

A I を用いたクラウドサービスに関するガイドブック

令和4年2月15日

総務省 情報流通行政局

デジタル企業行動室

目 次

はじめに.....	3
1. 本ガイドブックの目的・背景.....	3
2. 本ガイドブックの対象等	4
2. 1 本ガイドブックの対象.....	4
2. 2 A Iクラウドサービスの特徴.....	5
2. 3 想定する読者	5
2. 4 使われ方.....	6
2. 5 用語の定義.....	6
第 I 編 AIクラウドサービスの開発	8
1. AI開発のプロセス.....	8
2. 企画・計画フェーズ	9
3. 開発準備フェーズ	14
4. 開発フェーズ.....	16
4. 1 データの整備	16
4. 2 モデルの構築から学習、評価まで	21
5. 個別調整フェーズ	28
6. 運用・保守フェーズ	30
7. その他.....	32
7. 1 外部委託と外部連携	32
7. 2 環境問題への配慮	33
7. 3 A Iのエコシステムについて.....	33
7. 4 A Iの創作物に関する著作権の問題.....	33
第 II 編 AIクラウドサービスの提供	35
1. 提供前の段階で推奨される取組	35
2. 提供期間中において推奨される取組	40

はじめに

1. 本ガイドブックの目的・背景

現在、人工知能（ＡＩ）に関しては、技術水準が向上しつつあるのみならず、既に様々な商品・サービスに組み込まれ、ビジネスや生活といった様々な場面での社会実装が本格化しつつある。このような背景の中、普及が加速しているクラウドサービスに関しては、クラウドサービス事業者において、ＡＩの活用による自社サービスの付加価値向上を目指す取り組みが活発化しつつあり、今後もクラウドサービスを通じたＡＩの更なる普及が期待されるが、現実にはクラウドサービス事業者の３割程度しかＡＩを活用していない。調査の結果、クラウドサービス事業者は、ＡＩを導入したくても方法がわからず、また、様々な課題をかかえていることがわかってきた。

一方、ＡＩ機能を用いたクラウドサービス（以下、「ＡＩクラウドサービス」という。）の提供に当たっては、提供者においてＡＩの出力結果の精度等を利用者に約束することが難しいケースが多く、利用者にとっても、自身がリスクを負うサービスの導入には慎重にならざるをえない現状がある。また、利用者からの信頼獲得とトラブル発生の防止の観点において、提供者は、ＡＩの特性を踏まえつつ、自社のサービスの機能に係る性質や責任を利用者に正しく理解させる必要があるが、１つのサービスに多数の利用者が紐づくクラウドサービスにおいて、適切かつ効果的な情報提供の在り方が不明確であるという課題も存在する。

本ガイドブックは、これらの課題を克服するために作成したものであり、以下の２つを目的としている。

- ① ＡＩクラウドサービスの効果的・効率的な実現方法について具体的な開発の流れを体系的に整理するとともに、開発の各プロセスにおいて開発者が留意すべき点をまとめて提示すること。
- ② クラウドサービス事業者が利用者にＡＩクラウドサービスを提供する際に、利用者の信頼を獲得しつつ市場を拡大していくために推奨される自主的な取組を提言すること。

クラウドサービス事業者が本ガイドブックを参照することにより、ＡＩを使った新しいサービスや事業を開発するための指針となり、利用者に対する適切な措置を自主的に講じることにより、利用者を含む様々なステークホルダからの信頼が醸成され、それがＡＩやクラウドサービスの更なる需要につながり、国内におけるＡＩクラウドサービスの社会実装が促進されることを期待する。

本ガイドブックは上述のとおり、A I クラウドサービスを提供する者が、利用者の信頼を獲得しつつ市場を拡大していくために推奨される自主的な取組をガイドするものであり、何ら法的な拘束力を有するものではなく、また、当事者間における交渉や契約の自由を何ら制約するものではないことを付言する。

なお、A I の技術は現在も進化の途上にあるため、本ガイドブックの内容についても、今後の見直しにより、変更することがあることに留意されたい。

2. 本ガイドブックの対象等

2. 1 本ガイドブックの対象

本ガイドブックは、A I 機能を用いたクラウドサービスを対象とする。また、クラウドサービスのうちS a a S (Software as a Service) を対象とし、P a a S や I a a S は含まないものとする。

また、本ガイドブックでは、A I クラウドサービスの対象として、最終的な利用者が企業であるもの(B to B)を想定する。



<A I クラウドサービスの例>

- ・ 財務会計（クラウドサービス）＋不良債権の自動検知（A I 機能）
- ・ 名刺管理（クラウドサービス）＋文字認識（A I 機能）
- ・ 勤怠管理（クラウドサービス）＋退職予測（A I 機能）
- ・ 販売管理（クラウドサービス）＋クロスセルのアドバイス（A I 機能）
- ・ 故障管理（クラウドサービス）＋予防保全（A I 機能）
- ・ 上下水道の管理（クラウドサービス）＋漏洩危険個所の予測（A I 機能）

2. 2 AIクラウドサービスの特徴

1) 顧客別の受託開発のAIとの異なる特徴

AIクラウドサービスは、委託を受けて顧客別に開発するAIと比較して、以下のような特徴がある。

- ① 複数の顧客に提供するサービスであること。
- ② AIクラウドサービス提供事業者自らの投資で開発すること。（受託開発によるAIでは顧客が投資する。）
- ③ システムの要求条件や仕様はAIクラウドサービス提供事業者自らが決定すること。
- ④ 開発後にAIクラウドサービスを利用したい者が利用を申し込むという契約スタイルであること。
- ⑤ オープンなネットワークを使用すること。

2) 通常のクラウドサービスとの異なる特徴

通常のクラウドサービスと比較して、AIクラウドサービスには以下のような特徴がある。

- ① システムのアウトプットは「90%晴れる。」のように確率的な結果になり、常に同じとは限らない。
- ② アウトプットの根拠や理由の説明が難しい。ブラックボックス的になっていることが多い。
- ③ データ（学習用データおよび入力データ）がアウトプットに大きく影響する。
- ④ アウトプットの利用に人間の判断が入らないものがある。

2. 3 想定する読者

本ガイドブックは、読者としてクラウドサービス事業者を想定している。特にこれからAIクラウドサービスを提供しようと考えている企業を対象とし、すでにAIクラウドサービスを提供している企業にとっても課題解決の参考となることを期待している。また、企業の経営層から営業担当者、開発担当者等の幅広い読者を想定している。

2. 4 使われ方

本ガイドブックは、クラウドサービス事業者がA Iクラウドサービスを利用者に提供するに当たり、利用者からの信頼獲得とトラブル発生の防止の観点から、提供者において取り組むことが推奨される事項についてガイドするものである。

クラウドサービス事業者がこれらのガイドを、A Iクラウドサービスを利用者に提供する際の各プロセスに反映することで、A Iクラウドサービスの需要拡大と社会実装の加速が実現されることを期待する。

2. 5 用語の定義

本ガイドブックで用いる用語の定義は、以下に示すとおりとする。

なお、A Iの基本的な用語については、「A Iネットワーク社会推進会議」（平成28年度～）において作成された「国際的な議論のためのA I開発ガイドライン案」（平成29年7月28日 A Iネットワーク社会推進会議）（以下、「A I開発ガイドライン案」という。）の定義に従うものとする。

A I機能 : A Iの用途（A Iに実行させること）である「画像認識」、「分類」、「予測」、「文章生成」、「マッチング」、「会話」といった機能面を表現する場合には「A I機能」という言葉を用いる。なお、A Iそのものの定義には様々なものがあるが、本ガイドブックにおいては、機械学習を用いたA Iを念頭にしている。

事業者、顧客、利用者 : 「事業者」とは、クラウドサービス事業者を指し、「顧客」とは、クラウドサービスの利用を申し込んだ企業等を指す。また、実際にA I機能を使って作業を行う者（エンドユーザ）を「利用者」と呼ぶ。利用者は顧客の従業員に限らない。

精度 : A Iの出す結果が、どの程度正確であるかを表わすときに「精度」を用いる。なお、「品質」は、広い意味を持ち、人によって定義がまちまちであり、「性能」は、レスポンスタイムやスループットのような処理時間に関係する事柄に用いられることが多いため、本ガイドブックでは、正確性の意味では用いないものとする。

P o C (Proof of Concept) : A Iの開発に先立って、新しい技術やアイデアなどの実現可能性を確かめるために行う実証実験であり、概念実証ともいう。実運用で使用するA Iとは規模や使用環境が異なる。P o Cが失敗に終わった場合、本格的な開発には進まないという意思決定をすることになる。

トライアル : A I の開発後、サービスの販売を行う過程において、正式な契約をする前に、利用者が試しに使ってみる機会を設けることを「トライアル」と呼ぶ。

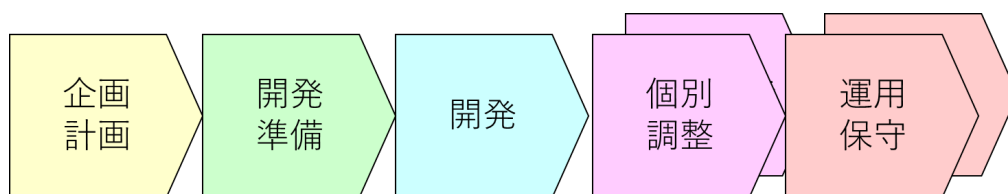
第Ⅰ編 AIクラウドサービスの開発

第Ⅰ編では、AIクラウドサービスを開発する際に留意すべき点をガイドする。

1. AI開発のプロセス

AIクラウドサービスの開発における課題や解決策を検討するにあたり、AIの開発プロセスのモデルを明らかにし、プロセスごとに課題や解決策を整理する。

本ガイドブックでは、AIクラウドサービスの開発プロセスを以下の5つのフェーズに整理する。



各フェーズの概要は以下の通り。

企画・計画フェーズ : 開発するAI機能を検討し、計画を立てる。人や資金などのリソースの割り当てを行い、開発の意思決定をする。

開発準備フェーズ : AIの開発環境を整える。新しい技術等についてはP o Cを行う。

開発フェーズ : データの収集・加工を行い、モデルの構築・（機械）学習・テストを行う。これらの作業は何度か繰り返し行われる。

個別調整フェーズ : 利用に先立って顧客がAI機能の精度を確認し、必要に応じて顧客別に追加学習等の調整を行う。

運用・保守フェーズ : 運用状況のモニタリングを行い、必要に応じて改善（モデルの修正、追加学習など）を行う。AI機能を使った作業等の運用自体は各顧客が行い、事業者は、その状況をモニタリングすることを想定している。

2. 企画・計画フェーズ

2. 1 実施事項

企画・計画フェーズでは、開発するA I 機能を検討し、A I 機能を実現するためのリソース配分を検討し、スケジュール、組織計画、資金計画等の実行計画を策定する。企画・計画フェーズの最終段階では、開発を行う意思決定を行う。

2. 2 留意すべき事項

1) A I の企画の目的

A I を企画するにあたっては以下のような点に留意する必要がある。

① A I を作ることが目的にならないようにすること。

クラウドサービスにA I 機能を追加する場合、「とにかく何でもいいからA I 機能を追加する。」など、A I を使うこと自体が目的になると、顧客のニーズと乖離し、追加したA I が使われないものになってしまう恐れがあるため、顧客のニーズや付加価値をベースとした企画を作成する必要がある。顧客のニーズを満たし、A I 機能を導入することが適切なものを選択する必要がある。

② A I ポリシーと整合性のとれた企画とすること。

企業は、どんなA I を作ってもよいというものではない。例えば、犯罪目的のA I の作成は許されない。企業のA I への取り組みにおける倫理についても整理して統制を行う必要がある。

A I を提供する企業は、A I への取り組みにおける基本的な考え方（以下、「A I ポリシー」という。）をとりまとめて公表する企業が増えている。「基本方針」、「指針」、「コミットメント」等の名前で呼ばれることもある。A I ポリシーをまとめて社内外に示すことは、企業の信用力の向上や社員への意識付けになるため、実施することが望ましい。

A I クラウドサービスを開発する各プロジェクトにおいては、このA I ポリシーに則った企画を行う必要がある。

A I ポリシーを策定するにあたっては、「人間中心のA I 社会原則」（平成 31 年 3

月 29 日 統合イノベーション戦略推進会議決定)、「A I 開発ガイドライン案」¹、「A I 利活用ガイドライン」(令和元年 8 月 9 日 A I ネットワーク社会推進会議)各原則を遵守する必要がある。

2) データの整備計画の重要性

A I の開発においては、学習に使用するデータが極めて重要であり、収集方法や収集範囲等について早くから十分な検討を行う必要がある。また、データの収集や加工に多くの稼働やコストを要するため、スケジュールの作成においてはデータの整備に十分な時間を見込んでおく必要がある。

3) クラウドサービスのシステム全体との関係

A I 機能を追加するにあたっては、追加部分だけでなくその関連部分、さらにはクラウドサービス全体に及ぼす影響などの検討を十分行う必要がある。

例えば、A I で利用するために既存のデータのフォーマットを変更する必要がある場合は、データベースの仕様まで変更するかどうかの検討が必要になる。

4) 採算性の確保

A I 開発のプロジェクトは、完成に至るまでのリスクが高い。P o C で終わってしまい、本格的な開発に至らないものも多い。A I クラウドサービスをビジネスとしてとらえた場合、採算性の確保も大きな課題である。採算性を確保するためには以下のようなアプローチが必要となる。

① Value Based アプローチ

企画・計画の段階で、顧客の付加価値（この機能が追加になると、顧客にはどのような価値が生まれるか。）を強く意識する必要がある。料金を検討する際は、顧客の人件費の削減に A I が寄与する等、顧客の付加価値が金銭的に評価できるとわかりやすい。開発にかかった費用をサービス料に上乗せするという料金体系では、顧客から支持されない可能性もある。料金体系についても顧客の付加価値をベースにして検討する必要がある。

¹ 「A I 開発ガイドライン案」の開発原則は、以下の 9 つである。

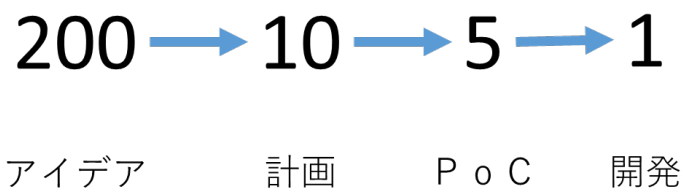
① 連携の原則、② 透明性の原則、③ 制御可能性の原則、④ 安全の原則、⑤ セキュリティの原則、⑥ プライバシーの原則、⑦ 倫理の原則、⑧ 利用者支援の原則、⑨ アカウンタビリティの原則

② 資金回収計画の立案

A I 開発の失敗事例には、投資の検討はしても回収の計画を立てていなかったことによるものがある。採算性リスクの軽減のためには、A I 開発においても毎回資金回収を意識した計画を作成して、実施結果のフィードバックを行うことが必要である。当初は経験不足により計画の精度が低くても、経験とともに精度を向上させるしくみを作っておくことが重要である。開発に係る人件費や環境整備費用の見積もりなどは経験とともに精度が向上する。

③ 多くのP o Cに挑戦できる意思決定プロセス

A I 開発におけるP o Cは失敗する確率が高く、実運用まで進むものが少ないが、1つのP o Cの失敗でA Iをあきらめるのではなく、多くのP o Cに挑戦できるような意思決定のプロセスが必要となる。

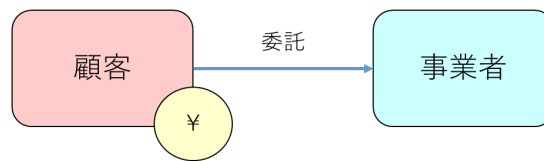


5) P o Cの計画について

P o Cは、新しい技術やアイデア等の実現可能性を検証するために実施するものである。A I 分野においては新しい技術が次々と開発されている状況にあり、P o Cによる実現可能性の検証はリスクの軽減という観点から必要なものである。

一方、受託開発によるA Iとは異なり、A IクラウドサービスのためのA I機能の開発は、事業者自身の投資によるものであることを考えれば、P o Cの実施については十分な検討が必要である。

< 受託開発の場合 >



< クラウドサービスの場合 >



受託開発との違い

AIで成功している企業へのヒアリングから、以下のような点への配慮が必要であることが判明している。

① P o C後のリソース配分についても計画しておくこと。

P o Cが失敗した場合はプロジェクトを中止する前提で、予めP o Cまでのリソースしか用意していなかった場合、たとえP o Cが成功したとしてもP o Cで終了せざるを得なくなることがある。P o C後のリソース配分についても計画しておく必要がある。

② P o Cの投資を小さくする工夫をすること。

P o Cの失敗はそのまま損失につながるため、できるだけ投資を小さくする工夫が必要である。具体的には、オープンなモデルの再利用やOPM（補助金、顧客による一部負担等）の活用が考えられる。また、普段から少なめの投資で研究開発を行い、実現可能性が見えた段階で大きな投資をするという工夫をしている企業もある。

6) AI 機能の分類

AI の企画において、フレームワーク思考により AI 機能を分類することは有効である。例えば、①認識、②予測・判断、③コミュニケーション、④制御、⑤生成の5つに分類することができる。

なお、将来さらに新しい種類の AI 機能が登場する可能性があるため、分類はこの限りではない。

- ① 認識（画像認識、音声認識など）
- ② 予測・判断（売上予測、シミュレーションなど）
- ③ コミュニケーション（自然言語処理、会話、翻訳など）
- ④ 制御（機械の制御など）
- ⑤ 生成（画像生成、文章生成など）

もう1つの軸としては、「人間の代替をするもの」、「人間の能力を拡張するもの」に分類することもできる。

7) AI クラウドサービスの事例

AI の企画においては、先進事例を収集してアイデアのヒントを得ることも有効である。

現在、AI クラウドサービスとして、以下のようなものが提供されている。

- ・ 会計・経理・財務（資金繰りアラームなど）
- ・ 人事・給与（退職予測など）
- ・ マーケティング（広告挿入など）
- ・ 販売（レコメンドなど）
- ・ 診断（欠陥の発見など）
- ・ 自動翻訳

3. 開発準備フェーズ

3. 1 実施事項

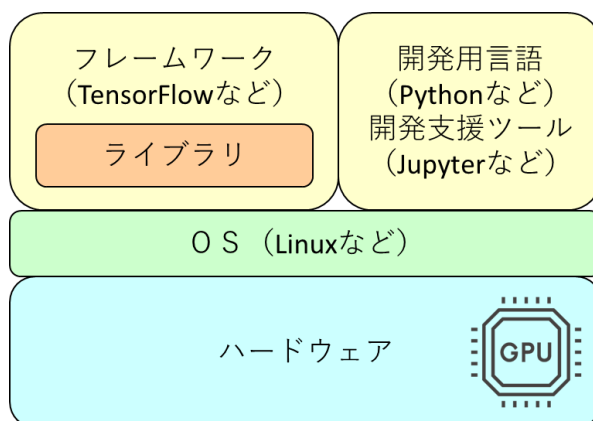
開発準備フェーズにおいては、A I 機能を開発するためのハードウェアおよびソフトウェアを準備する。また、新しい技術等の実現可能性の検証のためにP o Cを実施する。

3. 2 留意すべき事項

1) A I の開発環境の整備

A I 開発（特に機械学習）においては、G P U等を使った高速の演算処理機器を必要とする。高額ハードウェアの導入はビジネスの採算性にも影響するため、見積りの精度を高める必要がある。G P Uを共有サーバに入れて複数の開発メンバーで利用する場合は、利用の競合に配慮して導入数を決める必要がある。また、学習等に大量のデータを使う場合は、十分なストレージを用意する必要がある。

A I のソフトウェア開発環境は、フレームワークやライブラリが充実しており、1から作らなくてもよいが、環境の構築には時間を要する。



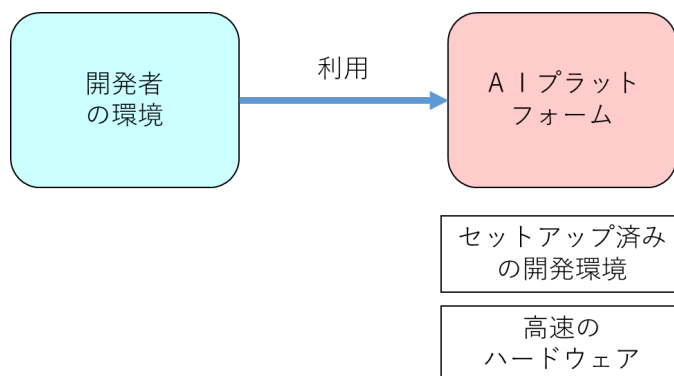
A I の開発環境例

2) AIプラットフォームの活用

AIの開発環境を提供するAIプラットフォームサービスが各社から提供されている。AIプラットフォームでは、GPU等を使った高速のハードウェアが用意されており、フレームワーク、ライブラリ等をインストール済みであり、開発支援ツールもそろっている。また、画像認識等の基本機能の各種APIも提供されている。

① P o CにおけるAIプラットフォームの活用

実験的な段階の開発環境については、開発効率や採算性を考慮し、必要に応じて、AIプラットフォームの利用が有効と考えられる。



② マルチクラウドの場合の留意点

AIプラットフォームを活用する場合は、マルチクラウド環境になることも多い。現在、マルチクラウドの利用は、クラウド全体の2割程度まで増えてきており、その課題も明らかになってきている。

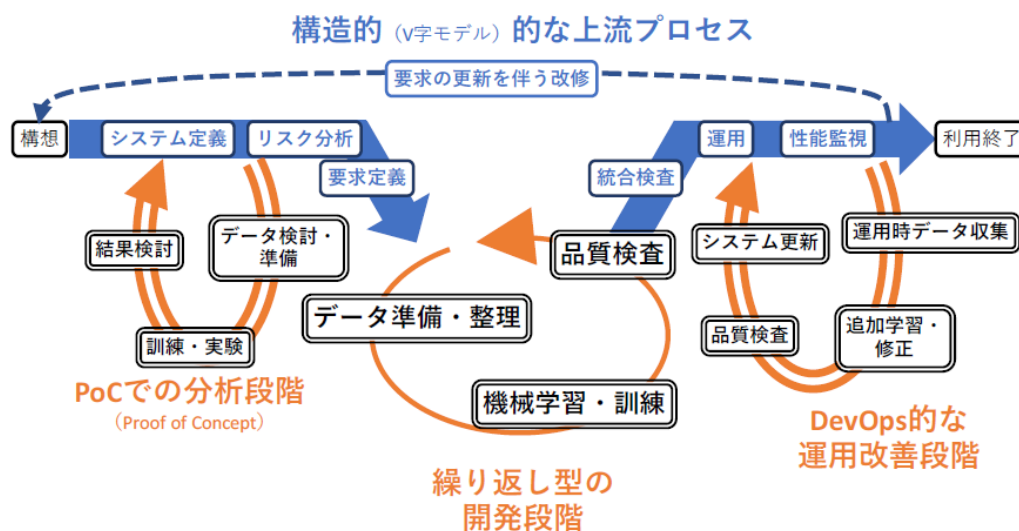
AI開発のために、マルチクラウド環境が必要になる場合は、通常のマルチクラウド環境と同様に、運用管理が複雑になる点やコストが増大する可能性についても十分検討する必要がある。

3) P o Cの実施

P o Cにおける作業については、開発フェーズにおける作業とほぼ同じものを簡略化し、小さい環境で行うものであるため、開発フェーズのところで記述する。

4. 開発フェーズ

開発フェーズは、データの整備、モデル構築、機械学習、テストなどを繰り返して行うフェーズである。



国立研究開発法人産業技術総合研究所「機械学習品質マネジメントガイドライン第2版 (2021年7月5日)」

4. 1 データの整備

4. 1. 1 実施事項

データの整備においては、機械学習に必要なデータを収集し、学習できる形式に加工する。また、学習用のデータと、評価用データやテスト用データを分けることも必要である。

4. 1. 2 留意すべき事項

AIの開発プロセス全体の中で、かなりの部分をデータ関連の作業が占めている。クラウドサービス事業者へのアンケート調査でも、データ収集を課題と考えている企業が最も多く、データが手に入らない、データの入手方法がわからない、データの収集や加工にかなりの時間がかかるという課題がある。

1) 適切なデータの収集方法の選択

データの収集方法はデータの種類によっても異なるが、以下のような方法の組み合わせによって行われている。

- ① 既存の顧客データを活用する。
- ② 新たにセンサーを設置して収集する。
- ③ 人海戦術で手入力する。
- ④ インターネットで自動収集する。(スクレイピングという。)
- ⑤ 公開されているデータセットを利用する。
- ⑥ データ提供事業者から購入する。

データの収集にあたっては、データの量や品質、収集コスト、収集時間などを考慮して適切な方法を選択する必要がある。

2) データの偏り防止

A I の出力結果は、収集するデータの偏りに大きく左右される。A I がどの程度正しい結果を出せるかという正確性の問題だけでなく、差別的な結果が出力されるなど、結果の公平性に影響を及ぼすこともある。データ収集においては、データの範囲・種類・分布などに留意する必要がある。「A I 利活用ガイドライン」においても、「適正学習の原則」及び「公平性の原則」において、データの質や代表性に留意すべきことが述べられている。

3) データ収集と著作権

日本の著作権法²は世界的に見ても進んでおり、学習用データの収集の自由がかなり認められている。具体的には、A I の学習データの収集には著作権者の承諾が不要になっている。著作権法を正しく理解することにより、効率的なデータ収集が可能となる。

なお、顧客のデータの二次利用については、著作権法上問題がない場合であっても、

² 著作権法（昭和四十五年法律第四十八号）第30条の4 著作物は、次に掲げる場合その他の当該著作物に表現された思想又は感情を自ら享受し又は他人に享受させることを目的としない場合には、その必要と認められる限度において、いずれの方法によるかを問わず、利用することができる。ただし、当該著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

二 情報解析（多数の著作物その他の大量の情報から、当該情報を構成する言語、音、映像その他の要素に係る情報を抽出し、比較、分類その他の解析を行うことをいう。第四十七条の五第一項第二号において同じ。）の用に供する場合

利用規約等で顧客の承諾を得ておくことが望ましい。

また、画像や映像データの場合、著作権の他に肖像権についても留意する必要がある。

4) データに関する法令の遵守

データ収集において特に留意すべき法令としては個人情報保護法と不正競争防止法がある。

<個人情報保護法>

収集するデータに「個人情報」が含まれている場合は、個人情報保護法の規程³⁴を遵守すること。

「個人情報」とは、生存する個人に関する情報であつて、①当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別できるもの、または②個人識別符号が含まれるものである。

<不正競争防止法>

データに「営業秘密」が含まれている場合は、不正競争防止法の規程を遵守すること。

不正な手段により営業秘密を取得し、不正取得した営業秘密を使用した場合、差止請求や損害賠償請求を受けることになる。

不正競争防止法における「営業秘密」とは、①秘密管理性、②有用性、③非公知性の3つの要件を満たすものである。

「秘密管理性」については、主観的に秘密と認識しているだけでは足りず、秘密を秘密でないものと区分して管理し、マル秘を表示するなど、営業秘密であることを明らかにすることが必要とされている。

AIの場合、「有用性」は満たされると考えられる。「非公知性」は、情報の保有者の管理下以外では一般的に入手できない状態を指す⁵。

<その他の法令等>

³ 個人情報保護法（平成十五年法律第五十七号）の主な関連条文：①個人情報（個人情報保護法2条1項）、②個人データ（同条6項）、③保有個人データ（同条7項）、④要配慮個人情報（同条3項）及び⑤匿名加工情報（同条9項）

⁴ 個人情報が含まれている場合のデータの取扱：①利用目的の特定、②利用目的の範囲内での利用、③利用目的の通知・公表、④不正の手段による取得をしない、⑤苦情処理の適切・迅速化とその体制整備のための努力

⁵ 福岡真之介「AIの法律」より。

上記の法令の他に、分野ごとにデータに関する規則があるのでそれらに従う必要がある。例えば、通信分野では、「通信の秘密」が日本国憲法（昭和二十一年憲法）第21条第2項や電気通信事業法（昭和五十九年法律第八十六号）第4条に規定されている。金融分野においては、銀行法（昭和五十六年法律第五十九号）、保険業法（平成七年法律第百五号）、金融商品取引法（昭和二十三年法律第二十五号）や「金融分野における個人情報保護に関するガイドライン」（平成29年2月 個人情報保護委員会・金融庁）などの規定にも留意する必要がある。医療分野においては「医療情報システムの安全管理に関するガイドライン第5.1版」（令和3年1月29日 厚生労働省）や「医療情報を取り扱う情報システム・サービスの提供事業者における安全管理ガイドライン」（令和2年8月21日総務省・経済産業省）等がある。

5) データの不具合に対する責任

収集したデータに不具合があり、損害が発生した場合の責任分担についても意識する必要がある。

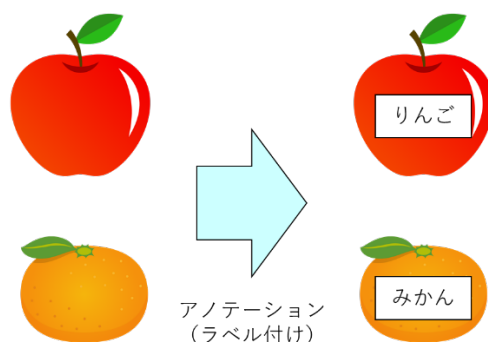
- ① データをデータ提供事業者から購入する場合は、購買契約における責任分担の規程に留意すること。

通常はデータ提供事業者の免責事項になっていることが多い。

- ② インターネット等でオープンなデータを収集する場合は、データの不具合についての責任も自社になることに留意すること。

6) データの加工における留意点

データの加工にもかなりの稼働がかかることを意識しておく必要がある。例えば、教師あり学習⁶を行う場合に、学習用データにラベル付けを行う作業（アノテーション）などには多くの時間がかかる。その他にも、データの欠損、形式の違い、単位の違いなど多くの原因によりデータの加工が必要になる。特に現場のデータは、AI等での活用を想定して作成されていないことが多い。データの加工をいかに効率化するかは課題の1つである。



データの加工の例

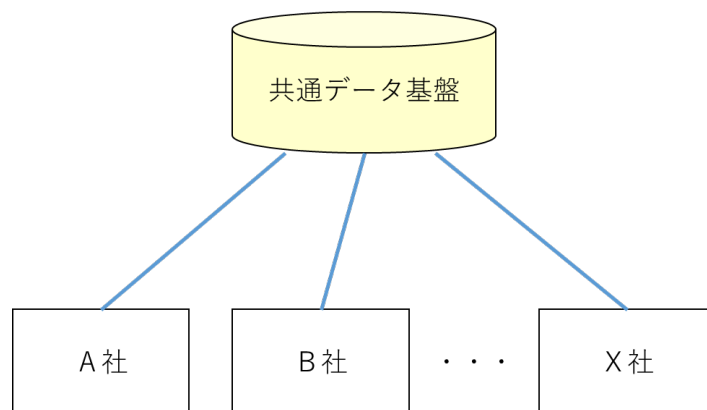
データの加工作業を支援（または自動化）するツールが出てきているので活用を検討する。

また、データを加工した場合の権利義務関係についても整理しておく必要がある。

⁶ 「教師あり学習」：データに正解データ（ラベル）を付加する作業をアノテーションといい、ラベル付きのデータで学習する方法を「教師あり学習」という。

7) 顧客データの共同収集への取り組み

クラウドサービス事業者は、自ら複数の顧客のデータを集めることが可能であるが、結果の精度を上げるためには、より多くのデータを集める必要がある。そのためには、複数のクラウドサービス事業者が共同で収集し、業界共通の「データ基盤」を構築する方法も有効である。

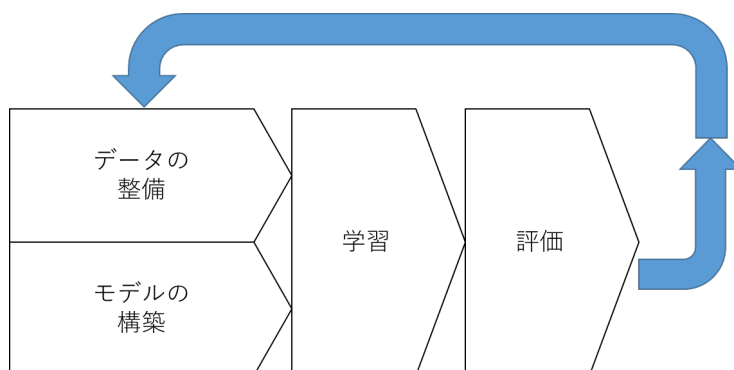


4. 2 モデルの構築から学習、評価まで

4. 2. 1 実施事項

AIの開発では、モデル（プログラム）を構築し、学習用データを使って学習を行う。学習済のモデルに対し、評価用データやテストデータを用いて、評価を行う。評価結果により、モデルを改善し、追加学習を行う。これらの作業は繰り返し行われる。一連の作業の結果、実運用に使われる最終的な「学習済みモデル」が生成される。

なお、P o Cフェーズにおける作業も同様であるが、規模や環境が実運用のものと異なる。



4. 2. 2 留意すべき事項

AIの開発においては、通常のシステム開発とは方法も課題も異なる面がある。

1) 開発手法の転換

AIシステムの開発は、従来のシステム開発と手法を変える必要がある。

① システム開発方法の転換

AIの特性上、演繹的なシステム開発方法から帰納的なシステム開発へ（プログラミングから機械学習へ）のパラダイムの転換が必要である。

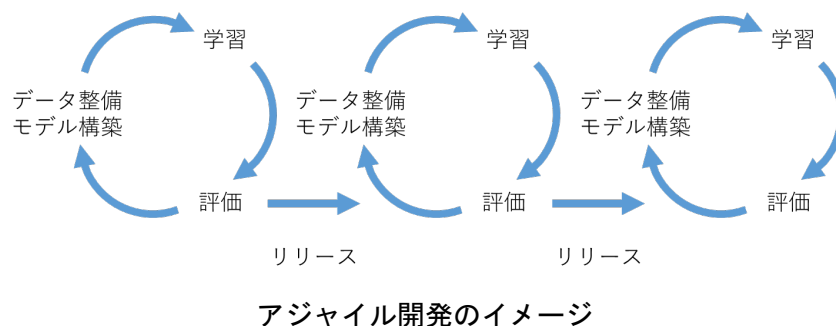


国立研究開発法人科学技術振興機構「AI 応用システムの安全性・信頼性を確保する新世代ソフトウェア工学の確立（2018 年 12 月）」

なお、開発の効率化の観点から、演繹モデルと帰納モデルを組み合わせた開発手法も研究されている。

② アジャイル型開発の導入

AIの開発には、ウォーターフォール型ではなく、アジャイル型の開発が向いていると言われている。



なお、A I 機能を含むシステム全体の開発においては、ウォーターフォール型の開発が採用されることもあるが、その場合はウォーターフォール型とアジャイル型の組み合わせとなる。

2) オープンソースの利用について

A I の開発においては、オープンソース⁷を活用する割合がかなり大きいため、オープンソースのリスクがクローズアップされる。オープンソースの利用は開発効率を大幅に高める一方、留意すべき点もある。

オープンソースを使用するにあたっては、いくつか留意する点がある。

① 責任分担の問題

オープンソースに問題（バグ等）があり、バグ等によってトラブルが生じたときは、オープンソースを利用したクラウドサービス事業者が責任をとることになる。また、オープンソース提供主体は、無償で提供しているので、通常は損害賠償を求めることはできない。実際には、オープンソース提供主体で使用許諾の条件を提示しているので、使用する前に十分確認する必要がある。

② 保守性の問題

⁷ よく使われている A I のオープンソースには、以下のものがある。

- ・ フレームワーク : TensorFlow (機械学習に用いるためのソフトウェアライブラリ)、Pytorch (Python (プログラミング言語) のオープンソース機械学習ライブラリ)、Keras (Python のオープンソースニューラルネットワークライブラリ) など。
- ・ ライブラリ : Numpy (Python の科学計算用ライブラリ)、Matplotlib (Python 及び NumPy のためのグラフ描画ライブラリ)、Pandas (Python のデータ解析ライブラリ) など。

また、フレームワークやライブラリ以外にも、A I のコミュニティでは多くの A I 機能がソースレベルで共有されている。

- ・ Github : Github の Git とは、プログラムのソースコードをバージョン管理するツール。Github は、Git をオンライン上で管理するウェブサービス。
- ・ Colab : ブラウザから Python を実行できるサービス。

オープンソースは、突然の修正やバージョンアップが行われることがある。互換性のない変更もある。オープンソースが変更になった場合のシステムの変更手順を決めておく必要がある⁸。

3) 品質向上／品質管理

AIは出力結果が一定ではないという性質を有するため、品質管理や品質保証が難しいという問題があるが、品質向上については最大限の努力を行う必要がある。

① 精度評価指標の充実

通常、AIの精度の評価には、正解率、適合率、再現率、F値（適合率と再現率の調和平均）などの指標が使われているが、学習に使用していないデータに対してどの程度正確な答えが出せるかという「汎化性能」や、データにノイズが含まれていても正確な答えが出せるかという「頑健性」についても評価を行う必要がある⁹。

4) 公平性の担保

AIの場合、結果の正確性に加えて、結果の公平性という課題がある。AIの出力結果が差別的な内容になって社会問題¹⁰になることもある¹¹。

AIの出力結果に対する公平性を担保するためには、不公平を生じる原因¹²を理解し、適切な対応を行う必要がある。

公平性を担保するには、データの偏りをなくすだけでなく、公平性の基準を明確に

⁸ 解決策として、オープンソースは使わない企業もいる。また、オープンソースに対するリスクを負う企業と契約するという形態をとっている企業もいる。

⁹ AIの品質管理については、国立研究開発法人産業総合研究所の「機械学習品質マネジメントガイドライン第2版（2021年7月5日）」及びAIプロダクト品質保証コンソーシアム（QA4AIコンソーシアム）の「AIプロダクト品質保証ガイドライン 2021.09版」（2021年9月15日）が参考になる。

¹⁰ AIの公平性に関しては、諸外国でいくつかの事件（犯罪者の再犯率の推測における人種差別等。）が起きている。

¹¹ 「人間中心のAI社会原則」では、「AIの設計思想の下において、人々がその人種、性別、国籍、年齢、政治的信念、宗教等の多様なバックグラウンドを理由に不当な差別をされることなく、全ての人々が公平に扱われなければならない。」としている。また、「AI利活用ガイドライン」の「AI利活用原則」にも「公平性の原則（AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービスの判断にバイアスが含まれる可能性があることに留意し、また、AIシステム又はAIサービスの判断によって個人及び集団が不当に差別されないよう配慮する。）」がある。

¹² 不公平の種類と原因：不公平には、「機会の不平等」と「結果の不平等」とがある。また、AIが不公平を生み出す原因には、①データバイアス、②サンプリング（標本抽出）バイアス及び③帰納バイアスの3種類がある。

し、それに合うアルゴリズムを選択する必要もある。また、設計意図、製造方法、テスト条件を明確にしておくことも重要である。

なお、公平性を担保するための技術の研究も行われているので、最新の技術動向に留意する必要がある。

5) セキュリティの確保

A I をターゲットにした犯罪¹³も発生しているため、A I の開発においては特別なセキュリティ対策が必要になる。特にA I クラウドサービスは、オープンネットワークを利用するため留意が必要である。

現在、A I をターゲットとした攻撃に対しては、様々な対策が研究されてきているが、最新技術動向を常に確認する必要がある。



A I をだまして道路標識を誤認識させる例

Kevin Eykholt 他, Robust Physical-World Attacks on Deep Learning Visual Classification (<https://arxiv.org/pdf/1707.08945.pdf>) より

¹³ 現在、A I をターゲットとした攻撃には、以下のようなものがあげられる。

- ・ Adversarial Examples (敵対的サンプル) : 入力データでA I をだまし、間違ったアウトプットを出力させる。
- ・ A I を乗っ取る攻撃
- ・ 学習データ汚染 : 学習データを改ざんし、A I に誤分類を起こさせる。
- ・ プライバシー侵害 : 学習データを盗み、学習データから個人を特定する。
- ・ 推論ロジックの改ざん

6) 説明可能性の向上

A I（特にディープラーニングを用いたA I）は、ブラックボックス的な構造のため、出力結果の根拠や理由を示すのが難しいという課題がある。

一方、「A I 利活用ガイドライン」の「A I 利活用原則」には、「透明性の原則」¹⁴及び「アカウントビリティ（説明責任）の原則」¹⁵があり、国際的な議論¹⁶にもなっている。A I を使ったシステムで説明責任を果たすには、説明可能性の向上が鍵となる。

A I の説明可能性は、アプリケーションの種類によっても異なる。使われたパラメータとその値を言われても、顧客や利用者には理解できない。（説明になっていない。）

① マーキングによる説明

A I のうち画像認識などは、根拠を示しやすい分野である。画像の中で判断根拠となった部分に印をつける等の工夫が行われている。

② 説明しやすい手法の選択

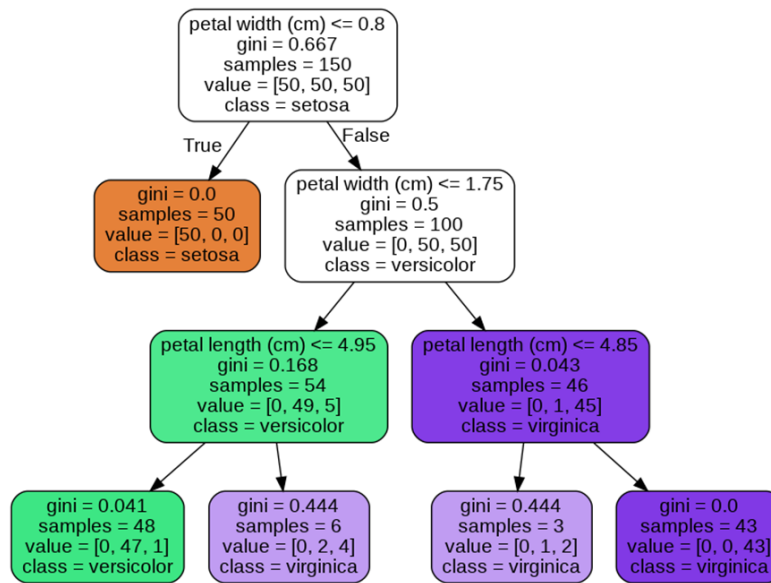
使われるA I の手法によっても説明可能性が異なるため、説明可能性の高い手法を選択するという方法もある。例えば決定木という手法は、根拠を示しやすいと言われている。ただ、決定木でも複雑になりすぎると説明しにくくなる。

¹⁴ 「透明性の原則」：A I サービスプロバイダ及びビジネス利用者は、A I システムまたはA I サービスの入出力の検証可能性及び判断結果の説明可能性に留意する。

¹⁵ 「アカウントビリティ（説明責任）の原則」：A I サービスプロバイダ及びビジネス利用者は、消費者的利用者及び間接利用者を含むステークホルダに対しアカウントビリティを果たすよう努める。

¹⁶ A I に関する説明責任については、国際的な議論となっている。

- ・ Google、IBM、Microsoftなどが立ち上げたPartnership on AIは、「A I のFAT(Fairness, Accountability, Transparency)」について議論している。
<https://www.partnershiponai.org/reflections-from-apm/>
- ・ E U のG D P R第22条に「A I の透明性」という条文が含まれている。
- ・ 「機械学習における公平性・説明責任・透明性ワークショップ（FAT/ML）」が2014年から毎年開催されている。



判断の根拠がわかりやすいアルゴリズム（決定木）の例

③ 説明可能なA I の技術動向に留意する

説明責任は、説明する手段がなければ果たせないため、各国で「説明可能なA I」（X A I : Explainable AI）¹⁷の技術開発^{18,19,20}が行われている。特に高い説明可能性が要求されるA I クラウドサービスを開発する場合には、「説明可能なA I」の技術動向を調査し、採用を検討する必要がある。例えば、審査系のA I クラウドサービスで、審査結果に対する不満が出ることが予想される場合には、高い説明可能性が要求される。

¹⁷ 現在、説明可能A I の実現には、現在4つのアプローチがある。

- ① Black Box Model Explanation（解釈可能なモデルの抽出）
- ② Model Output Explanation（出力の説明を生成）
- ③ Model Inspection（ブラックボックスの中身の検査）
- ④ Transparent Box Design（透明性のある学習器の設計）

¹⁸ 米DARPAは、X A I 研究開発投資プログラムで、機械学習に解釈性を持たせつつ精度を上げる研究を行っている。

¹⁹ 日本の大手企業（NEC、富士通等）も早くから取り組んでいる。（NECの「ホワイトボックス型A I」等）

²⁰ 個別企業だけでは研究開発をカバーしきれないため、各国が協調して国家プロジェクトとしても取り組んでいる。日本ではNEDOの「次世代人工知能・ロボット中核技術開発」（「説明できるA I」）で7件が採択されている（2019～2023年度）。

5. 個別調整フェーズ

5. 1 実施事項

クラウドサービスは複数の顧客に同じサービスを提供するのが基本であるが、顧客別に調整が必要な場合もある。特にA Iクラウドサービスの場合は、入力するデータの違いにより、顧客ごとに精度が異なることがあるため、顧客別の精度向上策が必要な場合がある。

5. 2 留意すべき事項

1) トライアルの実施

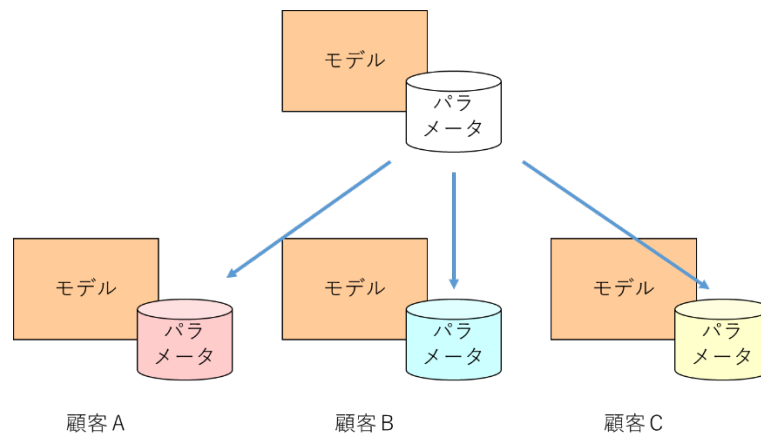
顧客によっては、A Iクラウドサービス開発時に学習に使用したデータの範囲外のデータを入力することもあり、十分な精度が確保できないことがある。また、処理に時間がかかることもある。正式な契約（利用申込）に先立ち、顧客に使ってもらい、十分な精度や性能が確保できるかなどを確認するトライアルの機会を設けることが望ましい。

トライアルを実施するための環境については、事業者側で用意する必要がある場合もある。

2) 顧客別追加学習による精度の向上

A Iの出力結果は、学習したデータや運用時に入力するデータにも依存するため、顧客ごとに精度に差が出ることもあり、何らかの顧客別対応が必要な場合もある。

しかし、クラウドサービスの本質は汎用的なサービスであり、顧客ごとにプログラムを修正しては、クラウドサービスのシェアリングエコノミーとしてのメリットが失われてしまう。これを解決するためには、プログラムは修正せずにデータの学習のみによる精度向上ができることが望ましい。顧客ごとに自分のデータを使って追加学習を行う「顧客別追加学習」という方法が考えられる。



顧客別追加学習のイメージ

モデルは変えずに、追加学習によりパラメータ（重み、バイアス）の値が変わる。

運用開始にあたってこの顧客別追加学習によって A I の出力結果の精度を向上させることが行われることが望ましい。例えば、音声認識の A I の分野では、運用を開始するにあたって操作者の声を学習させるということが行われている。

アプリケーションによっては、追加学習にかなり時間がかかることもあるが、モデルの一部のみに学習させるというテクニックもある。

顧客別追加学習を行うにあたっては、利用者が行う場合と事業者が行う場合とが考えられる。利用者が行う場合は、利用者に学習用の環境やツールを提供する必要がある。

なお、アプリケーションの種類によっては、顧客別の追加学習を行わないほうがいい場合もある。例えば、人事の採用を支援する A I で顧客別追加学習をすると、すでに在籍している人と同じような人ばかり推薦して問題となることがある。

6. 運用・保守フェーズ

6. 1 実施事項

運用・保守フェーズにおいては、データの入力など、実際の運用を行うのは顧客である。クラウドサービス事業者は、運用状況のモニタリングを行い、必要に応じてA Iの改善（モデルの修正や追加学習）を行う。A Iの改善には、顧客ごとの対応と、全顧客に対する対応とがある。

6. 2 留意すべき事項

1) モニタリングの重要性

A Iクラウドサービスでは、運用フェーズにおいて、入力データの変化や追加学習により出力結果が変化することもあり、事業者側でも運用状況をモニタリングする必要がある。

① 精度の変化の確認

A Iの場合、運用段階でも出力結果の精度が変化するため、常にモニタリングを行う必要がある。特に追加学習を可としているA Iサービスにおいては精度の変化が起きる可能性が高い。

② 原因分析

モニタリングの結果、出力結果の精度に問題が発見された際には、以下のような点に着目して分析を行う。

- ・ データの傾向の変化
- ・ レアケース
- ・ 過学習
- ・ モデルの変数（特徴量）不足

③ 利用者からの意見（苦情）・問い合わせの分析

A Iの問題点や課題は、明確に障害という形で現れるだけでなく、出力結果の偏り

による不都合などもある。障害だけでなく利用者からの苦情や問い合わせなどのインシデントについても分析する必要がある。

2) 運用期間中のA Iの改善

① 個別モデル修正より個別追加学習

A Iクラウドサービスの場合も通常のクラウドサービスと同様に、システムの継続的な改善やバージョンアップが行われる。この対応はすべての顧客に対する対応である。個別の顧客ごとにモデルを修正するというのは避けたい。

通常のクラウドサービスと異なるのは、日々を入力データを使った追加学習の有無である。追加学習を行うことにより、A Iの出力結果の精度を向上させられる可能性があるのは、運用開始時の場合と同様である。顧客別の対応にも有効である。

	全顧客	個別
モデルの修正	○	× (避けたい)
追加学習	○	○

② 利用者側の作業への影響も勘案すること

運用期間中にモデルの修正や追加学習を行う場合は、利用者側の作業の変更や稼働の増加などが発生する可能性があることに留意する必要がある。例えば、発熱チェック画像検索の精度が高まると、疑義者が増えてそれに対応する要員も増やさなければならなくなることがある。

7. その他

7. 1 外部委託と外部連携

AIクラウドサービスの開発において外部委託を行う場合は、プログラムの製造だけでなく、膨大な稼働がかかるデータ収集や加工についても外部委託することが考えられる。

1) プログラムの外部委託の留意点

AIのプログラムの開発の外部委託については、通常の外部委託における留意事項に加えて、品質管理（精度評価指標の充実）に関する合意が必要になる。

納品物の受入れ基準のうち、出力結果の正確性の評価について明確にしておく必要がある。学習に使用するデータの範囲及びテストに使用するデータの範囲とそれらに基づくテスト結果の精度目標について合意し、委託仕様書等に盛り込む必要がある。

2) データ収集・加工の外部委託

データ収集及び加工を含めて外部委託する場合には、以下のような点に留意する。

① 収集するデータの範囲・種類・量・分布等を明確にする。

データの偏りがAIの出力結果の精度を左右するため、収集するデータの範囲については詳細に検討し、委託仕様書等で明確に指示する。

② データの納品フォーマットを明確にする。

アノテーション（ラベル付け）を含むデータの加工を委託する場合には、最終的に納入されるべきデータのフォーマットを委託仕様書等で明確にする。

③ データ収集における法令の遵守等の義務を守らせる。

委託契約において、著作権法、個人情報保護法、不正競争防止法等の関連する法律の遵守を徹底させる。

3) 外部委託先との責任分担

外部委託先との責任分担については、通常の外部委託と同様に、あらかじめ契約書によって明確にしておくことが望ましい。

4) APIによるAI連携

一部のAI機能について、他社が作成したAIをAPI経由で利用する場合については、通常のAPI利用と同様、ライセンスの条件を十分に確認することが必要である。また、トラブル発生時の原因の切り分け方法や責任分担についても明らかにしておく必要がある。

7. 2 環境問題への配慮

環境への配慮やSDGsの達成のために、AIの開発においてもCO2の排出削減等の配慮が求められる。特にAI機能の開発においては、機械学習等に莫大なコンピュータ能力を必要とするため、消費電力の削減に留意する必要がある。具体的には、 unnecessary learning をできるかぎり排除し、少ないデータで学習できるアルゴリズムを採用する等の工夫が求められる。

7. 3 AIのエコシステムについて

社会全体のエコシステムという観点からは、1企業ですべてのAIを開発するのではなく、企業間で役割分担をすべきという議論もある。AIの人材不足の問題もあり、企業間の連携によるAI開発も検討すべきである。クラウドサービスはオープンなネットワークを通じてサービスを提供することが基本であり、ネットワーク上での連携がしやすい環境にある。

AIクラウドサービスを提供している企業の中には、すでに他社にAPIを提供している企業も多い。

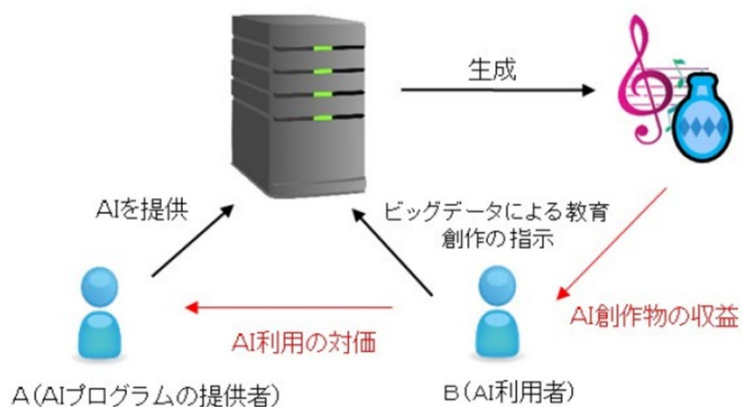
7. 4 AIの創作物に関する著作権の問題

AI技術の新しい分野として、画像の生成や作曲などができる「生成系のAI」も

開発されてきており、A I の創作物に著作権はあるか、という議論²¹²²が行われている。

コンテンツ・クリエイターによるAI利用

創作の道具としてのコンピューターの
延長線上で、AIを創作活動に利用



「次世代知財システム検討委員会報告書」²¹より

また、1から創作物を生成する場合だけでなく、元の画像データを加工して新しい画像データを作り出すような場合についても、今後検討していく必要がある。

²¹ 「次世代知財システム検討委員会報告書～デジタル・ネットワーク化に対応する次世代知財システム構築に向けて～」(平成28年4月 知的財産戦略本部検証・評価・企画委員会)でもまだ検討中が多い。

²² 現在の議論は消費者向けのコンテンツ(音楽、映像など)のビジネスが中心であるが、作曲家などの専門家を支援するシステムも含まれる。

第Ⅱ編 AIクラウドサービスの提供

第Ⅱ編では、提供者が利用者に、AIクラウドサービスを提供するに当たって推奨される取組を、提供前と提供期間中の2つの観点からガイドする。

1. 提供前の段階で推奨される取組

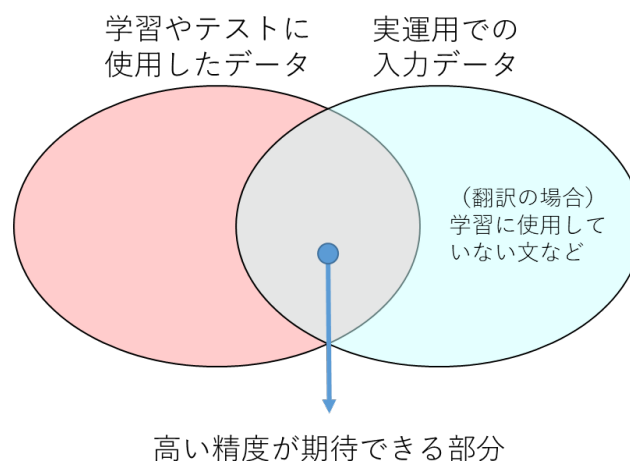
① AI機能の性質及び構成について説明する。

提供者はサービスの提供に当たり、利用者に対して、AI機能の性質及び構成について、AIの性質を踏まえつつ、十分かつ適切な説明を行う必要がある。または、利用規約やSLA、サービス仕様書等に明記する必要がある。

■ AI機能の性質について説明する例

AIの出力結果は、実際に入力されるデータの範囲等に依存することを説明する必要がある。

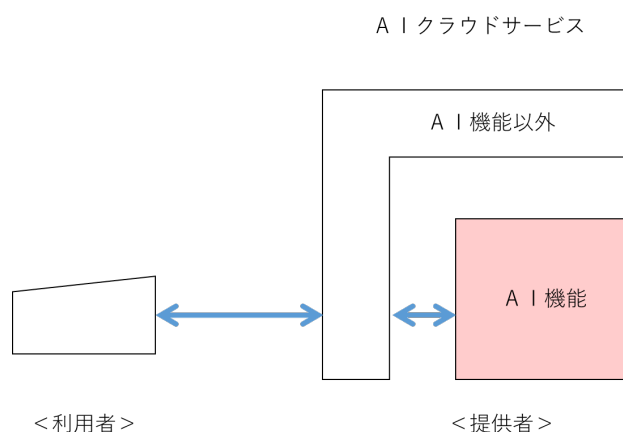
例えば、利用者に対してAI機能の精度を具体的な数字等で示す場合、その精度の根拠となった条件（学習やテストに使用したデータの範囲など）も併せて説明する必要があり、実際の利用状況において、これらの条件が満たされない場合（想定外のデータの入力がある等）は、事前に示した精度が発揮されない可能性があることも、利用者に説明する必要がある。



なお、A Iに入力されるデータを利用者でコントロールできる性質のサービスについては、上記のようなA Iの性質を踏まえ、提供者は、利用者がA I機能の便益を十分に享受できるよう、利用上の留意点等について適切な情報提供を行なうことが期待される。

■ A I 機能の構成について説明する例

A I 機能においては、A Iの出力結果を基に、要件定義された所定のプログラムが、利用者に価値を提供（通知や提案等）するものがある。その場合は、インシデント等が発生した場合に、その原因がA I機能を構成するどの部分にあるのかによって、対応の方法や期間が異なる可能性があるため、提供者は利用者に対して、A I機能の構成を事前に説明する必要がある。



A I 機能の構成例

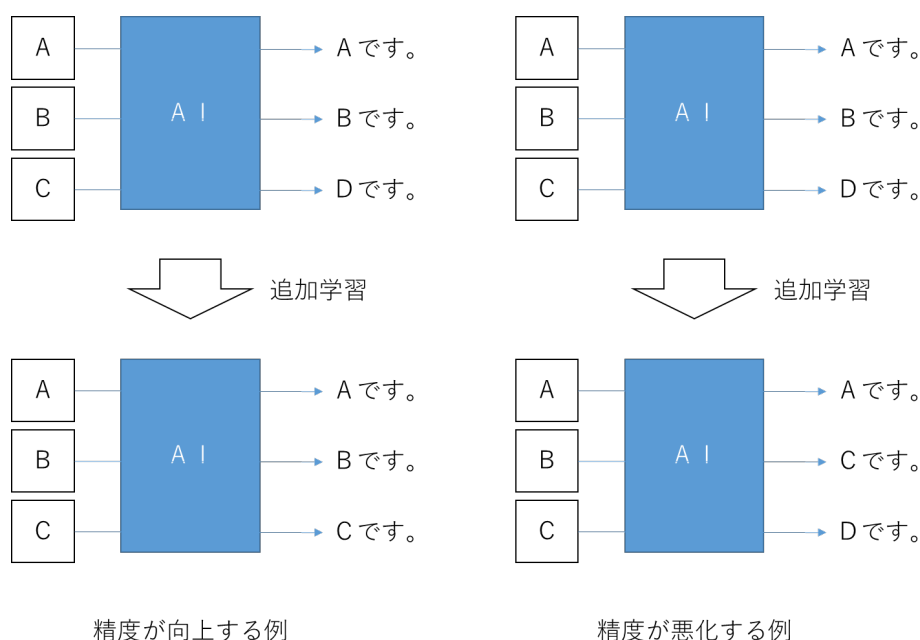
上記において、要件定義された所定のプログラムの不備等に起因する損害については、提供者の責任と判断される可能性がある（実際の責任の所在は当事者間で締結した契約等の内容によって判断されるものと考えられる。）。一方、A Iの出力結果によってもたらされた損害については、提供者の責任と判断することが困難である場合もあることから、利用者においても、そのリスクの受容可否を検討する必要がある。よって、提供者は、利用者におけるサービスの導入検討を支援する観点からも、A I機能の構成等について事前に説明することが求められる。

なお、A Iの出力結果によってもたらされた損害等への責任に関する考え方等については、③にて解説する。

② サービスの利用中にA Iの出力結果の精度が変化する可能性があることを説明する。

提供者は、サービスの提供中において、A Iに対する追加学習を行う場合等により、A Iの出力結果の精度がサービスの提供中に変化する可能性がある場合は、利用者に対して、その旨を説明する必要がある。または、利用規約やS L A、サービス仕様書等に明記する必要がある。

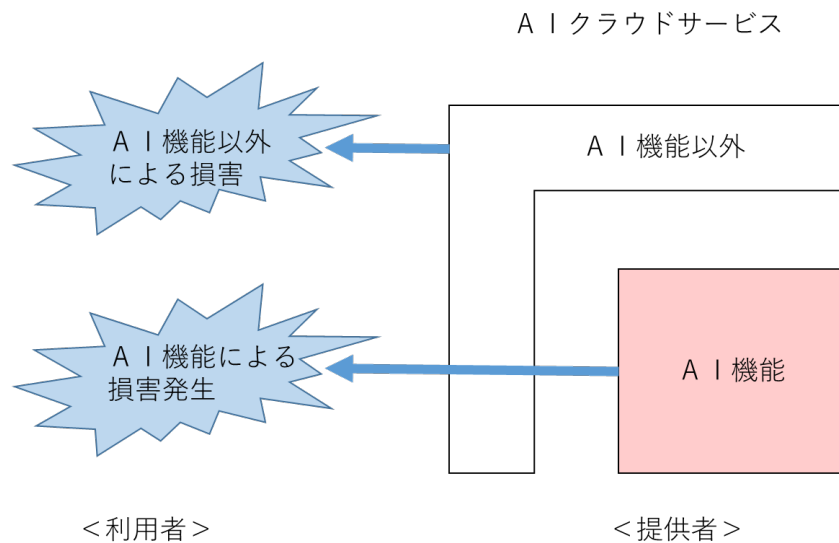
例えば、サービス提供中に、A Iの出力結果の精度の向上を目的とした追加学習を行う際に、結果的に、精度が悪化する可能性がある等のリスクがある場合は、そのリスクについても利用者に事前に説明することが望ましい。



③ A I機能に起因する損害等への責任について説明する。

提供者はサービスの提供に当たり、利用者に対し、A I機能に起因する損害等について、提供者が負う責任と利用者が負う責任（提供者における免責事項）を、それぞれ明確に説明する必要がある。または、利用規約やS L A、サービス仕様書等に明記する必要がある。その際、発生した損害が、A Iの出力結果に起因するか否かで、責任の所在が変わる場合は、サービスの構成上、どの機能に起因するものなのかを、利用者の知識等を配慮した上で、適切かつ具体的に示す必要がある²³。

²³ 一般的にビジネス利用者は、A Iの性質を理解した上で、A Iを利用するか否か、利用する場合はどのように利用するかを、自ら判断することが期待される。また、製造物責任法などの特別法で定められた場合を除く

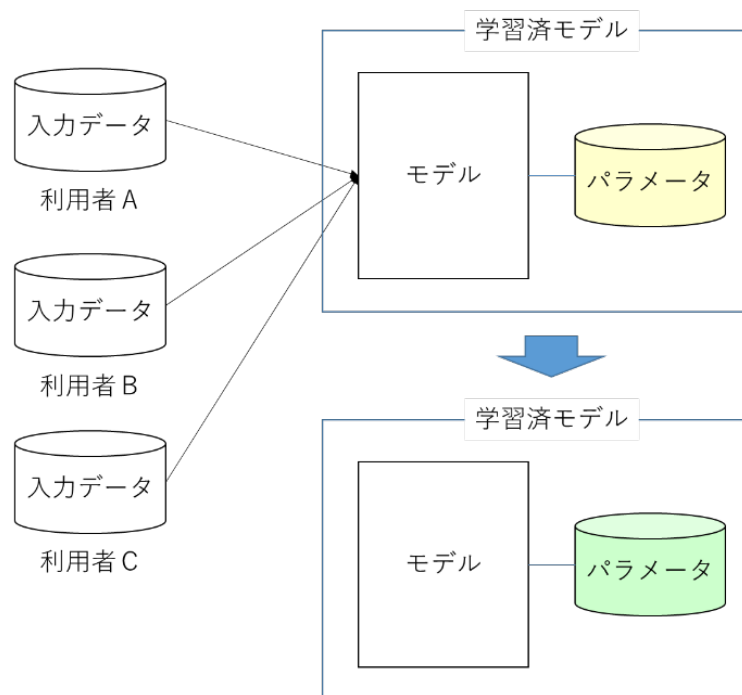


A I 機能による損害か否かにより責任の所在が異なる場合

④ 取得した利用者に係るデータの利用について説明する。

提供者はサービスの提供によって取得した利用者に係るデータを、A I 機能の精度の維持及び改善のための追加学習等に利用する場合は、その旨を利用規約やS L A、サービス仕様書等に明記し、かつ、提供時に利用者へ説明することが望ましい。また、その場合にA I の新たな学習済モデルが生成される場合は、それらに係る権利義務等についても、利用規約やS L A、サービス仕様書等に明記し、かつ、提供時に利用者へ説明することが望ましい。

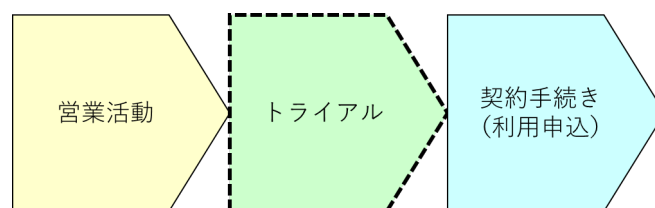
き、A I の性質を踏まえると、A I 機能の予見できない出力結果に起因する損害等について、提供者に責任を問うことは難しいと考えられる。



利用者の入力データを追加学習することにより精度が向上する

⑤ A I 機能の精度を利用者が確認するための機会を提供する。

提供者はサービスの提供に当たり、サービスの正式な契約の前に、利用者に対して、A I 機能の精度が、利用者の要求水準に達しているかを確認するためにトライアル（試用）の機会を提供することも有効である。なお、当該機会の提供に際して、利用者に費用等の負担が発生する場合は、その費用とサービスの正式な契約に基づく利用料等とを、契約等で明確に区別することが望ましい。



契約手続きの前に利用者によるトライアル（試用）を行うことも有効

2. 提供期間中において推奨される取組

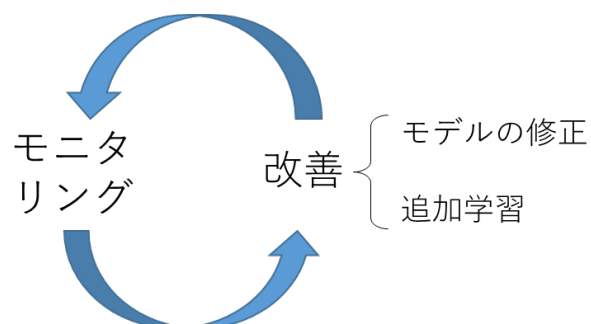
① A I 機能のうち、A I の出力結果の精度の維持及び改善に取り組む。

A I の出力結果の精度は、サービスの提供中における、A I への追加学習の有無や実際に A I に入力されるデータの質によって変化する可能性があることに留意し、提供者は、サービスの提供中も継続して、その A I の出力結果の精度の維持及び改善に努める必要がある。

具体的には、A I の出力結果について、用いる A I の技術水準、サービスに対する利用者の期待等を踏まえ、事前に一定の水準を定めた上で、その精度をモニタリングすることが期待される。

その際、その精度が事前に定めた水準を下回った場合には、モデルの修正や追加学習等により、その精度の改善に努めることが期待される。

なお、これらの取組の実施の有無については、利用者がサービスを選択するにあたり重要な情報となる可能性が高いことから、利用規約や S L A、サービス仕様書等に明記する必要がある。

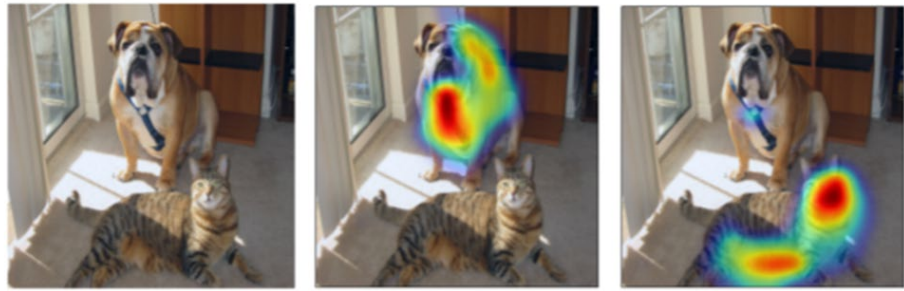


運用段階では精度のモニタリングと改善が必要

② A I の出力結果の根拠について説明できるよう努める。

サービスの提供中に、業務の性質上必要な場合やインシデントが発生した場合等において、利用者に A I の出力結果の根拠等の説明が必要な場合には、提供者は、サービスに用いられる A I の技術水準等を踏まえ、適切な説明を行なうことが期待される。特に公平な取り扱いが強く求められる審査系の A I の場合は、結果の根拠についての説明が必要となる。

第 I 編に記載した通り、説明可能性を高める技術の開発が進められてきている。例えば、画像認識の A I の場合、物体を認識した根拠となった部分を示す方法がある。



(a) Original Image (b) Cat Counterfactual exp (c) Dog Counterfactual exp
Figure 6: Negative Explanations with Grad-CAM

Ramprasaath R. Selvaraju 他、Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization (<https://arxiv.org/pdf/1610.02391.pdf>) より
この図で、(b)は犬がいると判断した根拠を、(c)は猫がいると判断した根拠を、色を変えて示している。