

# データ活用と課題と取り組み

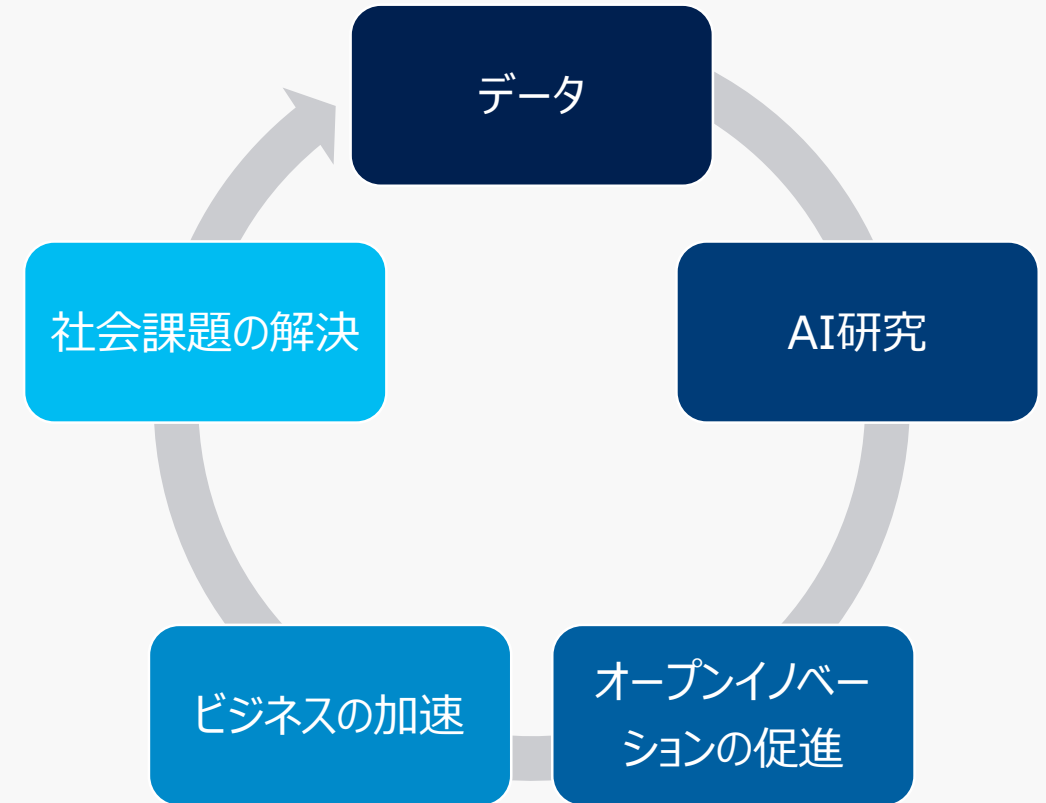
一般社団法人 AIデータ活用コンソーシアム 理事・副会長  
日本マイクロソフト株式会社 業務執行役員 NTO

田丸 健三郎

# 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・オブザーバー

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)
- 独立行政法人 情報処理推進機構 (IPA)
- 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

# コンソーシアムの主な活動

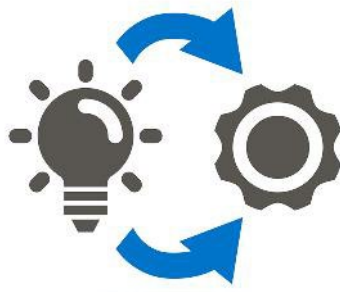


## 知的財産・契約

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。

### 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

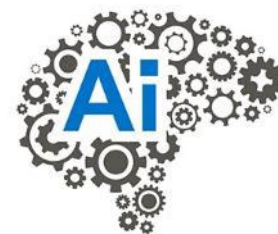


## AI研究

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。

### 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



## 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。

### 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



## 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。

### 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

# 重要性増す責任 データと責任と知的財産

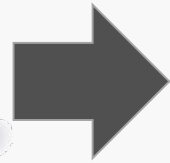


# AIが変えるデータ売買契約

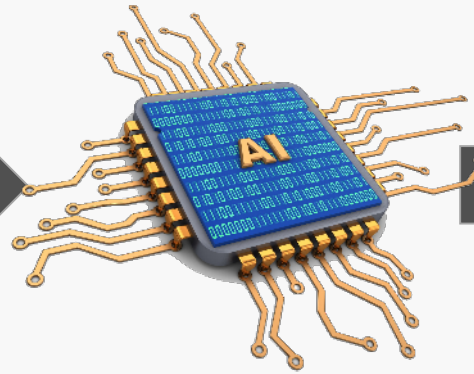
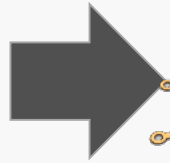
AIのビジネス活用により変わるデータの価値と  
複雑になるデータ取引・契約



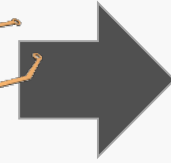
データ



学習



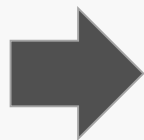
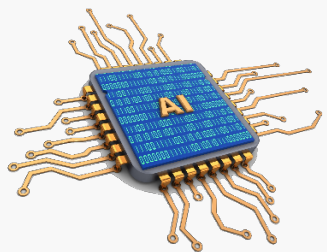
AI



ビジネスの創出



# 複雑な商流と権利



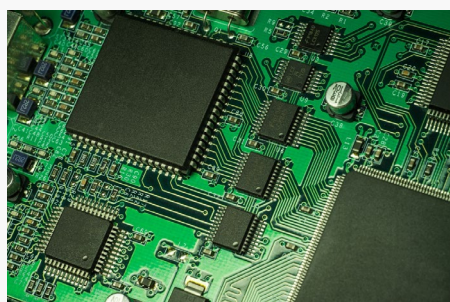
ライブラリ



最終製品・サービス



エンドユーザー



コンポーネント

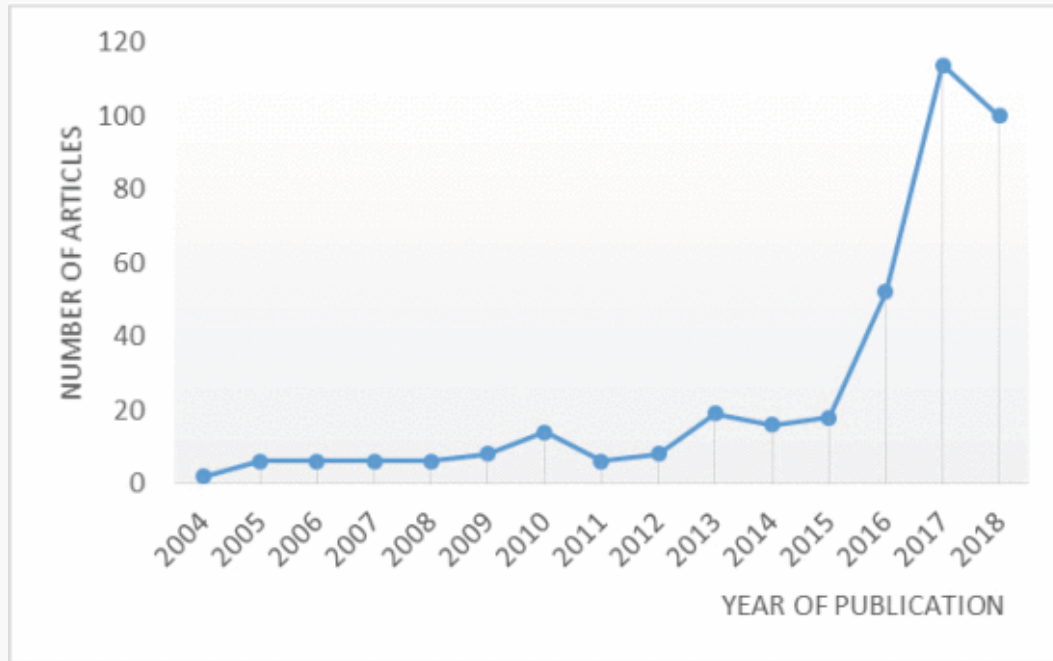


# 明確な製造物と責任

## ソフトウェア開発と責任



# "Right to Explanation"



Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)

## European Union regulations on algorithmic decision-making and a "right to explanation"

Bryce Goodman, Seth Flaxman

(Submitted on 28 Jun 2016 (v1), last revised 31 Aug 2016 (this version, v3))

We summarize the potential impact that the European Union's new General Data Protection Regulation will have on the routine use of machine learning algorithms. Slated to take effect as law across the EU in 2018, it will restrict automated individual decision-making (that is, algorithms that make decisions based on user-level predictors) which "significantly affect" users. The law will also effectively create a "right to explanation," whereby a user can ask for an explanation of an algorithmic decision that was made about them. We argue that while this law will pose large challenges for industry, it highlights opportunities for computer scientists to take the lead in designing algorithms and

## IGF 2018 WS #421 Algorithmic transparency and the right to explanation

**Theme:** Cybersecurity, Trust and Privacy

**Subtheme(s):** ALGORITHMS

**Organizer 1:** Civil Society, African Group

**Organizer 2:** Civil Society, Western European and Others Group (WEOG)

**Format:** Break-out Group Discussions - 60 Min

**Interventions:**

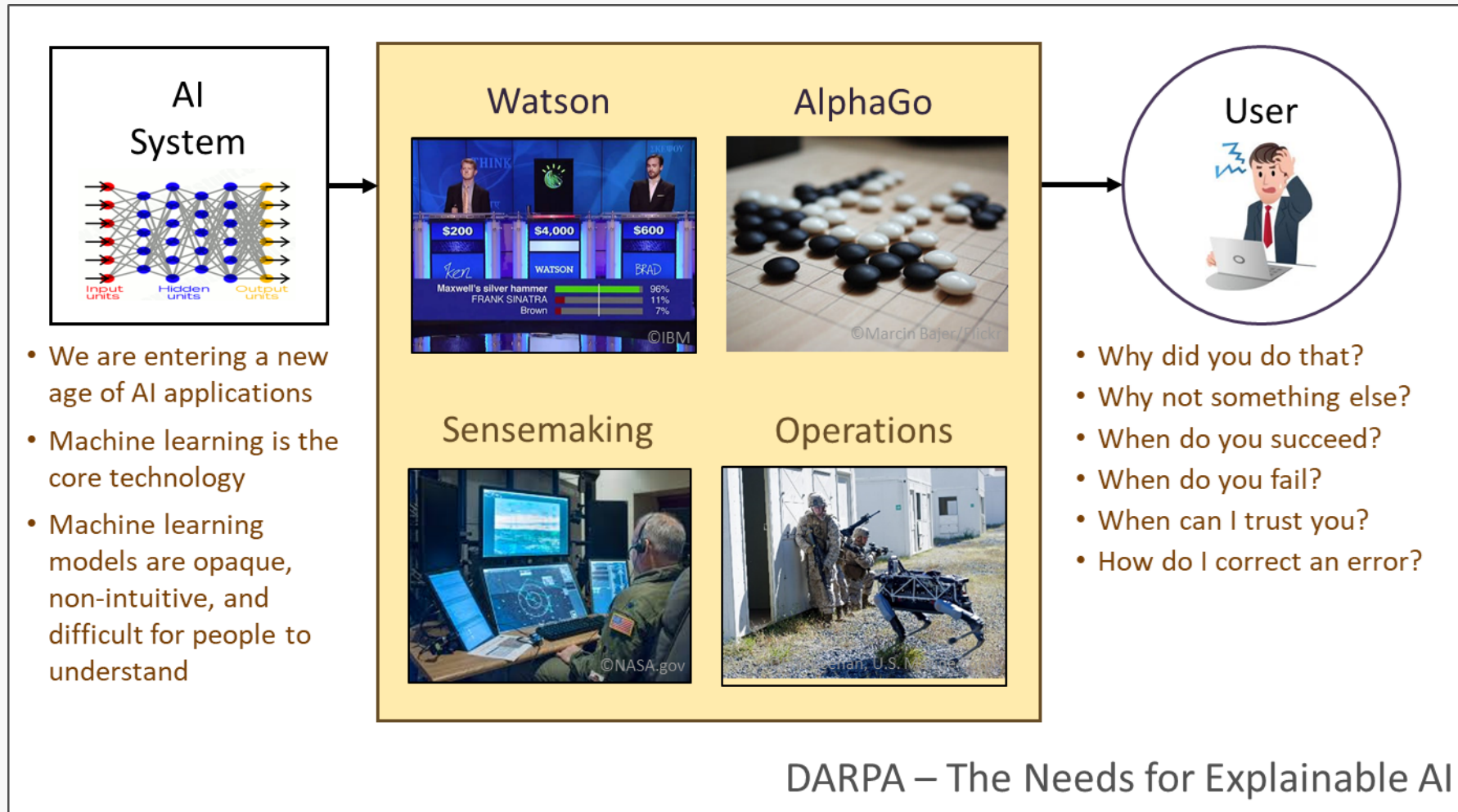
- Alex Comninos and Deborah brown (APC) will be moderators - Jelena Jovanovic (cyber security professional) will provide an overview of the concepts of algorithmic transparency, algorithmic justice, algorithmic bias and real life examples of the effects of algorithms from an information security perspective. - Vidushi Marda (Article 19) will provide an overview of the human rights aspects of automated decision making. She will focus on the GDPR Article 22 and the EU guidelines on Automated decision-making. She will provide a policy and human rights perspective.

The ICANN 2018 scientific programme includes the following workshops.

## Interpretable Methods for Machine and Deep Learning

**Description:** Machine learning-based methods and, in particular, deep neural networks have proven extremely successful in a wide variety of tasks but remain for the most part extremely opaque in terms of what they learn and how they use their acquired knowledge to make predictions. More specifically, without any insight into the network, the task of determining whether the system has properly integrated a specific concept is very difficult and the subsequent validation of the method for critical activities (e.g. driving, medicine,

# AIのビジネス活用と説明責任



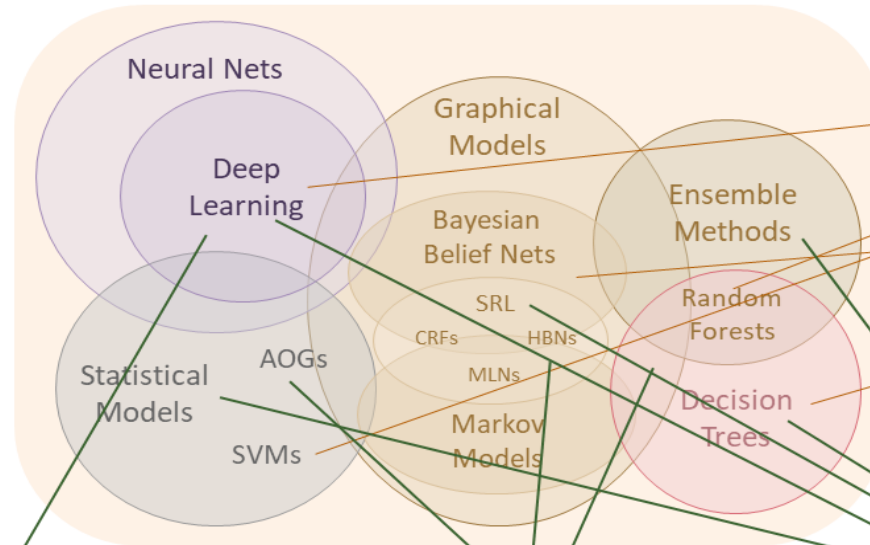


# B.1 Explainable Models

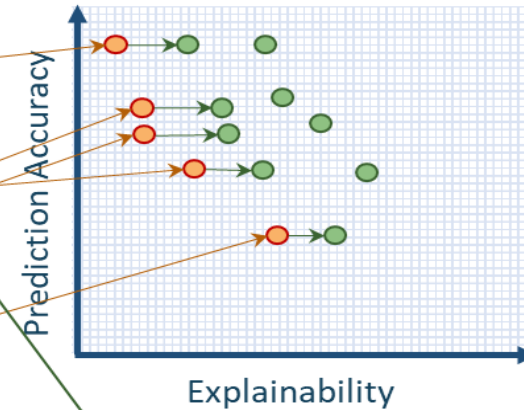
## New Approach

Create a suite of machine learning techniques that produce more explainable models, while maintaining a high level of learning performance

## Learning Techniques (today)



## Explainability (notional)



**Deep Explanation**  
Modified deep learning techniques to learn explainable features

The diagram shows a neural network with 'Input units', 'Hidden units', and 'Output units'. Specific units are labeled with 'Whiskers', 'Fur', and 'Claws', indicating the model has learned to identify these features.

**Interpretable Models**  
Techniques to learn more structured, interpretable, causal models

The diagram shows a decision tree with nodes containing numerical values, representing a structured and interpretable model.

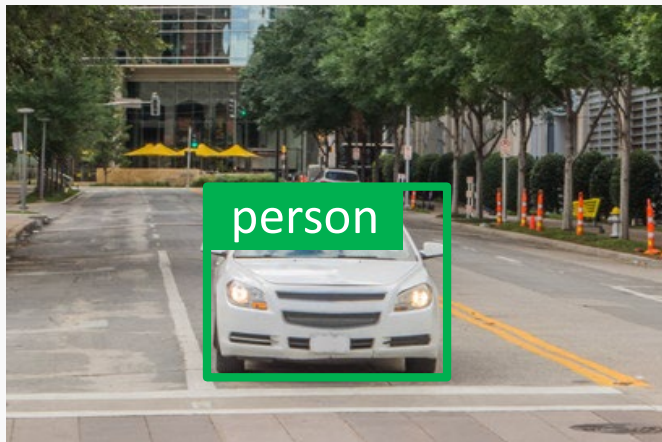
**Model Induction**  
Techniques to infer an explainable model from any model as a black box

The diagram shows a 'Model' box with a question mark, connected to an 'Experiment' box, illustrating the process of inferring an explainable model from a black box.

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

# AIが変えるデータの来歴管理

- トレーサビリティが求められる作業履歴
- 排除が難しい「誤り」と「悪意」によるアノテーション

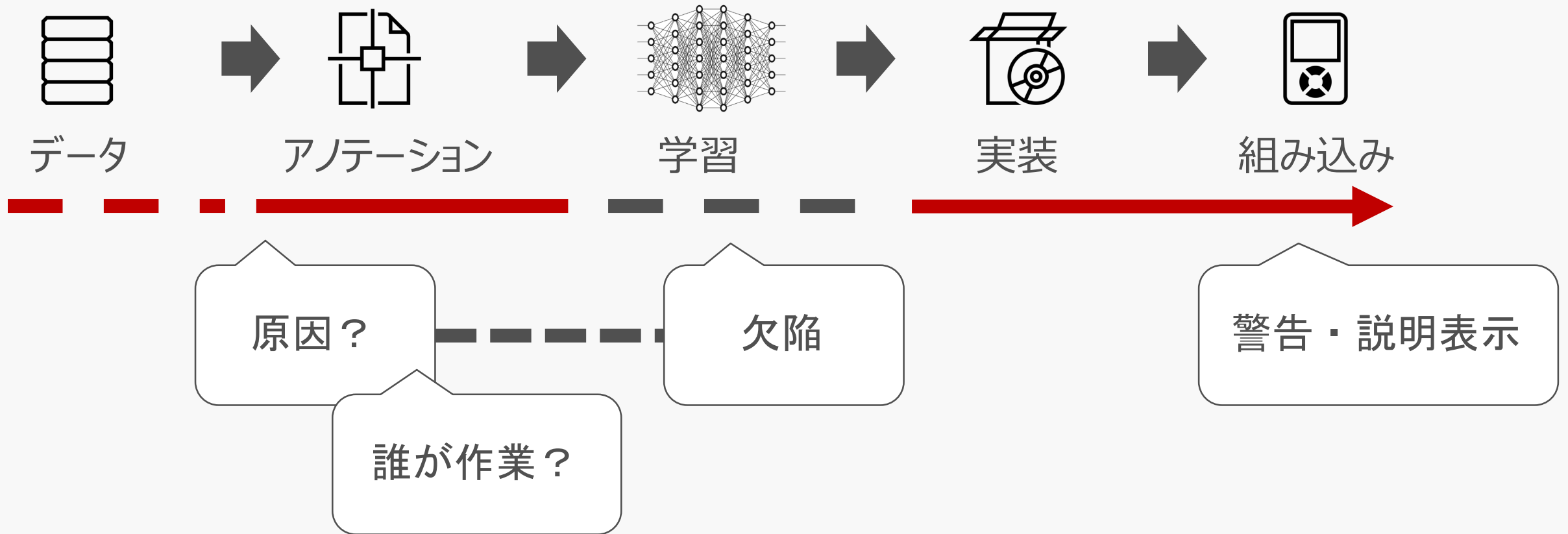


重要性を増す作業者の**身元確認**と**保証**



# AIが変えるデータに求められる責任

## AIを用いたシステムと責任



# AIの用途 × 保証 × 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



# 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

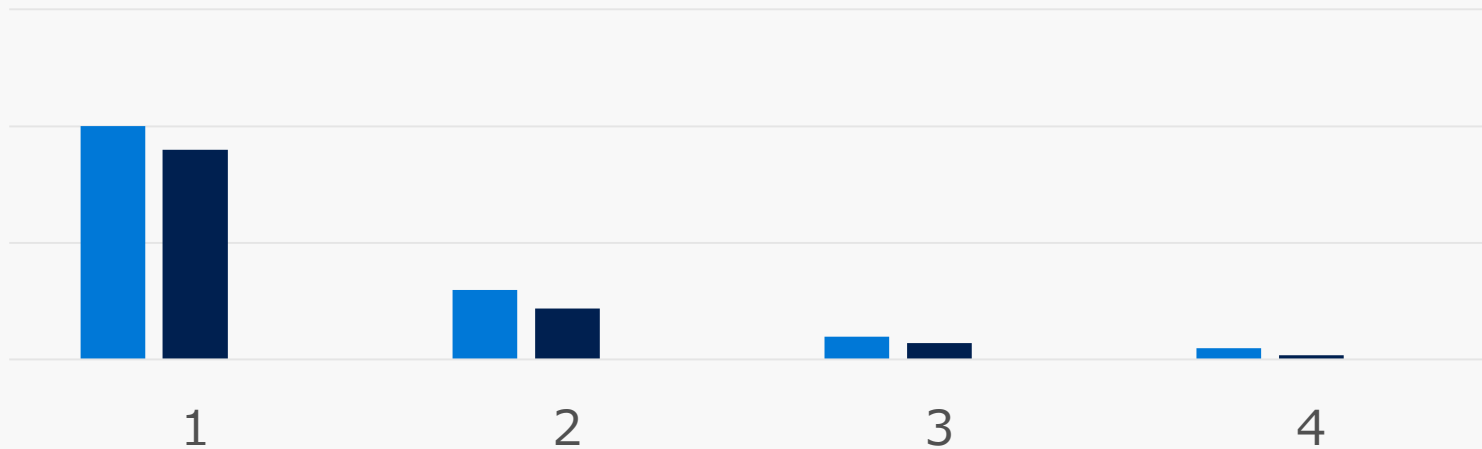
AI（便宜上、本書では、深層学習を含む機械学習を「AI」と表記する）の研究が進むとともに、AIの実社会への適用についても広く検討され始めています。一方で従来のルール、ロジックを基にしたアルゴリズム実装とは異なり、AIの推論過程を論理的に人が理解できる内容で説明することは現在の技術では困難と考えられています。加えて、学習に用いたデータの品質がAIの推論結果の品質に大きな影響を与え、データの偏り、不適切なアノテーションなどが誤った推論結果となって現れます。このような特性を踏まえ、政府情報システムにおいては、公平性、安全性、透明性およびセキュリティへの一層の配慮が重要です。

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が変えるデータの価値評価

- 時間と共に変化する価値
- 分布により異なる価値

## データの内訳と品質とマーケットバリュー



順位	2018年	2019年
1	マジ	り
2	すこ	パリピ
3	彼ピッピ	タピる
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

# 報告書 序文

- **AIDCプラットフォームにおけるデータ提供契約に関する報告書**
- **機械学習を中心とするいわゆるAI（人工知能）が機能するためには、信頼性の高い良質なデータが不可欠である。**そのような条件を満たす各種データに、必要に応じて簡便にアクセスできるデータプラットフォームが、デジタル社会における必須なインフラとして期待されている。一方でデータをAI・機械学習の学習に用いる際は、倫理、知的財産、製造物責任など多くのことを考慮しなくてはならず円滑なデータ取引を困難なものとしている。
- 一般社団法人 AIデータ活用コンソーシアム（以下「AIDC」という。）は、AIの研究と利活用において不可欠な各種データを収集し、それらを円滑かつ効率的に流通促進するためのプラットフォームとコミュニティを構築することで、日本におけるAIの研究と利活用をより一層加速させることを目標とし活動を行っている。AIDCの知的財産・契約検討ワーキンググループ（主査：渡部俊也教授。以下「本WG」という。）は、AIDCのデータ基盤を介したデータの流通において、データ提供者とデータ利用者の間に適用される契約の定型化と、当該契約の作成にかかるコストの低減化を目指し、当該契約の内容について、別紙1のメンバーで、**2019年9月から2021年11月までの期間、全24回の部会を開催し、検討を重ねてきた。**

AIDC プラットフォームにおける  
データ提供契約に関する  
報告書

初版  
2022年2月17日  
一般社団法人 AI データ活用コンソーシアム  
知的財産・契約検討ワーキンググループ

# 一般社団法人AIデータ活用コンソーシアム 知的財産・契約検討WG名簿

(本WG参加当時の肩書・所属を記載。敬称略)

## 主査

- 渡部 俊也 東京大学 未来ビジョン研究センター 教授

## 委員

- 阿部 豊隆 TMI総合法律事務所 弁理士・Cal州弁護士
- 柿山 佑人 TMI総合法律事務所 弁護士
- 榊原 颯子 TMI総合法律事務所 弁護士
- 柴野 相雄 TMI総合法律事務所 弁護士
- 下川 和男 イースト(株) 取締役会長
- 田丸 健三郎 日本マイクロソフト(株) 業務執行役員 ナショナルテクノロジーオフィサー
- 土肥 里香 TMI総合法律事務所 弁護士
- 野呂 悠登 TMI総合法律事務所 弁護士
- 舟山 聡 日本マイクロソフト(株) 業務執行役員 政策渉外・法務本部 副本部長 弁護士
- 松村 将生 (株) エクサウィザーズ 弁護士
- 村上 諭志 TMI総合法律事務所 弁護士
- 村田 真 慶應義塾大学大学院 政策・メディア研究科 特任教授
- 李 秀元 (株) ブロードバンドタワー 取締役執行役員 経営戦略・DC事業担当

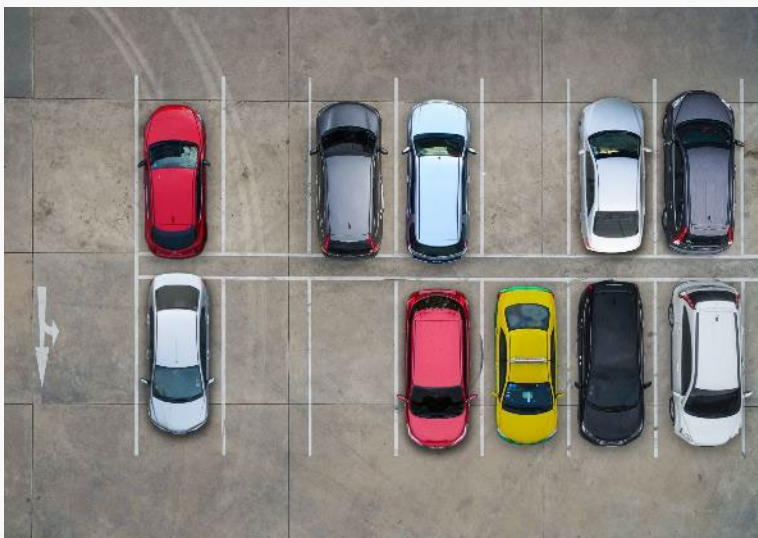
## アドバイザー

- 青木 大介 金融庁 総合政策局総合政策課フィンテック室
- 阿部 学 国立研究開発法人科学技術振興機構
- 岩島 真理 国立研究開発法人科学技術振興機構
- 菊地 陽一 特許庁 総務部企画調査課
- 佐藤 涼 国土交通省 総合政策局モビリティサービス推進課
- 関 優志 経済産業省 経済産業政策局 知的財産政策室
- 坪井 恵理佳 国立研究開発法人宇宙航空研究開発機構
- 坪内 優佳 特許庁 総務部企画調査課
- 天白 隼也 金融庁 総合政策局総合政策課 フィンテック室
- 永野 志保 特許庁 総務部企画調査課
- 西原 栄太郎 独立行政法人情報処理推進機構
- 橋本 優里花 経済産業省 経済産業政策局 知的財産政策室
- 日置 巴美 三浦法律事務所 弁護士
- 松尾 浩司 国立研究開発法人科学技術振興機構
- 三木 康平 金融庁 総合政策局総合政策課 フィンテック室
- 森川 蓉子 国立研究開発法人宇宙航空研究開発機構
- 山本 和男 内閣府 政策統括官付上席政策調査員
- 吉田 浩子 国立研究開発法人宇宙航空研究開発機構
- 渡邊 遼太郎 経済産業省 経済産業政策局 知的財産政策室 室長補佐 弁護士

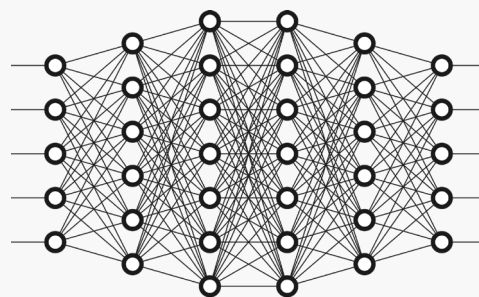
# データと知的財産



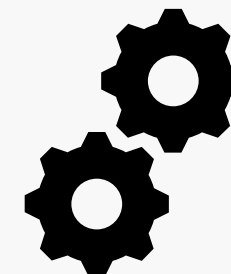
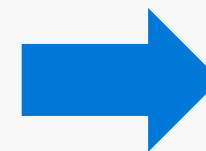
# 例) 画像とラベルの二つのデータを使用



画像データ



学習



推論



Image(ファイル名)	xmin	ymin	xmax	ymax	label
shutterstock_107834711.jpg	279.5211031	573.5263251	868.5476314	1236.99047	Worker
shutterstock_107834711.jpg	963.1650996	630.8133642	1288.328995	1238	Worker
shutterstock_68568742.jpg	8.99531494	431.8333282	466.0906185	1275	Worker
shutterstock_68568742.jpg	386.1322635	466.5000153	940.510151	1275	Worker

アノテーションデータ

# AIベンダー、エンジニアが注意すべき法的事項

## アノテーション作業、クレンジング作業

- アノテーション作業、クレンジング作業により、**どのような権利が発生**するか。
- **著作権**が発生するか？
- **著作者**は、**著作者人格権と著作権**を享有する（著作権法17条1項）。
- 著作者人格権と著作権の享有には、**いかなる方式の履行も不要**（同条2項）。

# 著作物の創作と言えるか？

- **著作者**：著作物を**創作**する者をいう（2条1項2号）。
- **著作物**：思想又は感情を**創作的に表現**したものであつて、文芸、学術、美術又は音楽の範囲に属するものをいう（同項1号）。
- アノテーション作業、クレンジング作業は、**著作物を創作**する作業といえるか。



# 「限定提供データ」としての保護

不正競争防止法2条7項

- この法律において「**限定提供データ**」とは、**業として特定の者に提供する情報として電磁的方法**（電子的方法、磁気的方法その他の他人の知覚によっては認識することができない方法をいう。次項において同じ。）により**相当量蓄積**され、及び**管理されている技術上又は営業上の情報**（**秘密として管理されているものを除く。**）をいう。

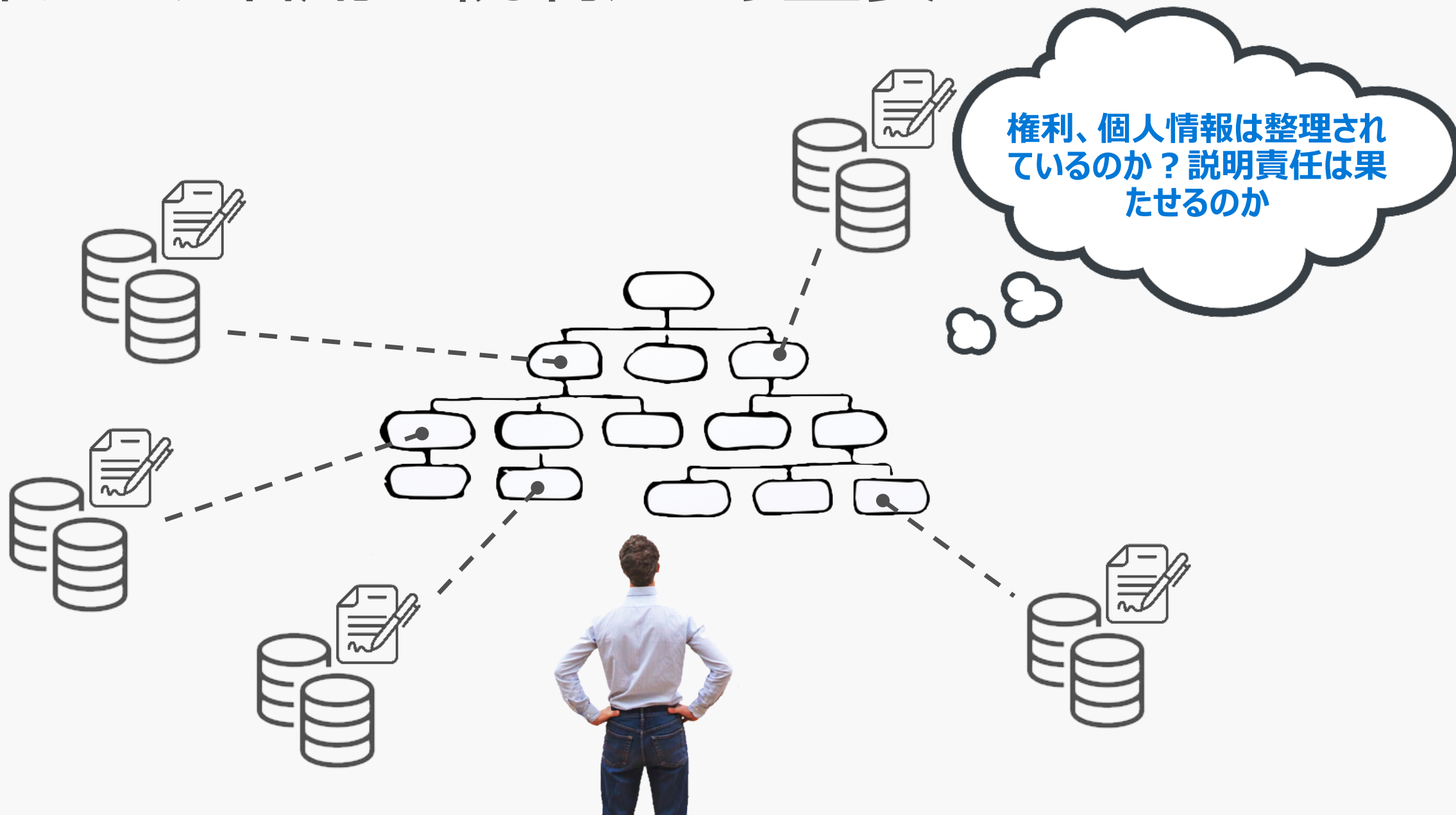
# 特別法による保護以外に保護する方法

- **契約**による保護。
- **契約**において、**権利、義務**という**ルール**を定め、それによって、自己の権利を保護する。
- **信用できる特定のプラットフォーム**上で、一定のルールに従って、**アノテーションデータ、クレンジングデータを流通させることが重要**。

# AIが変えるデータの 価値とデータ流通基 盤の構築



# 外部データ活用の統制がより重要に



# 『AIDC Data Cloud』

## 2022年3月サービス開始

- 登録無料
- 様々な課金モデルに対応
- 動的契約モデルへの対応

<https://www.aidcdatacloud.com/>

データさがす	購入	販売	このサイトについて	お問い合わせ
検索	「検索」して登録する	「販売」して登録する	ヘルプ	
申し込み (CSV 用)	購入ガイド	販売ガイド	利用規約	
検索	検索ガイド	販売ガイド	プライバシーポリシー	
PDF	購入データ	販売データ	お問い合わせ	
	お気に入り		よくある質問	
			特定商取引法に基づく表示	

# 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**

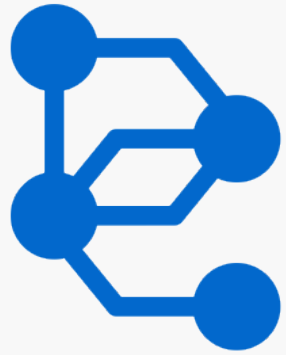


データに求められる  
**来歴と保証**



**時間**と**内訳**により  
変わるデータの**価値**

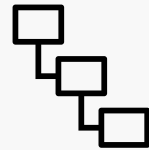
# ① 『AIDC DataCloud』が実現するデータ流通



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

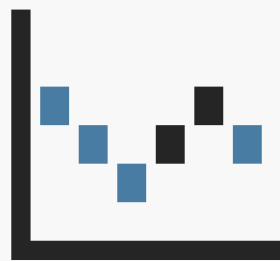
データ販売代金から基盤維持に要する費用を手数料として課金

2～3%を想定

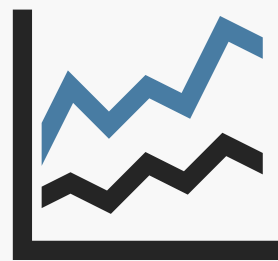
# 様々な課金モデル



期間



ワンタイム



ボリューム

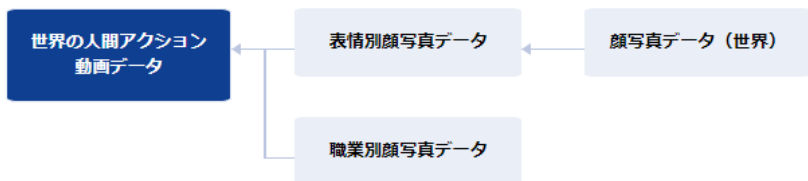


レベニューシェア



# 様々な契約条件に対応

## 本データまでの来歴



## 本データからの派生



Data Cloud

すべてのデータ ▼ キーワード 🔍

お知らせ マイデータ 販売契約 申請確認 (構築中) お問い合わせ ログアウト

ホーム / データカタログ契約項目の登録

### データカタログ契約項目の登録

販売時に契約書を作成するためにあらかじめ記載項目を登録します。

---

#### 登録データ名称

契約書掲載のデータ名

---

・契約種元入力

---

データ種類の選択  オリジナルデータ  アノテーションデータ  クレンジングデータ

データの内容を入力 (改行は出来ません)  ID  パスワード  暗号化

オリジナルデータの対象期間  例) 〇年〇月〇日～〇年〇月〇日に取得したデータ  
※ データの性質に合わせた単位 (日単位、時・分・秒単位、年・月単位、〇区、〇駅後等)、できるだけ具体的に記載してください。

データの分布・振り  例) アンケート情報について、自然言語で記述する場合「日本人以外の回答を除外している。」  
自然言語以外で記述する場合、リスト等で割合を示す。

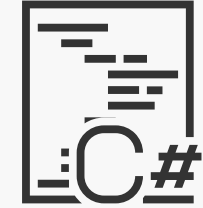
ボリューム (データ件数・数値等)  例) アンケート情報の場合、アンケート回答者の人数、地下鉄の運行データの場合、車両の数や駅の幅数、出発時刻と到着時刻の経路等。

欠損・消去の履歴  例) 地下鉄の運行データについて一定区間以降のデータを削除している場合、その旨を記載する。

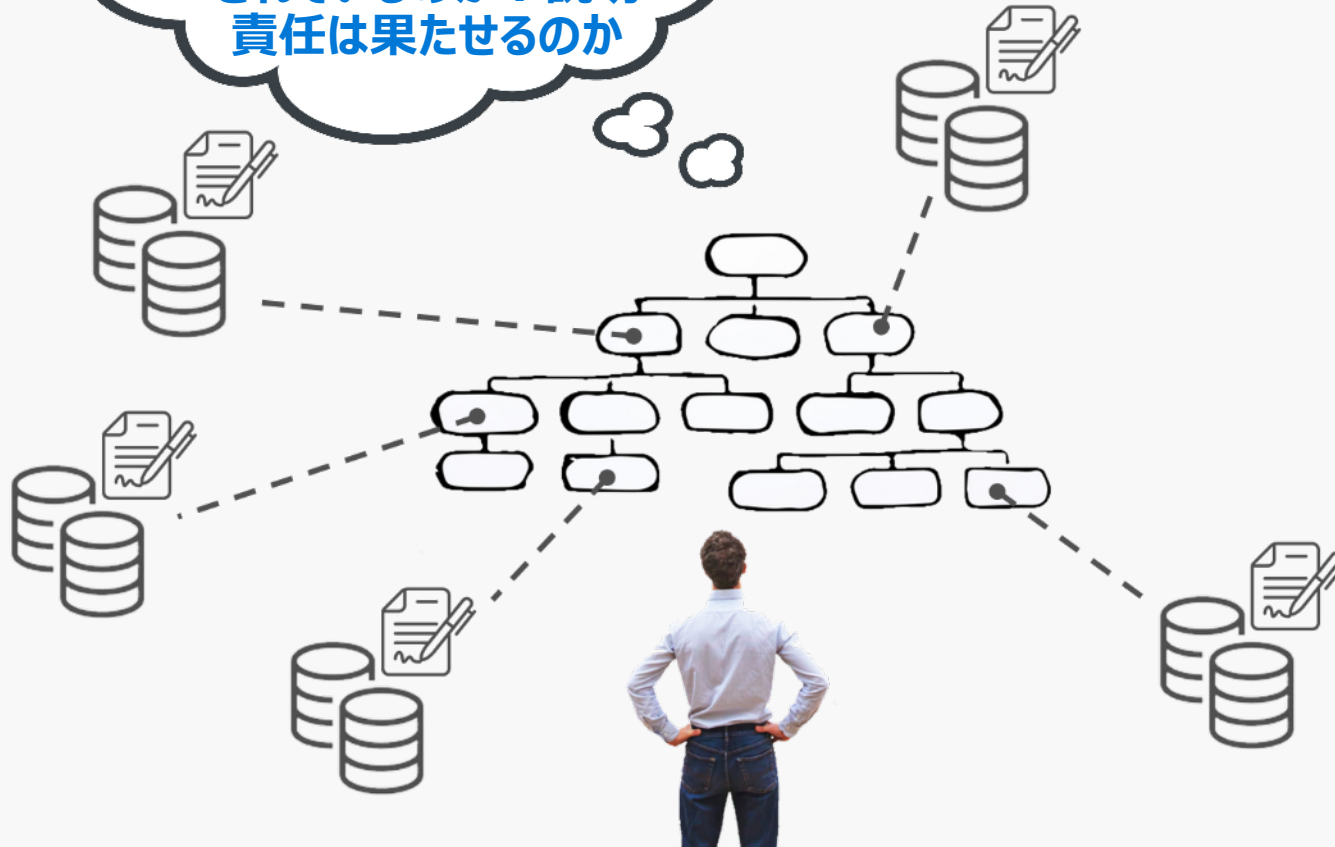
## ② 『AIDC DataCloud』を会員に無償ライセンス ～ 社内データ統制の実現を支援 ～



Data Cloud



権利、個人情報  
は整理されている  
のか？説明責任  
は果たせるのか



### 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

# 来歴の表示と派生データ対応

The screenshot shows the product page for '人間のアクション動画データセット' (Human Action Video Dataset) on the AI Data Consortium website. The page includes a search bar at the top, navigation tabs, and a main content area with a video player, a '利用を計算する' (Calculate Usage) calculator, and a 'カートに入れる' (Add to Cart) button. The price is listed as ¥330,080. Below the calculator, there are tabs for '利用' (Usage), '販売' (Sales), '解析・分析データの公開' (Data Release), and 'ライセンス' (License). The '利用' tab is selected, showing details about the dataset's version (v1.0), creator (山田 太郎), and other metadata.

The screenshot shows the Q&A section of the website. It features a question and answer format. The question asks about the average size of the data and whether it includes data from the '2020追加 (男性のみ)' (2020 Addition (Male Only)) dataset. The answer states that the average size is 20MB, but some data can be up to 100MB, and the '2020追加 (男性のみ)' dataset contains 215 videos.

## 来歴





# **AI Data Consortium**

AIデータ活用コンソーシアム

<https://aidata.org>