資料2

AI学習データの提供促進に向けたアクションプラン ver1.0

2023.11.7

内閣府 科学技術・イノベーション推進事務局

1. はじめに

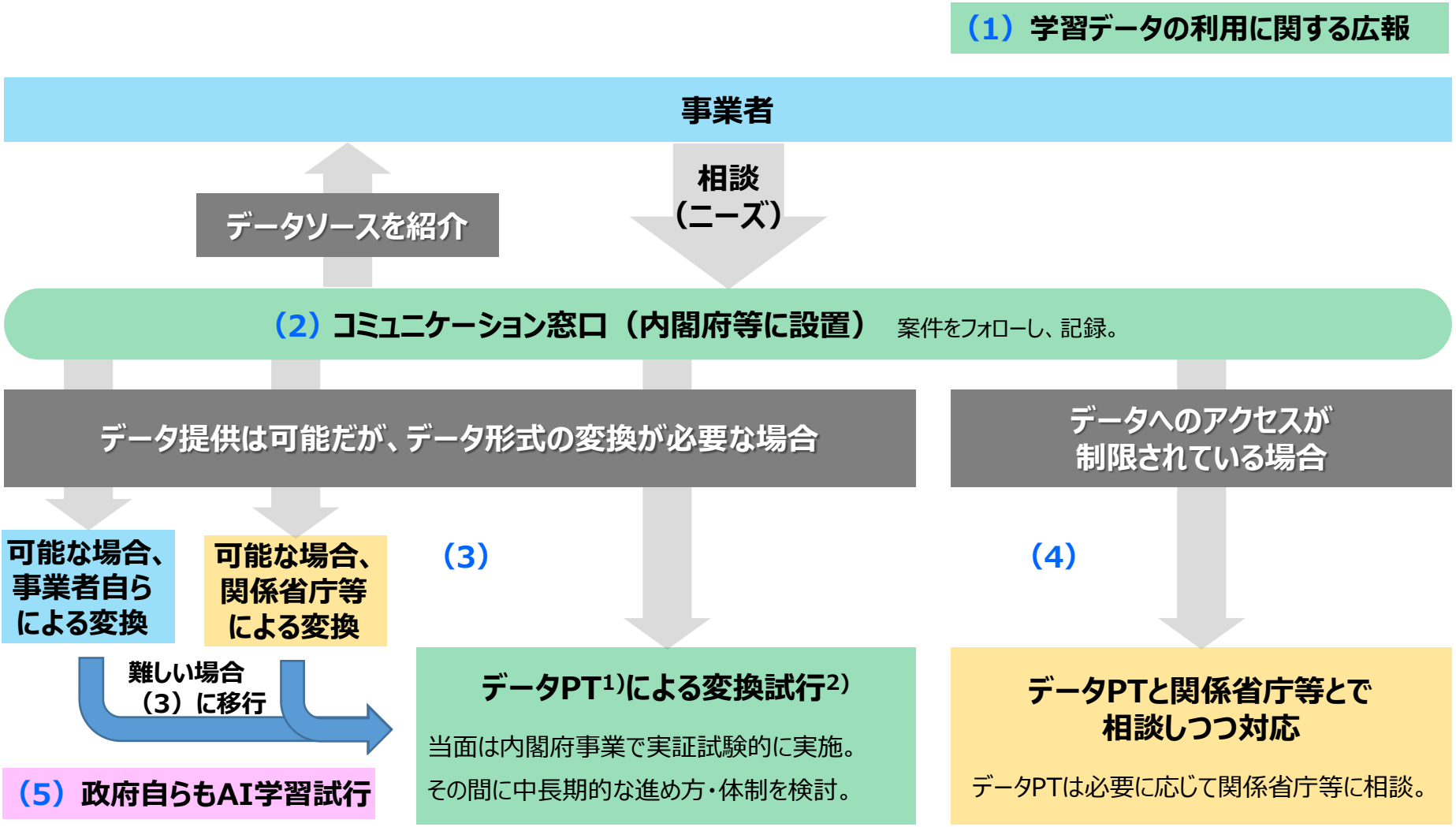
AIモデルの性能には、AIが学習するデータの量と質が影響する。
また、今後、多様なAIモデルが開発され、多様なデータが必要になると考えられる。

政府等が保有するデータの多くは、
①作成者、作成時期、作成場所等が明確。
②著作権などの権利処理が不要。公開、二次利用等の承認が得られている。
③正確な情報（内容・文法が正しい）。
④不適切な情報を含まない。個人情報が含まれていない、匿名化処理されている。
を満たすと考えられ、学習データとして有用と考えられる。
また、多様な分野のデータがあり、AI学習に利用できる可能性が高い。

一方で、政府等保有データは、pdf形式で公表され、直ちにAI学習に用いることが難しい場合もある。
また、データへのアクセスが制限されているケースもみられる。

このため、政府等保有データに関して、AI開発者からのニーズに応じてAI学習データとしての提供を促進するため、ここにアクションプランを策定した。

2. アクションプランの概要



1) 内閣府、デジタル庁によるプロジェクトチーム（PT）をAI戦略チーム下に設置。

2) 膨大なデータを扱うため、完璧な変換は技術的に難しいことに留意が必要。

3. (1) 政府等保有データのAI学習データへの提供に関する広報

目標

○様々な機会を用いて、政府等保有データのAI学習データへの提供促進について広報し、日本のAIモデル開発力の強化に資する。

取組み

○政府に登録された2万2千を超えるデータセット等を紹介するe-Govデータポータルのリンクを、AI戦略会議、AI戦略チームのHPなどに掲載【内閣府、デジタル庁】

○政府等保有データについて、e-Govデータポータルへの登録を関係省庁に呼びかけ【内閣府、デジタル庁】

○AI関係者が集まる各種のイベント等において、e-Govデータポータル、AI Japan（国研）のデータポータル、内閣府等のコミュニケーション窓口を広報【内閣府、デジタル庁】

○国研の利用可能なデータのうち、ダウンロード可能なURLのリスト化、国研のWEBサイトへの掲載【関係省庁等】

3. (2) AI開発者向けのコミュニケーション窓口の設置

目標

○AI開発者からのAI学習データに関する相談（ニーズ）を受け付け、対応する[1)]。

○事業者ニーズ等を集約し、今後の政策検討に貢献する。

取組み

○AI開発者向けに、専門性の高い職員によって構成されるコミュニケーション窓口を内閣府に開設し（業務フローを用意）、AI学習データの提供等を希望する者からの相談（ニーズ）等を受け付け、関係省庁等と連携して対応【内閣府】

○関係省庁や関係機関にも同様の窓口の開設を働きかけ【内閣府、関係省庁】

○窓口に寄せられたAI学習データに関する事業者ニーズ等の情報を把握・蓄積し、分析【内閣府】

1) 当面は、実証試験的な観点から有意義と考えられる案件を優先する可能性がある。また、AI開発者からの相談が処理能力を超えた場合、対応をお断りする可能性がある。

3. (3) マシンリーダブルでないデータの形式変換

目標

○AI学習用データとしてのニーズがあるデータに関して、マシンリーダブルでない（PDFやJPEG形式の）データをマシンリーダブルな（テキストやhtml）形式に変換する。

取組み

○AI戦略チームの傘下にデータPTを設置【内閣府、デジタル庁】

○マシンリーダブルなデータ形式の定義の明確化【データPT】

○公開データのマシンリーダブルな形式への変換を内閣府事業として試行、その他の省庁等が協力可能なケースや具体的進め方についても検討 【データPT、国立印刷局】

○試行結果等も踏まえ、中長期的なデータ形式変換の進め方・体制、費用等のルールを検討し、事業スキーム等を構築【データPT、国立印刷局】

○政府等保有データのうち、AI学習データとしてのニーズが具体的に寄せられたデータに関して、マシンリーダブルになっていないデータを分類・整理し、リスト化【データPT、関係省庁】

○マシンリーダブルな形式への変換を政府全体の取組みとすることを目指す【内閣府、デジタル庁】

3. (4) アクセス制限のあるデータに関して、適格な申請者に対する迅速な提供

＜目標＞

○AI学習データとしてのニーズがあるデータに関して、アクセス制限の有無を把握し、アクセス可能な場合には、適格な申請者に対しては迅速に提供する。

＜取組み＞

○先行的事例として、NICT保有の日本語データに関して、年明けを目途に共同研究の形で提供を開始できるよう検討中【総務省】

○政府等保有データのうち、 AI学習データとしてのニーズが具体的に寄せられたものに関して、アクセス制限の有無を確認して、アクセス可否を検討【データPT、関係省庁】

○アクセス可能な場合には、アクセスに必要な条件・書類等を策定し、適格な申請者に対しては迅速に提供【データPT、関係省庁】

3. (5) 政府自らもAI学習を試行

＜目標＞

○業務効率化、サービス向上、より創造的な業務への挑戦（政策的な文書案の作成等）

＜取組み＞

○政府自らも公的機関が保有するデータ等を用いてAI学習を試行【デジタル庁】

○政府が関与するプロジェクトや事業の中でAI学習を実施する場合、AI学習データに関する課題とその解決方法等に関する知見を集約【関係省庁】

4. 工程表（現時点での想定）

2023		2024		2025	
11	12	1	6	1	4

アクションプラン

統合イノベーション戦略2024

(1) 学習データに関する広報

(2) コミュニケーション窓口の設置

(3) データ形式変換の試行※ → 本格実施

(4) アクセス制限のあるデータの提供（可否の検討、条件等の策定）

(5) 政府自らもAI学習

※ マシーンリーダブルな形式への変換については、「デジタル社会の実現に向けた重点計画」等への反映を含め政府の取組みを今後検討予定。