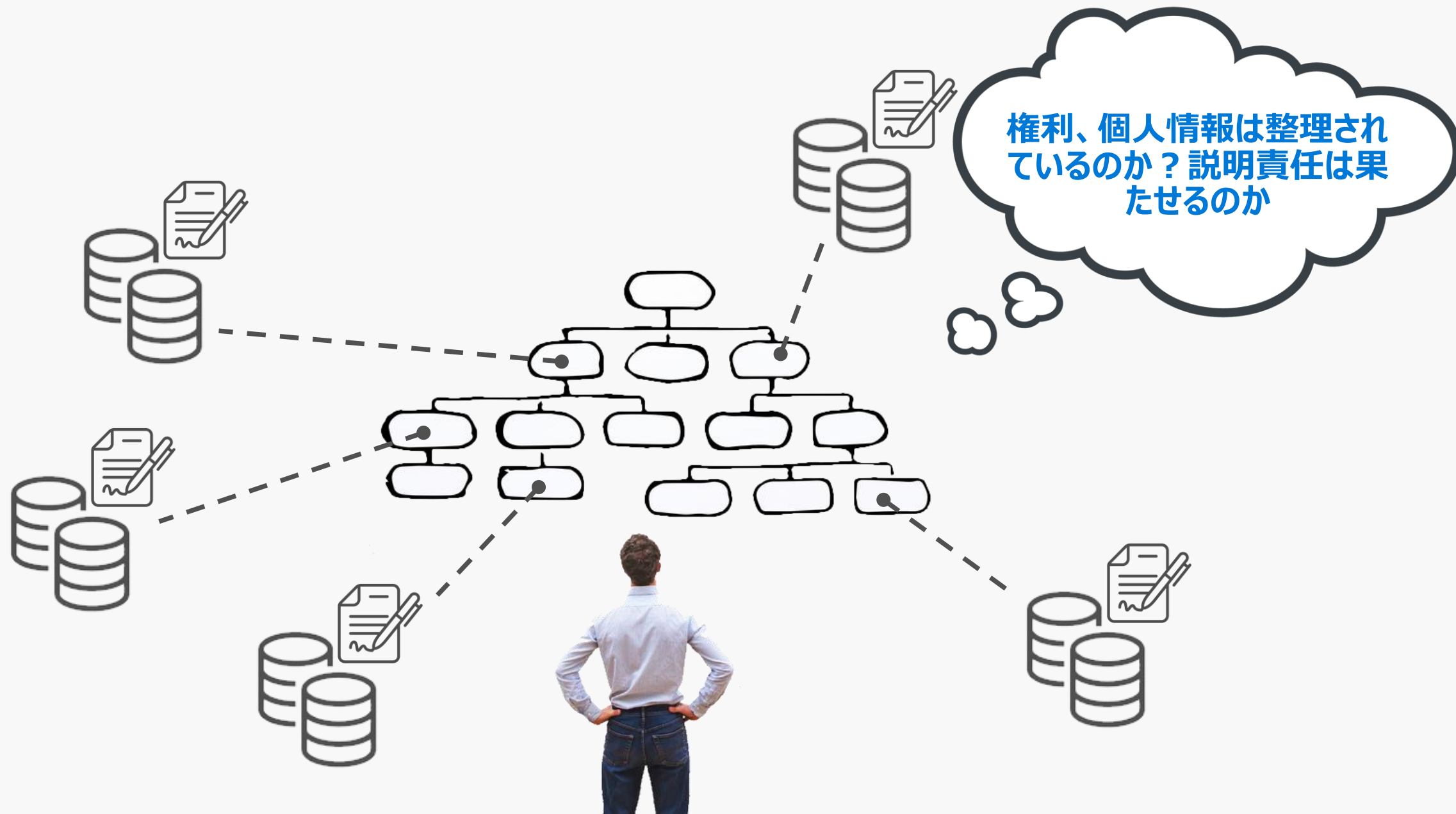


## 『AIデータ活用コンソーシアム』のデータ基盤を 組織内システムとして転用支援

- データ基盤構築に要するノウハウの活用
- 開発コストの低減



# 外部データ活用の統制がより重要に



# 目指す姿とチャレンジ④： ニーズに応じたAIデータ収集・提供のモデル

## 媒介のみの提供モデル

## データ管理型の提供モデル

データ利用者

データ利用者

Web API

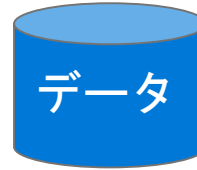
Web API

### AIデータ活用基盤

- 契約手続・管理
- メータリング・課金
- 認証・セキュリティ
- テクニカルQA等

### AIデータ活用基盤

- データ管理
- 契約手続・管理
- メータリング・課金
- セキュリティ
- 認証・セキュリティ
- テクニカルQA等



Web API

Web API

Web API

### データホルダー

- データホルダーによるデータのメンテナンス



### データホルダー

- AIデータ活用基盤をデータストアとして使用
- データの収集、更新の責のみを負う



**AI Data Consortium**

AIデータ活用コンソーシアム

AIデータ活用コンソーシアムが  
構築、提供するサービス



# 理事・協働メンバー

## 会長

- ・ **長尾 真**  
元京都大学総長 / 元独立行政法人情報通信研究機構理事長  
/ 前国立国会図書館長 / 前国際高等研究所所長

## 副会長

- ・ **渡部 俊也**  
東京大学 未来ビジョン研究センター 教授/副センター長 大学執行役・副学長 / 工学系研究科技術経営戦略学専攻 教授
- ・ **杉山 将**  
理化学研究所 革新知能統合研究センター センター長 / 東京大学 大学院新領域創成科学研究科 教授
- ・ **井佐原 均**  
豊橋技術科学大学情報メディア基盤センター センター長・教授
- ・ **田丸 健三郎**  
日本マイクロソフト株式会社 業務執行役員 NTO

## 理事

- ・ **井佐原 均**  
豊橋技術科学大学情報メディア基盤センター センター長・教授
- ・ **石井 靖乃**  
日本財団 ソーシャルイノベーション本部 公益事業部 部長
- ・ **小野 雄次郎**  
一般社団法人 日本支援技術協会 代表理事
- ・ **越塚 登**  
東京大学大学院 情報学環・学際情報学府 情報学環長 教授
- ・ **坂村 健**  
東洋大学情報連携学部INIAD学部長 / cHUB (学術実業連携機構) 機構長 / YRPユビキタス・ネットワークング研究所 所長 / 公共交通オープンデータ協議会 会長

- ・ **下川 和男**  
一般社団法人 日本電子出版協会 副会長 / イースト株式会社 取締役会長
- ・ **杉山 将**  
理化学研究所 革新知能統合研究センター センター長 / 東京大学 大学院新領域創成科学研究科 教授
- ・ **田丸 健三郎**  
日本マイクロソフト株式会社 業務執行役員 NTO
- ・ **長尾 真**  
元京都大学総長 / 元独立行政法人情報通信研究機構理事長  
/ 前国立国会図書館長 / 前国際高等研究所所長
- ・ **藤原 洋**  
株式会社ブロードバンドタワー 代表取締役 会長兼社長 CEO
- ・ **前川 喜久雄**  
国立国語研究所 音声言語研究領域 教授
- ・ **安岡 孝一**  
京都大学 人文科学研究所 教授
- ・ **柳原 尚史**  
株式会社Ridge-i 代表取締役社長
- ・ **吉川 健一**  
株式会社ブルックス 代表取締役社長
- ・ **渡部 俊也**  
東京大学 政策ビジョン研究センター 教授

## 団体

- ・ **石井 靖乃**  
日本財団 ソーシャルイノベーション本部 公益事業部 部長
- ・ **小野 雄次郎**  
一般社団法人 日本支援技術協会 代表理事
- ・ **下川 和男**  
一般社団法人 日本電子出版協会 副会長

## 協働

- ・ 公共交通オープンデータ協議会
- ・ 一般社団法人 日本支援技術協会
- ・ 一般社団法人 日本電子出版協会

## 協賛

- ・ Deep Learning Lab

## アドバイザー

- ・ 文部科学省 (MEXT)
- ・ 経済産業省 (METI)
- ・ 内閣府 (CAO)
- ・ 日本銀行 (BOJ)
- ・ 総務省 (MIC)
- ・ 国土交通省 (MLIT)
- ・ 宇宙航空研究開発機構 (JAXA)
- ・ 特許庁 (JPO)
- ・ 国立研究開発法人 科学技術振興機構 (JST)
- ・ 金融庁 (FSA)
- ・ 一般社団法人 日本特許情報機構 (JAPIO)





<https://aidata.or.jp>