

# AIとデータを取り巻く課題とソリューション 『AIDC Data Cloud』

日本マイクロソフト株式会社 業務執行役員

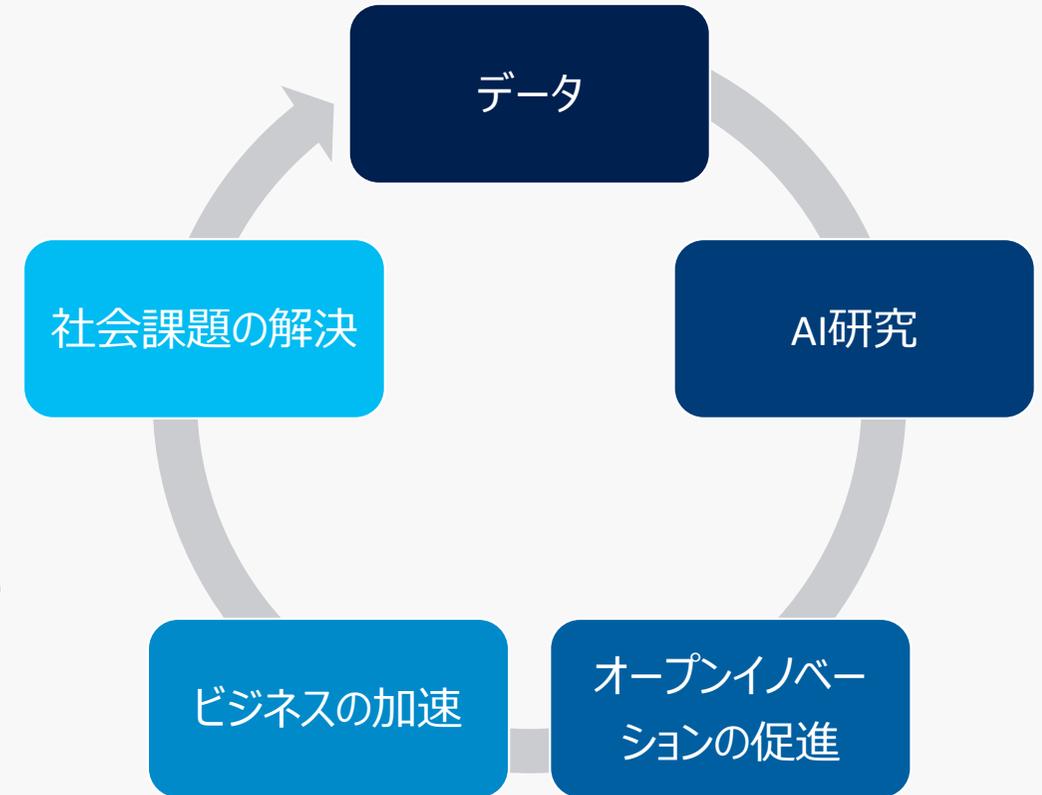
田丸 健三郎

<https://aidata.or.jp>

# 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**

# AIによる社会課題解決への貢献

様々なデータの提供、共有基盤の構築を通して、AI研究と社会課題の解決のための取り組みの接続を実現。



脳性麻痺、聴覚障害者の話す言葉を認識できるASR（Automatic Speech Recognition）エンジンは存在しない



支援者の多くは、発話困難な障害者の話す言葉には共通点があると感じている。



会話データの収集と最新のAI研究による発話困難な障害者のコミュニケーション、社会参画を支援

**例：障害者と健常者の円滑なコミュニケーションの実現**

# AIデータ活用コンソーシアム設立 2019年3月6日



## 発起人企業・団体

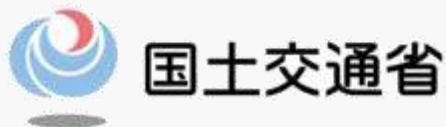
- 京都大学
- ギリア株式会社
- 国立国語研究所
- 一般社団法人 日本電子出版協会
- 東京大学
- 東洋大学
- 豊橋技術科学大学
- 日本財団
- 一般社団法人  
日本支援技術協会
- 日本マイクロソフト株式会社
- 株式会社ブリックス
- 株式会社ブロードバンドタワー
- 理化学研究所
- 株式会社Ridge-i

# 連携パートナー

知財・契約・データ

研究・ソリューション

プラットフォーム



- 文部科学省 (MEXT)
- 経済産業省 商務情報政策局 情報経済課 (METI)
- 日本銀行 (BOJ)
- 総務省 (MIC)
- 特許庁 (JPO)
- 国土交通省 (MLIT)
- 宇宙航空研究開発機構 (JAXA)
- 国立研究開発法人 科学技術振興機構 (JST)
- 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
- デジタルアーカイブ学会
- 東京大学
- 京都大学
- 理化学研究所
- 大阪大学
- 他

# メンバー



AI Data Consortium

AIデータ活用コンソーシアム

## 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)

# 不足するデータとAIの可能性

# 国内における課題（データ）

～ 圧倒的に不足日本固有のデータ～

- AI研究およびビジネス活用を前提とした業種を超えた多様なデータの共有、取引基盤の欠如
  - 先行する米国、中国
- 圧倒的に不足する日本固有のデータ（海外データによる偏ったAIモデル）
  - 多様性に富む日本語のデータ（自然言語データ）
    - 地域差、年代差、性差、業界差
    - 日常会話、ビジネス会話
    - 日々増え続ける固有名詞、変化し続ける用法
  - 日本文化固有の画像データ
    - 海外で作成された画像データセットは日本固有の物を含まない（茶碗、鳥居、だるま、赤べこ、けん玉）
  - 日本の地域固有データ
    - 道路交通、気象、地理、電力・水道、人口動態



ASEAN特許審査協力  
(ASPEC)



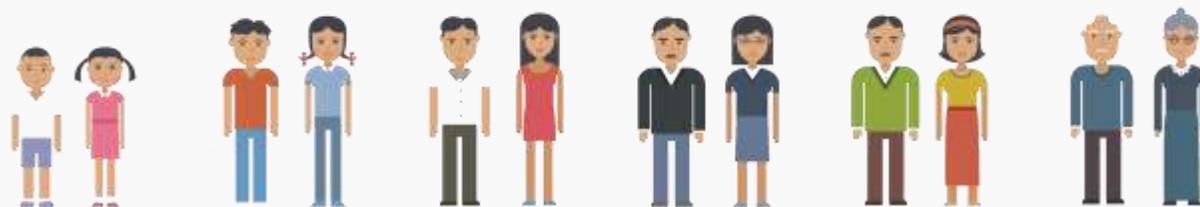
# 重要性を増す複数のモダリティによるモデル構築

- マルチモーダル深層学習で重要となる複数の異なるデータ
- ヘテロジニアスなデータによる複合モデルの構築
- 継続的な更新が必要なデータ

テキストデータ、音声データだけでは「やばい！」の意味を評価できない。行動データとの組み合わせで初めて「やばい！」の評価ができる。

やばい！  
このラーメンやばいよ！

美味しい！  
このラーメン美味しいよ！



変化し続ける言葉。10年前の10代と現在の10代では使用する言葉が大きく異なる。AIモデル品質の維持、向上には継続的なデータ収集、更新が欠かせない。

例：字面だけでは判断できない意味抽出によるイノベーション

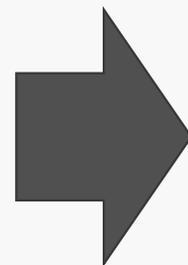
# ヘテロジニアス データとAIによる都市課題の解決

気象、交通、モバイル空間統計、GISなどの様々なデータによるモデル構築とサービスの提供を実現



東京都下水道局『東京アメッシュ』  
数分後の降雨状況の予測を可能に

<http://tokyo-ame.jwa.or.jp/>



都市の動態モデルによる人・物の予測

混雑前の移動（時間差移動）、混雑した経路の回避

**例：交通事業者横断的な最適化された運行スケジュールによる混雑緩和**

# AIによる社会課題解決への貢献

様々なデータの提供、共有基盤の構築を通して、AI研究と社会課題の解決のための取り組みの接続を実現。



脳性麻痺、聴覚障害者の話す言葉を認識できるASR（Automatic Speech Recognition）エンジンは存在しない



支援者の多くは、発話困難な障害者の話す言葉には共通点があると感じている。



会話データの収集と最新のAI研究による発話困難な障害者のコミュニケーション、社会参画を支援

**例：障害者と健常者の円滑なコミュニケーションの実現**

# 重要性増す責任 AIとデータの関係

# データ流通基盤に求められる様々な要素と課題



知的財産



製造物責任



倫理・公平性



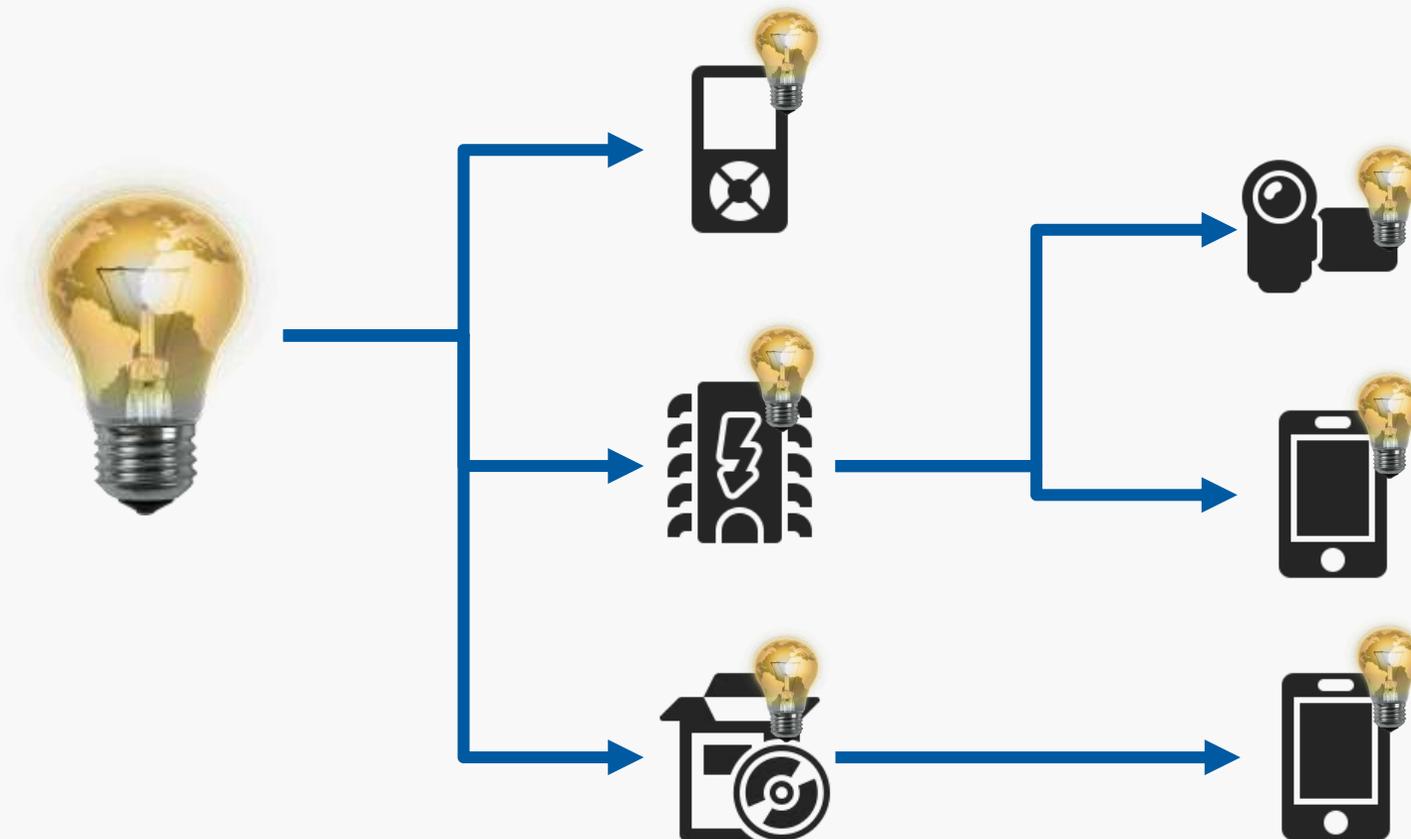
変化する価値



統合データ活用

# 再販とライセンス

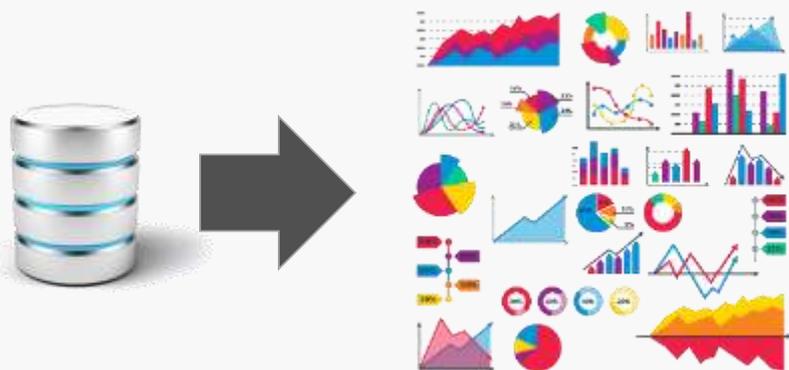
- 最終製品に組み込まれるテクノロジー（特許）
- 商流により異なるライセンス（契約、価格など）



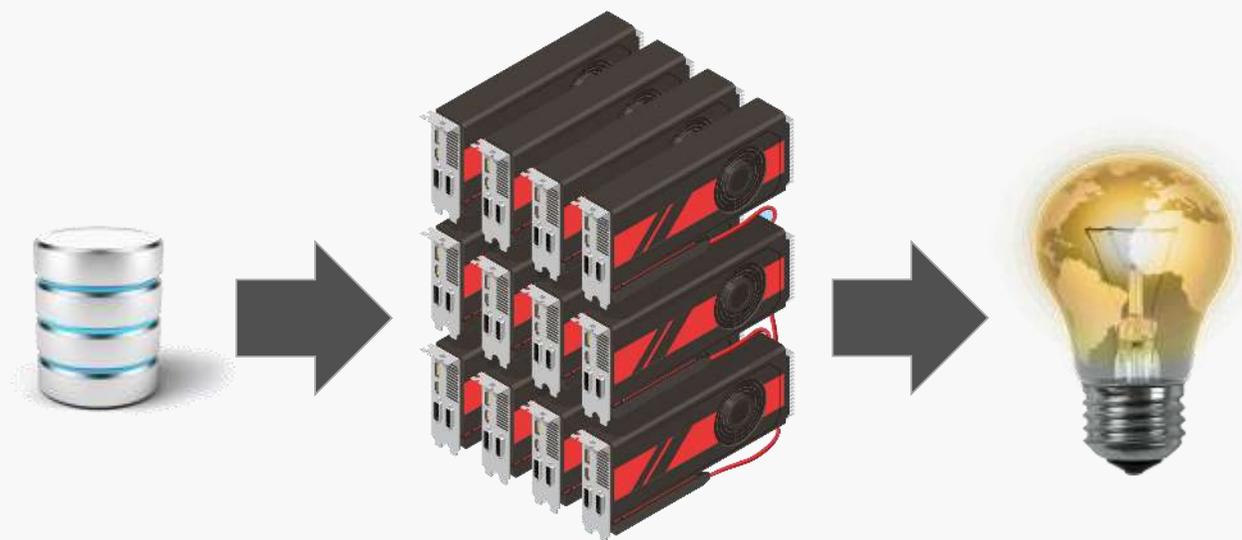
- 使用契約
- ライセンシー
- 許諾内容・数量
- 売買経路

# 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。
- データが形を変え永続し、新たな価値を創出する。



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

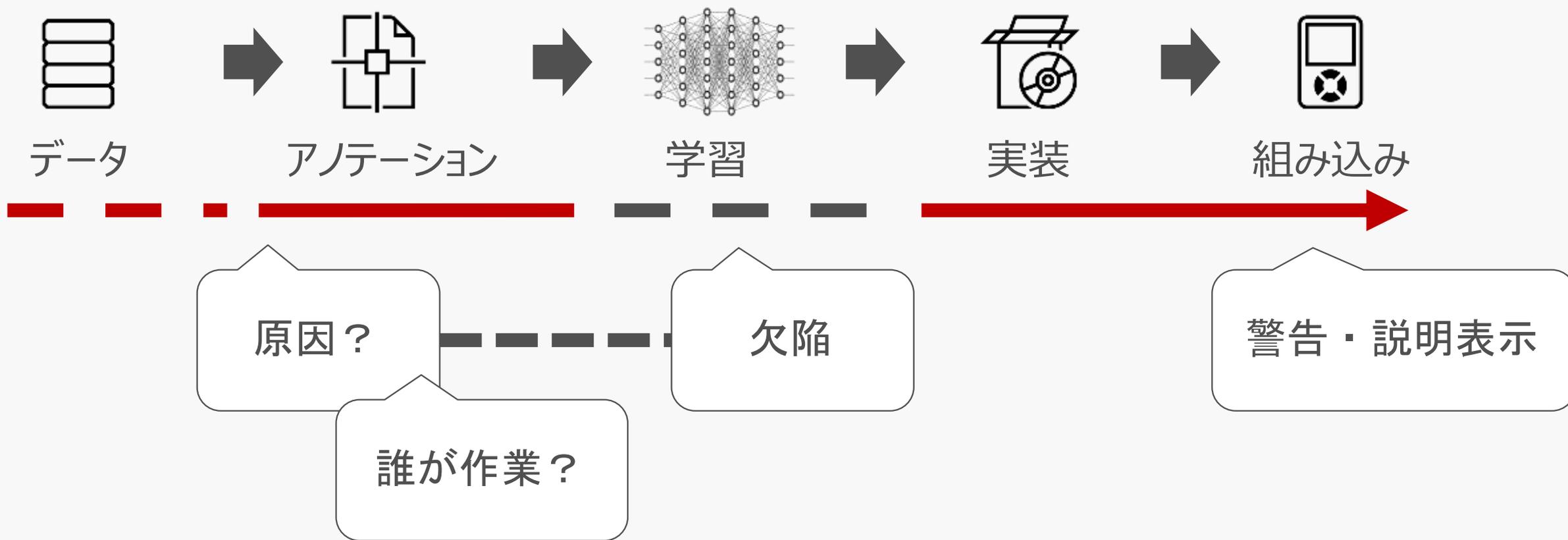
# 明確な製造物と責任

## ソフトウェア開発と責任



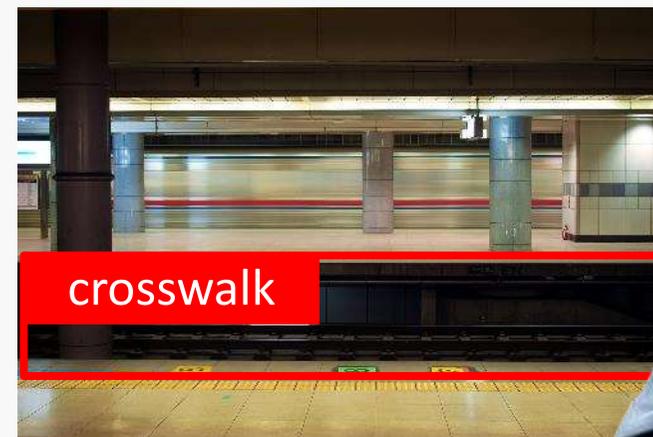
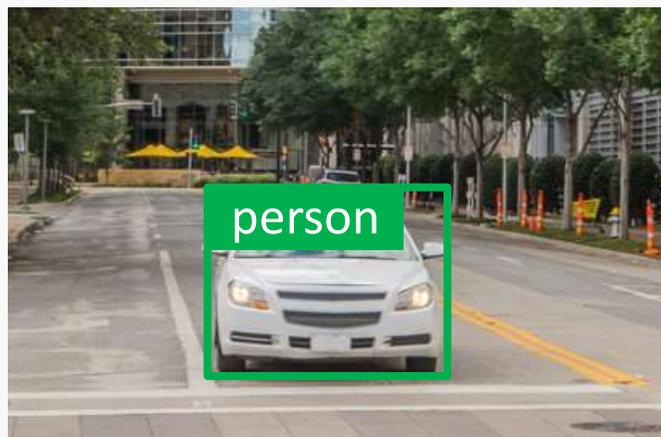
# 想定される責任範囲

## 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。

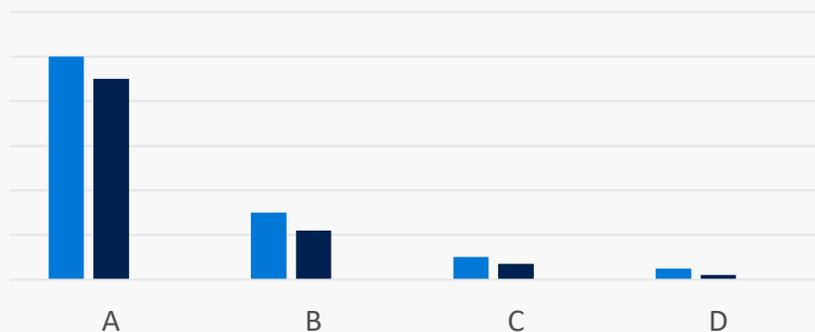
用途により異なるデータに求める保証レベル

# 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



### 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

➡ 時間と共に変動するデータの価値

やばい！  
このラーメンやばいよ！

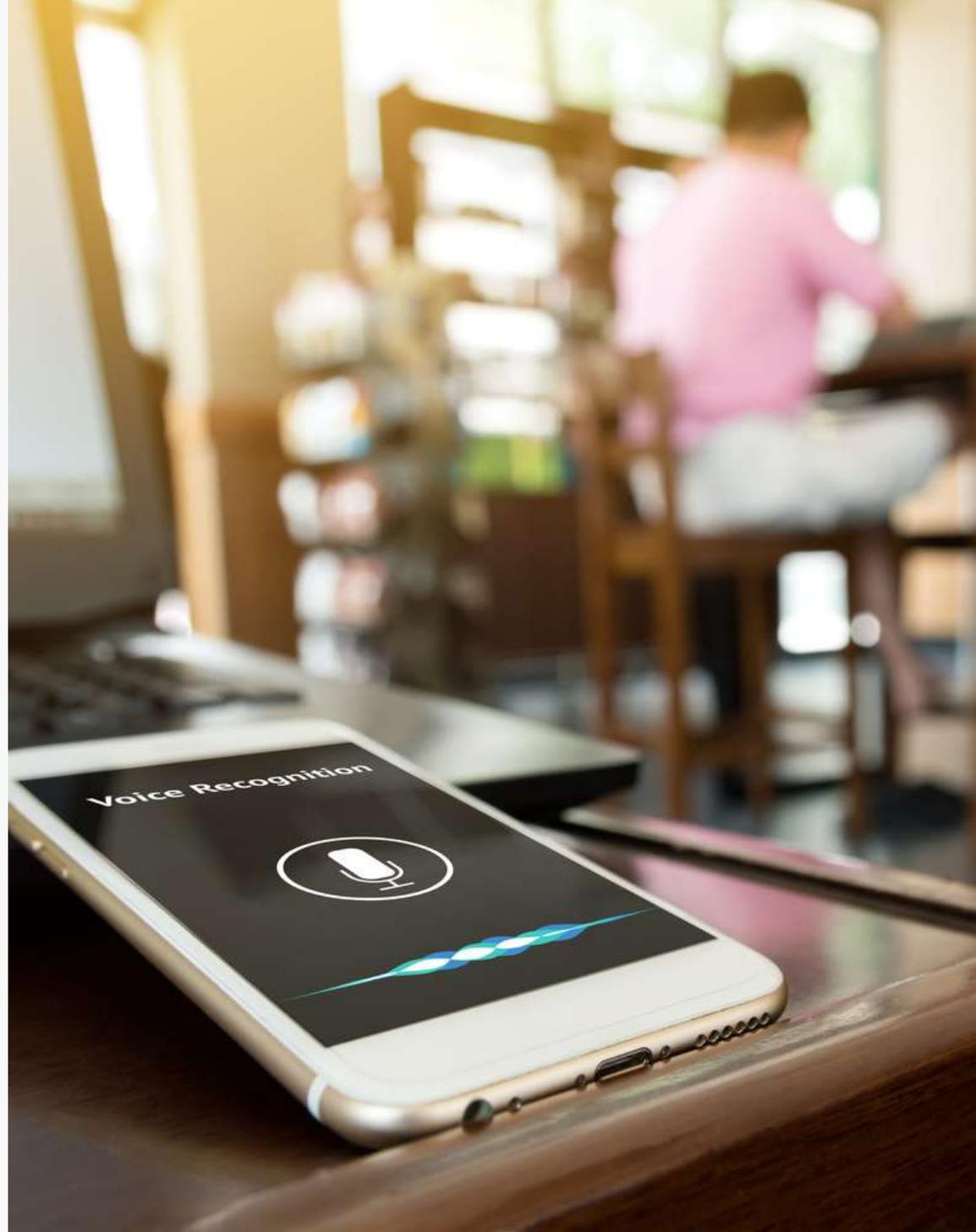
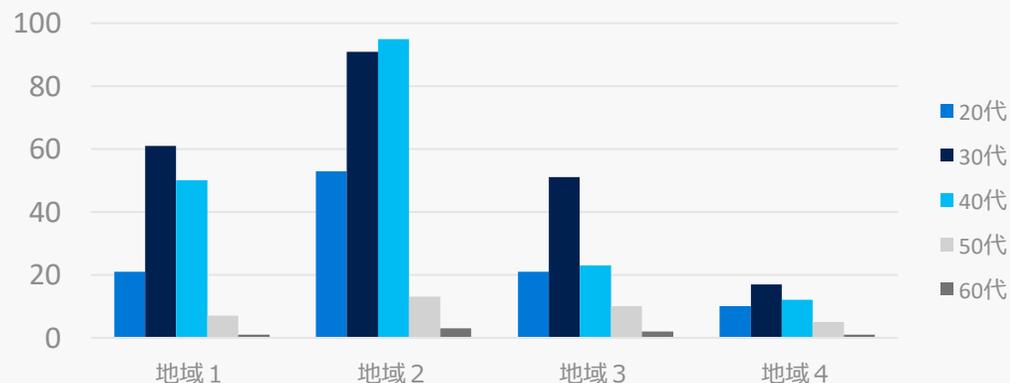
美味しい！  
このラーメン美味しいよ！

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で重要性を増す音声認識を含むAI活用
- 高まる地方のニーズと少ないデータ

例：音声データ分布



# データと計算リソースの統合

- ビッグデータの増加
  - より困難になるオンプレミスでのデータ処理
- 複数の異なるデータの統合と活用の増加
  - 例) 都市における人・モノの動態モデル (電車・バス、天候、道路、SNSなど)



# 利用者側の責任とは

「AIシステムにおけるデータ取り扱いの基本的な考え方」は、学習に使用するデータとAIシステムの関係を**責任**という観点から考察している。

政府 CIO 補佐官等ディスカッションペーパー

2020年6月

## AI システムにおけるデータ利用の特性と 取扱い上の留意点

2020年6月

田丸健三郎<sup>1</sup>、満塩尚史<sup>1</sup>、柳原尚史<sup>2</sup>、高木幸一<sup>3</sup>、  
西村毅<sup>1</sup>、梅谷晃宏<sup>1</sup>、楠正憲<sup>1</sup>、細川努<sup>1</sup>

### 要旨

AI（便宜上、本書では、深層学習を含む機械学習を「AI」と表記する）の研究が進むとともに、AIの実社会への適用についても広く検討され始めています。一方で従来のルール、ロジックを基にしたアルゴリズム実装とは異なり、AIの推論過程を論理的に人が理解できる内容で説明することは現在の技術では困難と考えられています。加えて、学習に用いたデータの品質がAIの推論結果の品質に大きな影響を与え、データの偏り、不適切なアノテーションなどが誤った推論結果となって現れます。このような特性を踏まえ、政府情報システムにおいては、公平性、安全性、透明性およびセキュリティへの一層の配慮が重要です。

本書ではAIを政府情報システムまたは政府が提供するサービス等で活用する際に、リスク度に応じて考慮すべきAIの学習データに関する透明性の確保、偏り（バイアス）の排除、データ加工の来歴保管の必要性、権利関係についてAIを用いたシステムに求める検討事項および考え方を記述しています。

# AIとデータを取り巻く課題の解決

## 組織内外で相互運用可能なデータ基盤

- **大きなデータ基盤開発コスト**
- 求められるデータ基盤に関する深い技術知識
- 限られた対応可能なソリューションベンダー

## 外部データ活用リスクの顕在化

- AIにおける**製造物責任**が大きな課題に
- AIで求められるデータ来歴
- 難しいAIにおけるデータと倫理と保証の関係

## 難しいAI研究者との相互連携、協働

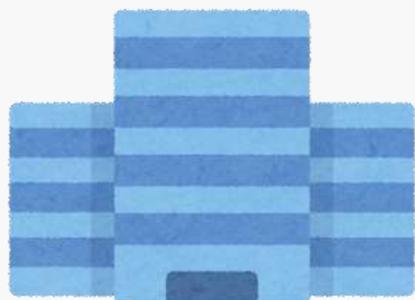
- 難しいAI研究者、エキスパートとの連携構築
- 難しい協働の為の仕組み構築

## AI学習用データ契約モデルの不在

- AIの学習で用いることを前提としたデータ流通、**契約モデルの不在**
- 技術特許にも似た**権利、知財**の問題

# 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

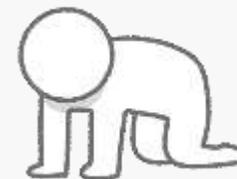
③ データが使い捨てではなく、付加価値を持つ事を把握

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

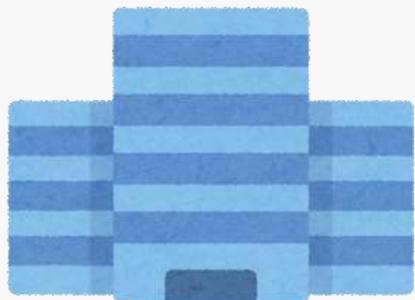


AI研究者  
スタートアップ



# 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



# AI開発サイクル

1. 問題の定義

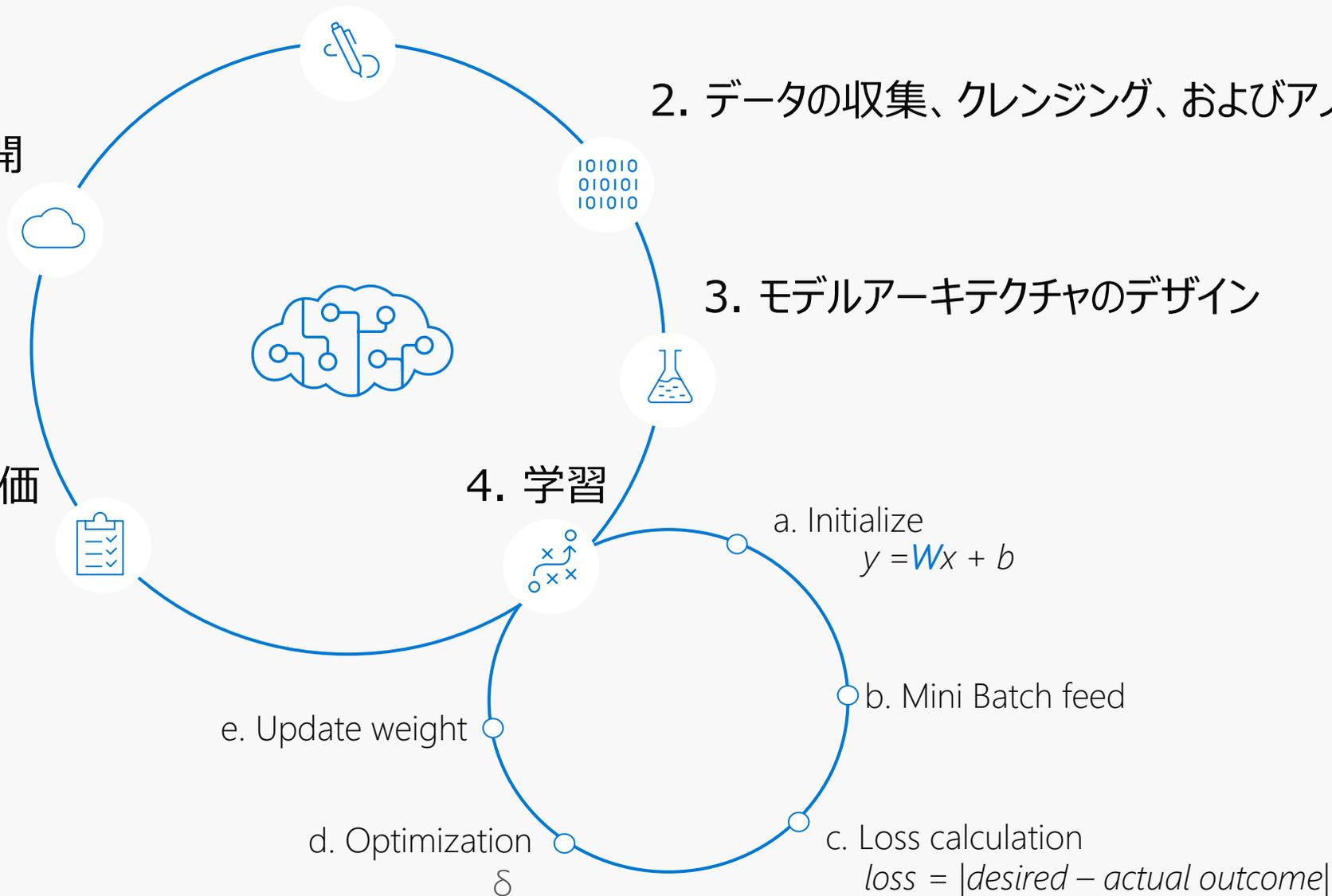
2. データの収集、クレンジング、およびアノテーション

3. モデルアーキテクチャのデザイン

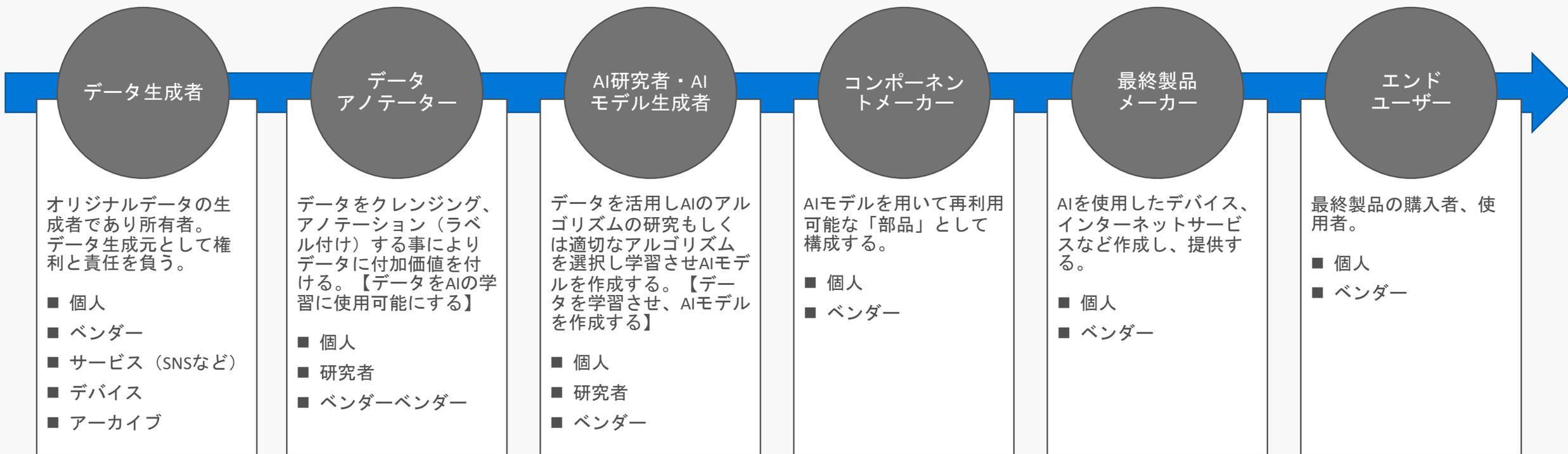
4. 学習

5. テストおよび評価

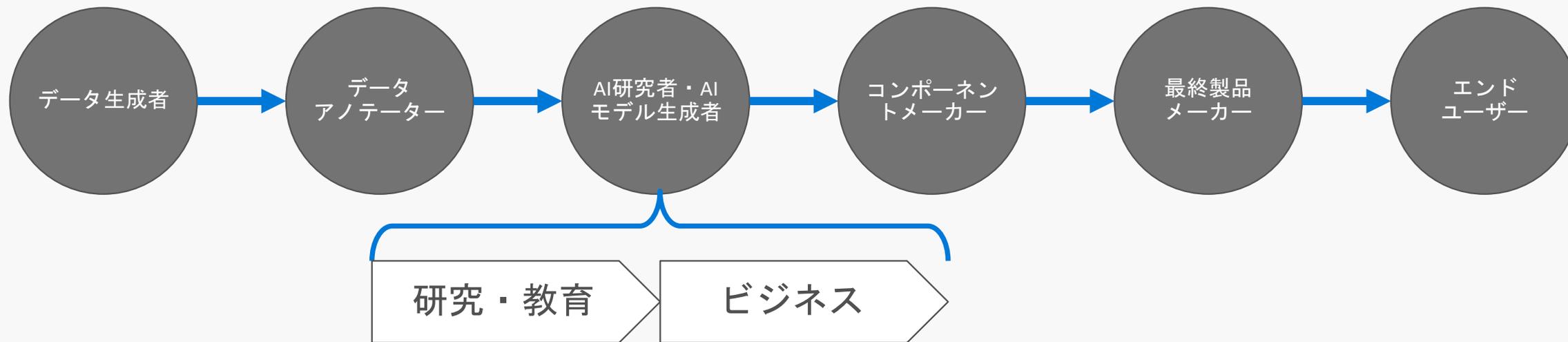
6. 展開



# AIにおけるデータの役割

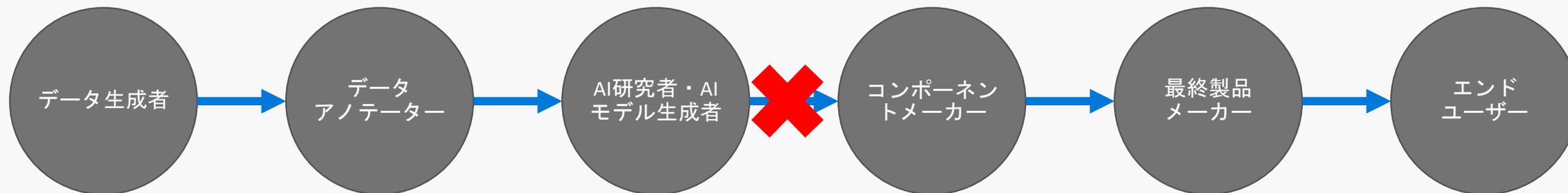


# 想定される商流（例）



- 活用フェーズによる変化する使用条件

# 想定される商流（例）



- 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。（購入データが最終的には活用されない。）

# コンソーシアムの取り組み

# 当団体が取り組む3つのテーマ

AIのためのデータ活用における課題の検討および解決のため作業部会（WG）を設置、AIの研究開発、ソリューション実現に寄与する。

## 知的財産

学習で使用するデータ、AIにより生成されるデータの権利関係の検討。

学習結果であるモデルの知財をどのように考えるのか。また、生成系アルゴリズムによる学習データと生成されたデータにおける権利関係の検討を行う。

## 法令

海外で進むゲノム、個人情報を含むデータ活用における法令、ガイドラインの調査、および検討。

国際的に競争力のあるAI研究、およびデータ活用における課題の検討、提言、法令解釈について作業を行い、AIにおける国際競争力強化に貢献する。

## データ共有基盤

多種・多様なデータを共有する上で求められるサービス基盤の検討。

完全にオープンなデータから、機密性の高いデータ、NDAなどの契約が必要となるデータ、一部有償のデータをホストする上で求められる基盤の検討。また、データ活用における利便性についても検討。

## 契約の標準化、関係省庁との調整

想定される知財、個人情報、プライバシーなどを考慮した手続きの確立。円滑なデータ、AIモデル取引基盤の実現

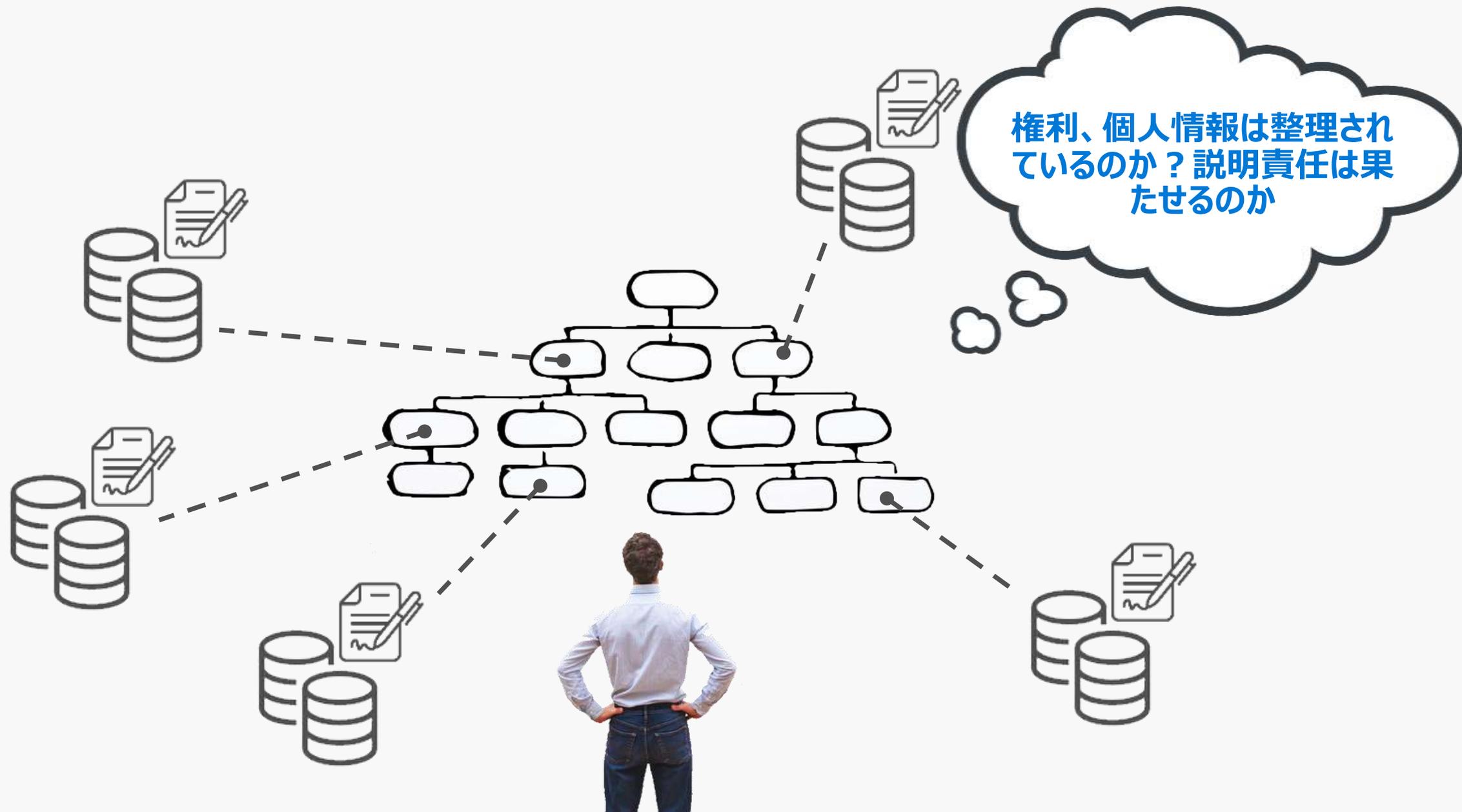
## データ収集・活用

基盤となるデータ収集、およびデータホルダーリレーションの構築。AI研究、イノベーションによるビジネス活用の推進。

## AIモデル、データ共有・取引基盤の構築

多様なトランザクション要件を満たした、多様なデータのストア・共有、AIモデル実行基盤の構築

# 外部データ活用の統制がより重要に



# 求められるAI利用を想定したデータ取引の仕組み



**商流**により異なる  
データ売買**契約**

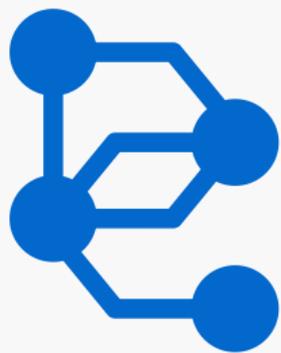


データに求められる  
**来歴と保証**



**時間**と**内訳**により  
変わるデータの**価値**

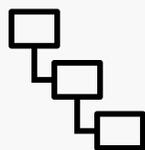
# ① 『AIDC DataCloud』が実現するデータ流通



Data Cloud



様々商流に対応した契約テンプレート



来歴情報に対応したデータカタログ



柔軟な課金モデル

データ販売代金から基盤維持に要する費用を手数料として課金

2～3%を想定

# 目指す姿とチャレンジ 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

# 技術要素



## APIサービス (データ、AI・分析モデル)

- Web API (動的 Web API によるデータ種別にとられないAPIサービス)
- メータリング (ユーザー、組織毎の使用状況の計測)
- 両方向 (データ取得、データストア)



## データストア

- 動的スキーマによる動的DBコンテキストスイッチ (データの種別、構造を問わないデータストア)
- 使用率、負荷に応じた動的配置 (フラッシュ〜コールド)



## データカタログ

- 内容、分類 (Secret, Confidential, Public)
- カテゴリ、タイプ、スキーマ、オーナー、ソース
- ライセンス、課金



## ID

- Active Directory, LDAP v3などによるIDおよび役割をベースとして管理
- (ID、および役割ベースのセキュリティ)



## Directory連携

- 相互認証のサポートと相互認証されているディレクトリサービスとの透過的なID連携



## データ可視化・Visualization

- データプレゼンテーション (グラフ化、簡易集計など)
- 汎用データポータル (定型データポータル)



## セキュリティ

- ID、役割ベースのセキュリティ管理
- 全ての操作ログ
- データの確からしさ (元データ、クレンジング等による変更履歴)



## プリセットAIモデル

今後の更新で対応予定

- リポジトリ内データによる汎用学習済みモデル
- 一般に公開されている汎用AIモデル



## 深層学習・Notebook

- TensorFlow/Keras/PyTorchなどによる深層学習環境 (VM)
- Jupyter NotebookなどのPoC、学習環境
- マルチノードによる分析・学習のスケールアウト

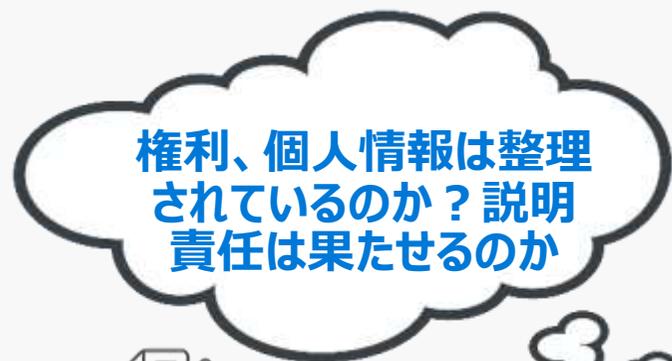


## IDEベース機械学習

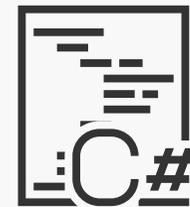
- Web IDE/GUIによるデータ分析・学習のフローデザイン、学習、展開
- Web UIベースのモデル評価
- 学習済みモデルのWeb APIの自動化された展開

## ② 『AIDC DataCloud』を会員に無償ライセンス

～ 社内データ統制の実現を支援 ～



Data Cloud



### 削減できる3つのコスト

- 複雑な契約モデルの作成
- 来歴に管理に対応したデータ共有基盤の開発
- データ統制

# 書籍データの可能性

## データとしての価値とビジネス

# 部会のご紹介

# 知的財産・契約検討WG

**主査：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】**

**運営：阿部豊隆 TMI法律事務所（弁理士） パートナー**

AI（深層学習）・ML（機械学習）におけるデータ活用は売上分析のような従来のデータ活用と異なり、使用したデータが学習結果として永続化される点において大きな違いがあります。データが形を変え永続化される事は、従来のデータ活用とは大きく異なる価値をデータにもたらす事に他なりません。

AI・MLにおけるデータの商取引は、テクノロジー特許と多くの類似点があります。テクノロジー特許が商流、最終製品の数・単価などにより契約、価格が変わるのに対して、データのAI・ML活用では特許が持つ特性に加えてデータの希少性、ドメイン、鮮度など、より多くの要素、開発・ユーザーシナリオにより契約・価格が異なります。加えて、データそのものの知的財産だけでなく、個人情報、プライバシーなど様々な法令、ガイドラインとの整合性も考慮した検討が必要となります。

知的財産・契約検討WGでは、商流、最終製品の形態などを考慮した契約手順、契約テンプレート、そしてデータ流通基盤で使用するスマートコントラクトを実現を目指します。

# AI研究WG

**主査：杉山 将 【革新知能統合研究センター (AIP) センター長 / 東京大学大学院教授】**

AI研究WGでは、AI研究に資するデータの検討とそれらデータを用いた社会課題の解決方法について検討を行います。北米、中国では積極的なデータ収集により地域に根差したAI研究による社会課題の解決が積極的に進められています。AI研究WGでは、他のWGと連携し、データの流通・活用によりAI研究の促進を通じた社会課題の解決に取り組みます。



脳性麻痺、聴覚障害者の話す言葉を認識できるASR (Automatic Speech Recognition) エンジンが存在しない



支援者の多くは、発話困難な障害者の話す言葉には共通点があると感じている。



会話データの収集と最新のAI研究による発話困難な障害者のコミュニケーション、社会参画を支援

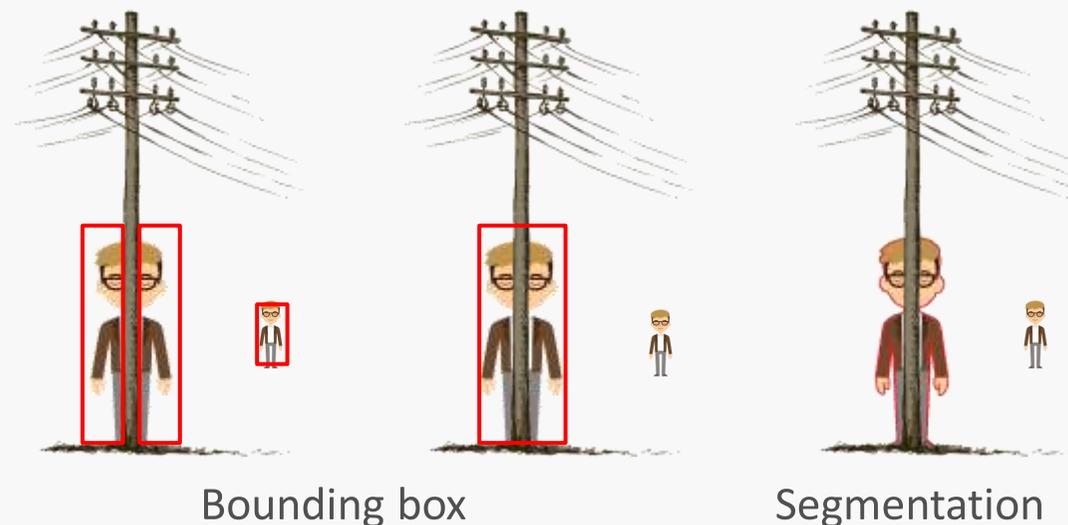
# データ流通・活用WG

**主査：井佐原 均 【豊橋技術科学大学 教授／情報メディア基盤センター長】**

AI（深層学習）・ML（機械学習）におけるデータ活用において、求めるデータの所在、所有者を探し、契約に至るまでの道のりは非常に長いものとなっています。これは、データ提供側は個社ごとに商流の確認、契約内容および価格の交渉などをしなくてはならず、必要となるその煩雑な作業と費用から、データ提供に躊躇するデータ所有者も少なくありません。一方では保有しているデータを周知できず活用できないデータ所有者も少なくなく、データ所有者と利用者のネットワークから活用、有用な事例の創出を行います。

この他、データにはセンサーなどのデバイスから直接生成される（生）データ、クレンジング、アノテーションされたデータなど様々な種類が存在します。同一のデータであっても、利用目的に応じてクレンジング、アノテーションの基準も異なります。

データ流通・活用WGでは、データ所有者、アノテーター、そしてデータ利用者によるデータ流通・活用を促進に資する活動を行います。



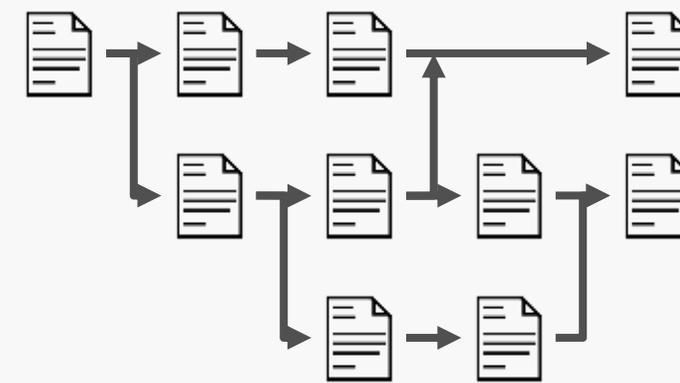
多岐に渡るアノテーション基準

# データ基盤WG

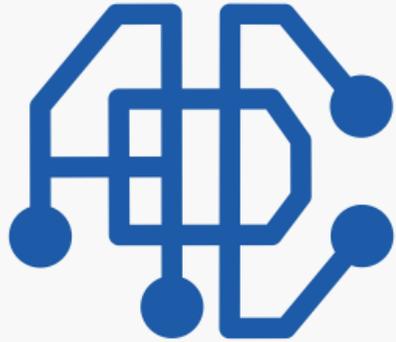
**主査：田丸 健三郎 【日本マイクロソフト株式会社 業務執行役員 NTO】**

データ流通・共有の為のサービス基盤の構築は、当コンソーシアムにおける重要な活動の一つです。データ流通基盤には、目的とするデータを見つけるためのカタログサービス、様々な種類のデータ（構造化データ、非構造化データ、ストリーミングなど）に対応したデータストア媒介サービス、容易なデータ分析を可能にする機械学習・深層学習サービス、データ流通の為のスマートコントラクトの仕組みを構築、提供します。また、AI（深層学習）・ML（機械学習）を用いたソリューションであっても負う製造物責任。データの透明性、来歴管理、クレンジングやアノテーションの作業確認と作業内容の記録など基盤には様々な要件が求められます。

これらの要件を満たすためには、強固なセキュリティ、ID管理などの認証基盤の構築が必要なるだけでなく、認証機関と連携したデータ提供者、編集者の確認と保証の仕組みが重要となります。データ基盤WGでは、これら様々な要件を満たす基盤のアーキテクチャの検討を行い、会員と協働し実装を行います。また、基盤は会員限定オープンソースとして実装を行い、会員における組織内データ流通・活用基盤の実現を可能にします。



リビジョンと来歴管理



# AI Data Consortium

AIデータ活用コンソーシアム

<https://aidatacon.com>

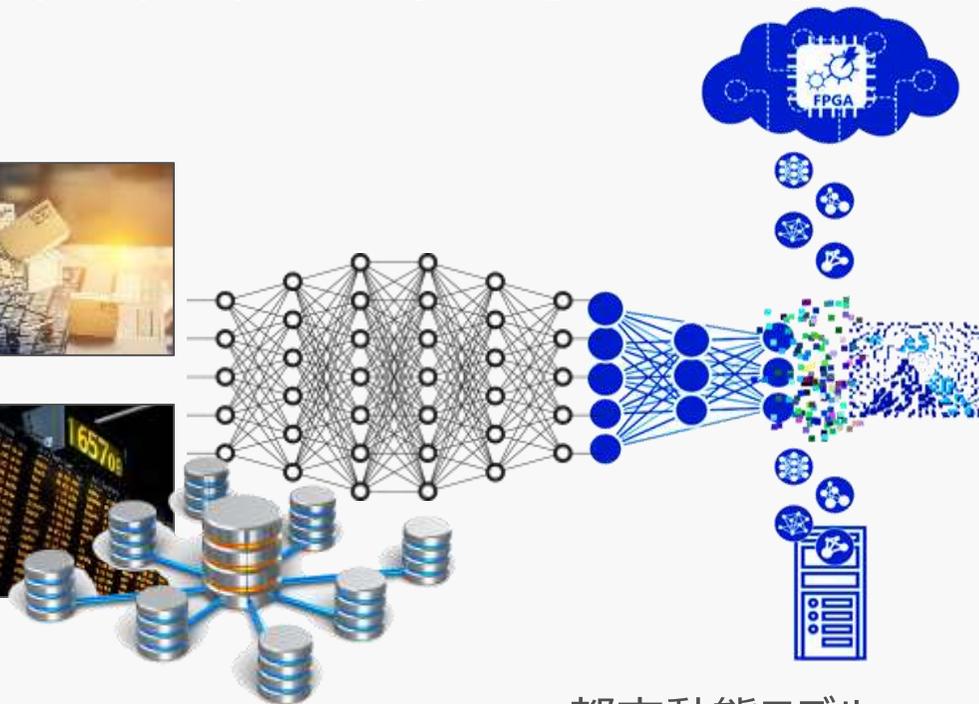
# Appendix

ユースケース

# SmartCity / Society 5.0とAIデータ活用

## データの統合と活用

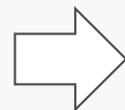
- 複数の異なるデータをワンストップで手続き、活用
  - 複数の異なるデータ検索、検索、契約、学習から推論までをワンストップで実行



例) 都市の動態モデルによる人・物の予測

混雑前の移動（時間差移動）、  
混雑した経路の回避

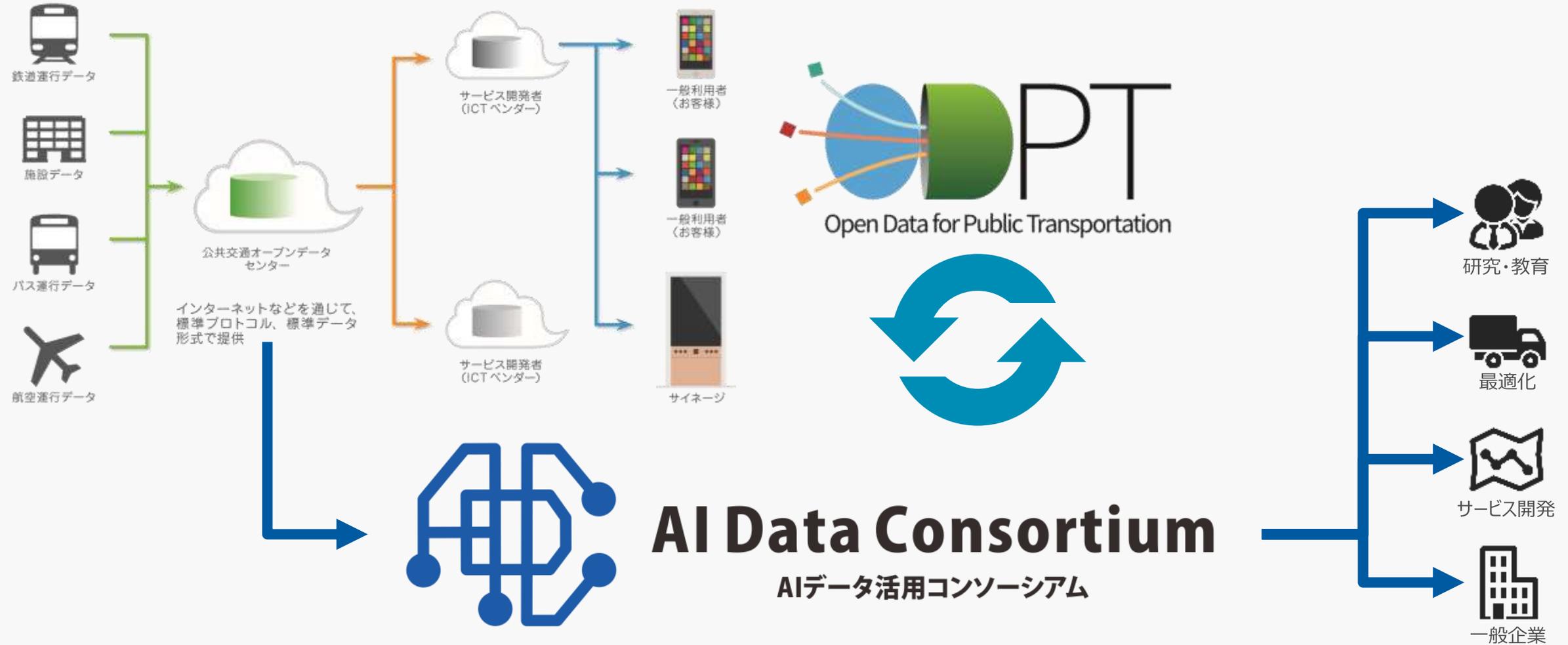
気象、モバイル空間統計、交通、GISなど



- 都市動態モデル
- 鉄道・バスダイヤ最適化
- (イベントなどによる大希望移動時の) ルート案内最適化

# 公共交通とAIデータ

## 公共交通オープンデータ協議会と連携しプロトタイプスタディ



複雑な商流・用途に対応した諸手続き、メータリング、法令対応により、円滑なデータ流通を実現

## 障がいのある方に向けたAIデータ活用

# コンピュータ操作に困難のある方に向けた データ活用プロジェクト

- ソフトウェアやデバイスの発展により、障がいのある方もコンピュータを使ってコミュニケーションをしたり、仕事をされることが一般的に
  - その方の動かせる部位にあわせたスイッチ
  - 視線制御機能
  - 操作支援ソフトウェア
    - スイッチや視線で操作しやすいインターフェース
    - 限定的操作予測機能を備えたソフトウェア (Pete) も
- 一般の方に比べて、一つの操作に時間がかかることが課題



# コンピュータ操作に困難のある方に向けた データ活用プロジェクト

- 高度な操作予測を行うため、多くの操作ログデータを集めて操作に困難のある方の操作支援に活用



## 日本支援技術協会の役割

- 必要なログデータ研究
- ログ収集ツール開発
- ログデータ提供先との協議

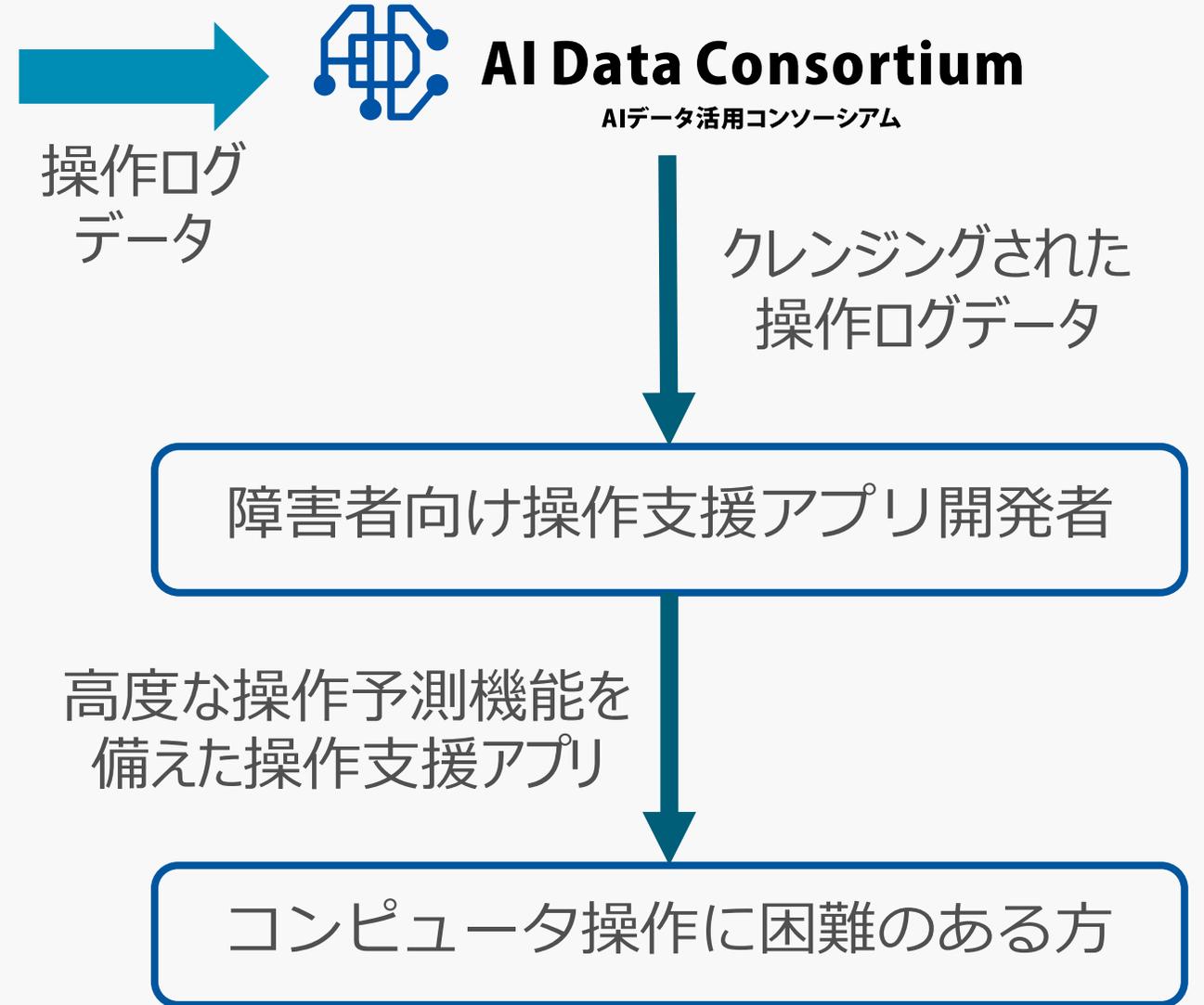
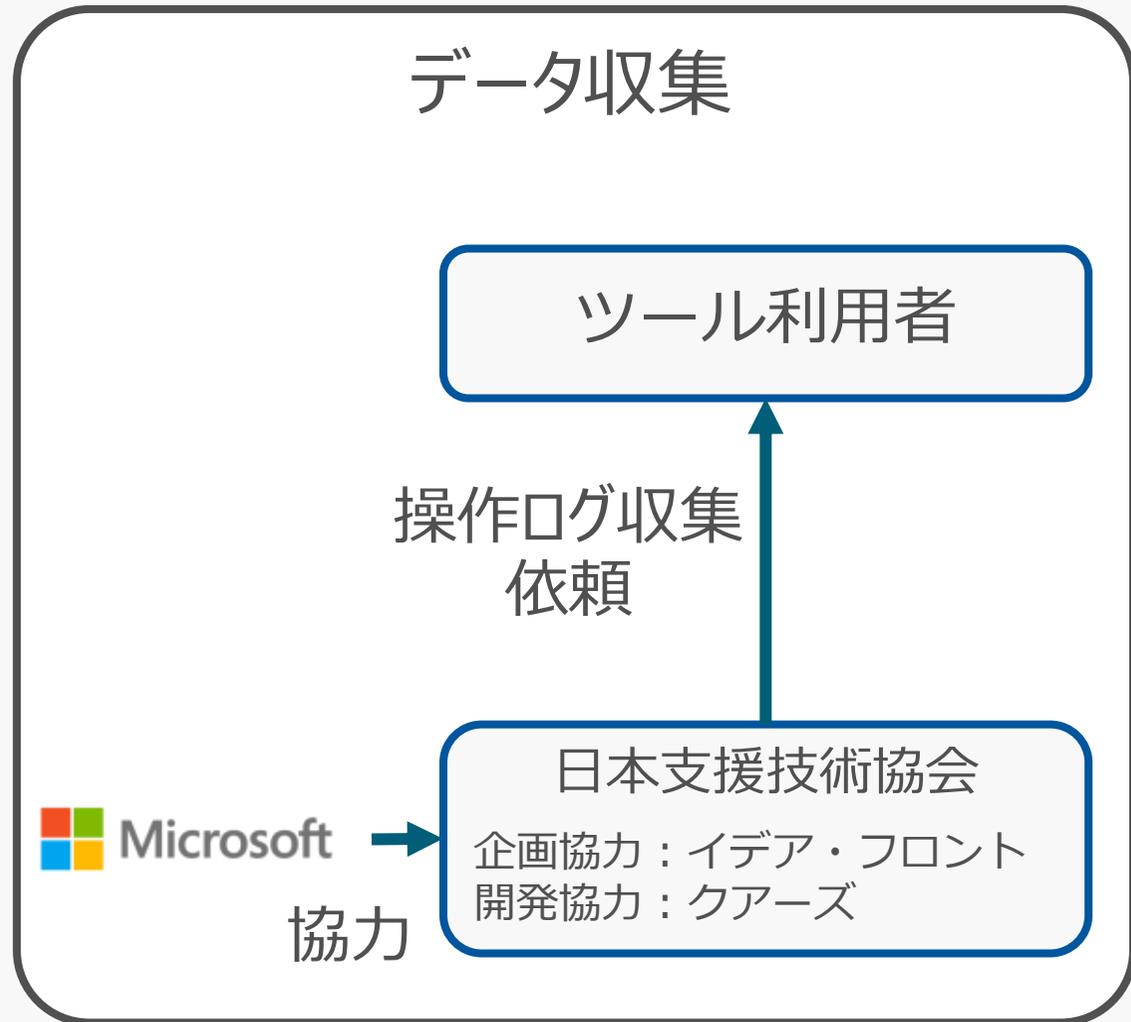


## AIデータ活用コンソーシアムの役割

AI Data Consortium  
AIデータ活用コンソーシアム

- ログデータ収集ツール提供
- ログデータ保管・提供
- ログデータ収集・保管・提供に関するルールづくり

# 実施体制



## 高齢者に向けたAIデータ活用

# 高齢者の音声データ活用による 高齢者とロボットのAIによる会話促進

- 高齢者にとって「会話」は認知機能を低下させないために非常に重要
- 高齢者施設などで高齢者がロボットとAIで会話することは人手不足の観点からも有効
- 高齢者の音声データは少ないため音声認識の精度が低い傾向



# 高齢者の音声データ活用による 高齢者とロボットのAIによる会話促進

- 高齢者に特化して音声データを収集し、高齢者と会話するロボットなどに活用



- 日本の高齢者の音声データ収集



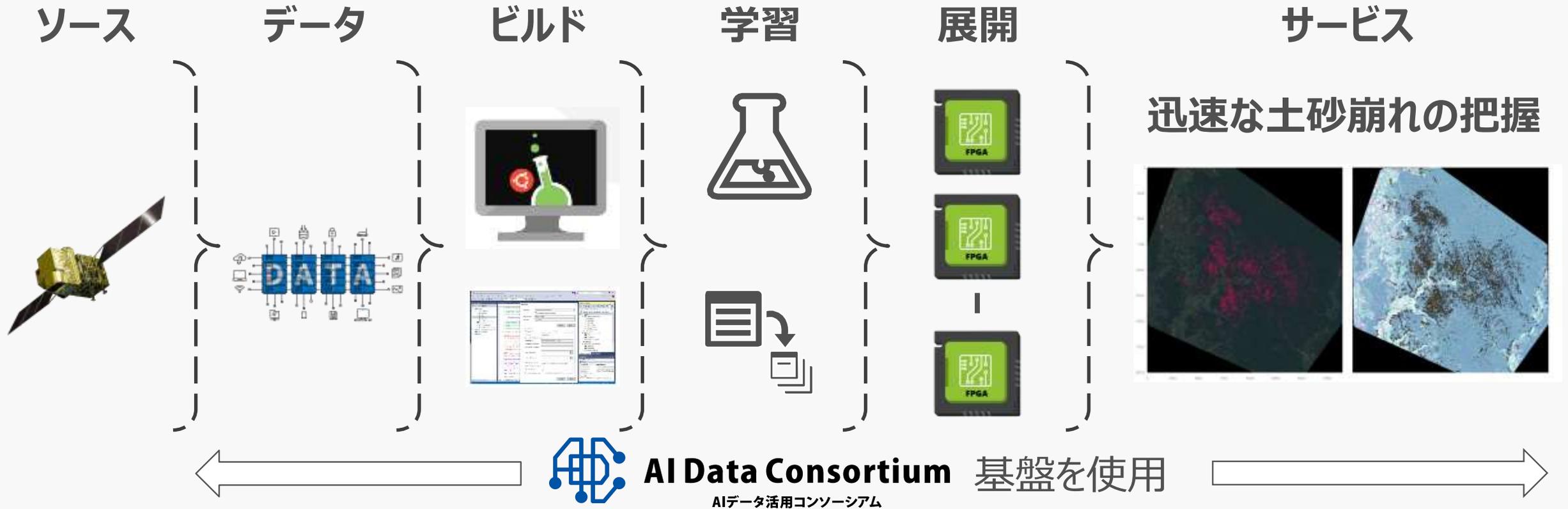
- 高齢者とコミュニケーションするロボットへのAIの追加
- 高齢者施設での実証研究



**AI Data Consortium**  
AIデータ活用コンソーシアム

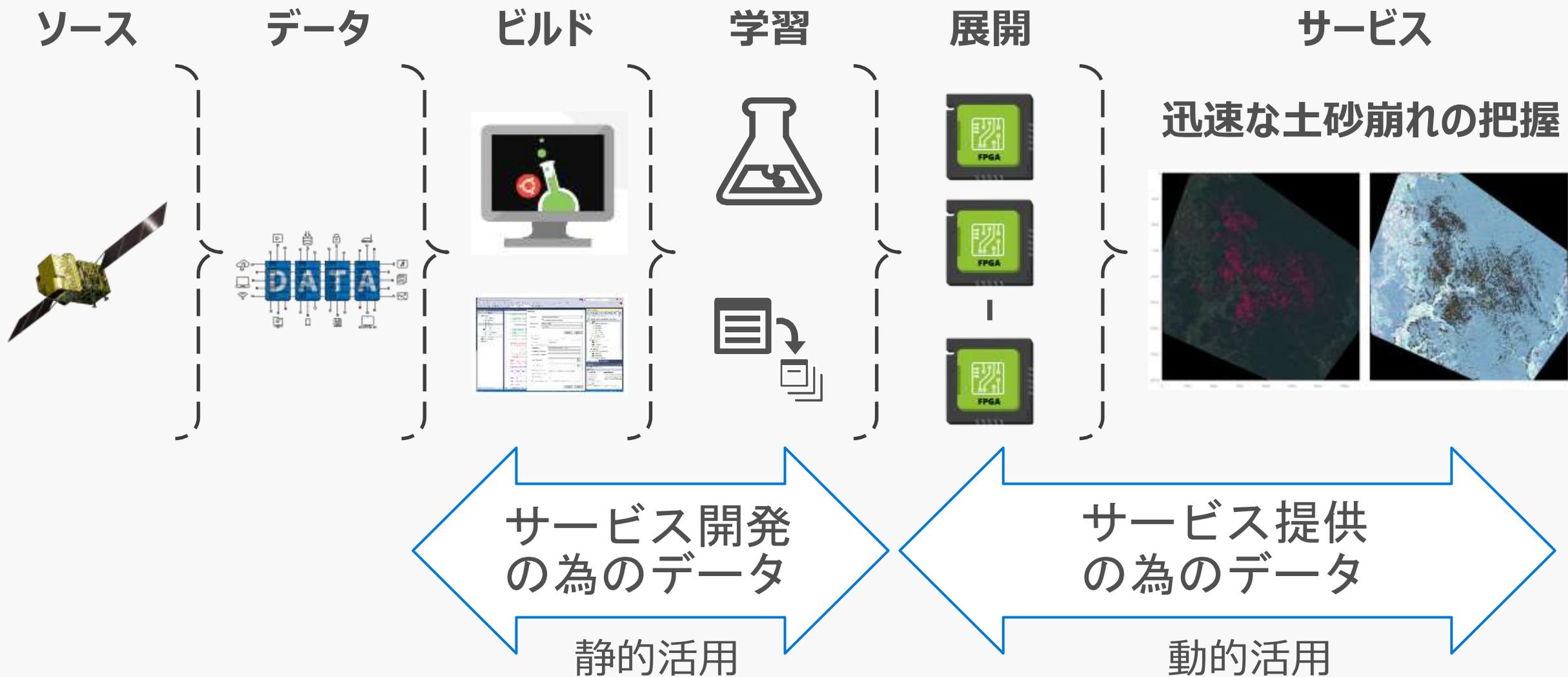
- 高齢者の音声データの保管・ATR への提供
- 保管・提供に関するルールづくり

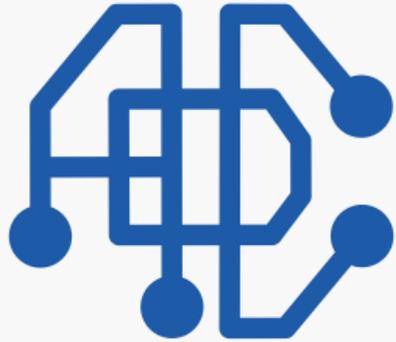
## データのパイプライン処理とサービス提供



- (自治体) 迅速な被災状況の把握による住民支援
- (保険会社) 迅速な支払額予測と顧客対応

# フェーズにより異なるデータ契約





# AI Data Consortium

AIデータ活用コンソーシアム

<https://aidatacon.com>