資料2

開発者・利用者から見たAI利活用における議論

#2019/02/05　AIガバナンス検討会

小池　誠

自己紹介

- 小池　誠
  - 農家
  - AI技術を用いたキュウリ選別機を開発
    - AIを用いた画像認識でキュウリの等級を判断
  - １年間ほどAIを使ったシステムを利用

開発者・利用者の視点から見たAI利活用

前提（1）

- **AIとは**
  - ここでは機械学習（主にディープ・ラーニング）を用いた手法について考える
  - AGIはあくまで概念として扱う（実装方法が不明なので）
- **従来システムとAIシステムの違い**

従来システム	AIシステム
❑ 仕様から演繹的に構築	❑ データから帰納的に構築
❑ モジュール化し評価・保証をシンプルにできる	❑ モジュール化不可能で評価・保証が超複雑
❑ 入力は限定されている	❑ 入力は膨大（現実世界，ビッグデータ）
❑ 意図しない入力に対する挙動を特定可能	❑ 意図しない入力に対する挙動は未知（ブラックボックス）

AI技術は，入力が膨大で複雑な問題・網羅的な評価ができない問題（主に現実世界の問題）を解くために利活用されることが予想（期待）される

前提（2）

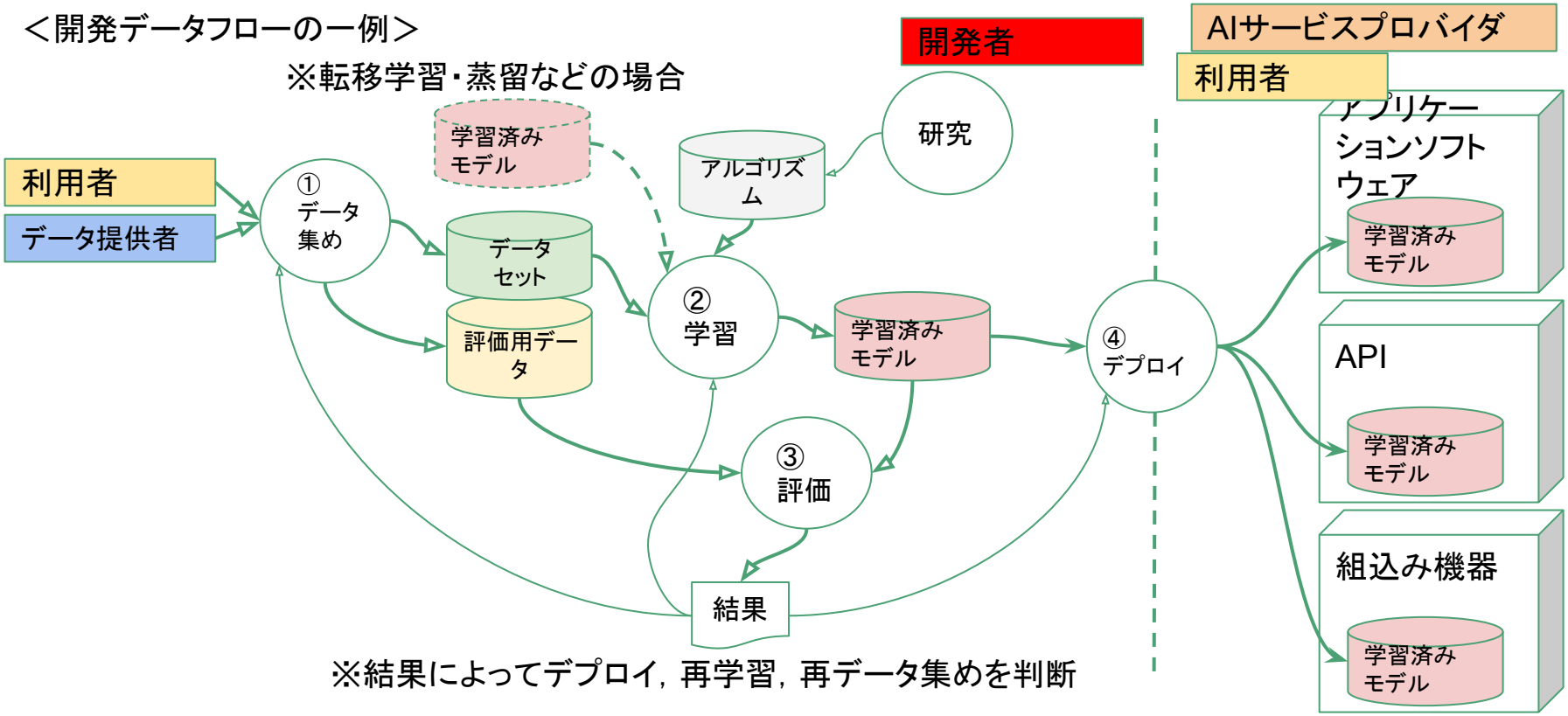
- AIが社会に受け入れられ利活用されるためには
  - 安全：品質，制御可能，予測可能など
  - **安心**：安全性，公平性，プライバシー，リテラシーなど

AIシステム/サービスの**利用者は，データ提供者を兼ねる**場合が多いと考えられる．
利用に関してだけでなく，自身の生み出したデータが**どのように利用されているのか，それはコントロール可能であるか**という観点を含めた“安心”が必要だと考える．

※例えば，「愛猫家が集めた猫画像から，猫の人気判断AIが作られ，人気＝値段の低い猫を自動的に殺処分するために活用されていた」なんてことが起こるかもしれない．AI開発者は，データとデータから生成されたモデルの利用先をデータ提供者に知らせる必要があるのではないか

AIシステムとは

- AI利活用原則案での「AIシステム」と主体の役割



- ❑ AIシステム，AIソフトがどの部分を指すのか？
- ❑ 利用者の指す範囲が広すぎる？（エンドユーザと事業利用者を同列で扱うべきか）

目次

- 人間とAIの役割分担
- AIのアップデートと実現可能性
- データの扱いとトレーサビリティ
- アカウンタビリティとリテラシー

人間とAIの役割分担

- キュウリの等級判断における役割の変化

＜人間＞
なし

＜AI＞
等級判断
の全責任

＜人間＞
判断基準の調整
最終チェック
判断の責任
＜AI＞
判断のサポート

作業場の明るさや季節によって，判断精度が不安定.
たまに，明らかな大間違いをするな･･･不安.
（**AIモデルの品質保証の不確実性に起因する問題**）
結論：AIは使えんな･･･

AIのサポートで作業効率が4割アップした！最終的には人間がチェックするから **安心･安全**.

<バイヤー>今までどおり人間が判断してるなら**安心**（**責任の所在**）

人間とAIの役割分担における論点

［適正利用の原則］

利用者は，人間とAIシステムとの間及び利用者間における適切な役割分担のもと，適正な範囲及び方法でAIシステム又はAIサービスを利用するように務める.

利用者視点

論点：適切な役割分担，範囲，方法とは？

- 基本的には，最終利用者はAIの機能/性能，使われる状況などを考慮し，利用者が柔軟に設定すべき
- しかしながら，最終利用者が適切な判断を下すためにはAIの素性や異常時の振る舞いなどの情報が必要であり，**開発者/サービスプロバイダはできる限りリスク情報を提供・公開し適正利用に協力する義務**があると考える

AIのアップデートと実現可能性（1）

- AIのアップデート（モデルの更新）が必要な場合
  - 学習で用いたデータの分布と運用で用いるデータの分布に差異が出てきた場合
  - 最終利用者の環境の変化/嗜好の変化により，過去のデータがノイズになる場合（主にレコメンドシステムなど）
  - 意図しない振る舞いが確認された場合
    - 開発時には発見できなかった不適切な振る舞いなど

など

- ❏ AIシステムは仕組み上，動かしてみないとわからない部分も多い，且つ，事前に網羅的な評価を行うことも難しいため，リリース後に意図しない動作が確認される可能性が高いと思われる
- ❏ アップデートはAIシステム／AIサービスを運用する上で，継続的に取り組む必要がある課題である

AIのアップデートと実現可能性（2）

- 迅速なアップデートと実現可能性
  - 再学習には質の良い最新のデータが大量に必要になるが用意できるか？レアケースの場合は？
  - アップデートによってデグレが発生しないか？
    - 事前に評価不可能で，やってみないとわからない
  - 現状のAI技術では難しい問題の場合は？
    - 例：絵画やアートと猥褻画像との判断（意味理解の問題）
  - AI組込み機器などを確実にアップデートする仕組みは？
    - ユーザが実施せず，リスクを抱えたまま運用されるケース
    - 旧AIと新AIが混在することで発生するリスクなど

など

- ❏ AI開発者・プロバイダは，アップデートに務めるべきであるが，AIの運用される環境・状況によっては困難な場合もあることを想定しておくべき
- ❏ （今後もAI技術は進歩すると思うが）現状のAI技術でできること・できないことは，正しく把握しておくべきではないか

AIのアップデートに関する論点

[適正利用の原則]

利用者は，人間とAIシステムとの間及び利用者間における適切な役割分担のもと，適正な範囲及び方法でAIシステム又はAIサービスを利用するように務める.

利用者視点

論点：現状のAIが対応しきれない問題を正確に把握しておく必要があるのではないか？現AIのできること，できないことの把握.

- 開発者・プロバイダは，AIシステムで発見された，又は，利用者から報告があったリスク情報を直ちに正しく通知すべきではないか
- 利用者は，開発者・プロバイダからのリスク情報を迅速に確認するよう務めるべきではないか

AIのアップデートに関する論点

[安全の原則]
利用者は，AIシステム又はAIサービスの利活用により，アクチュエータ等を通じて，利用者等及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する.

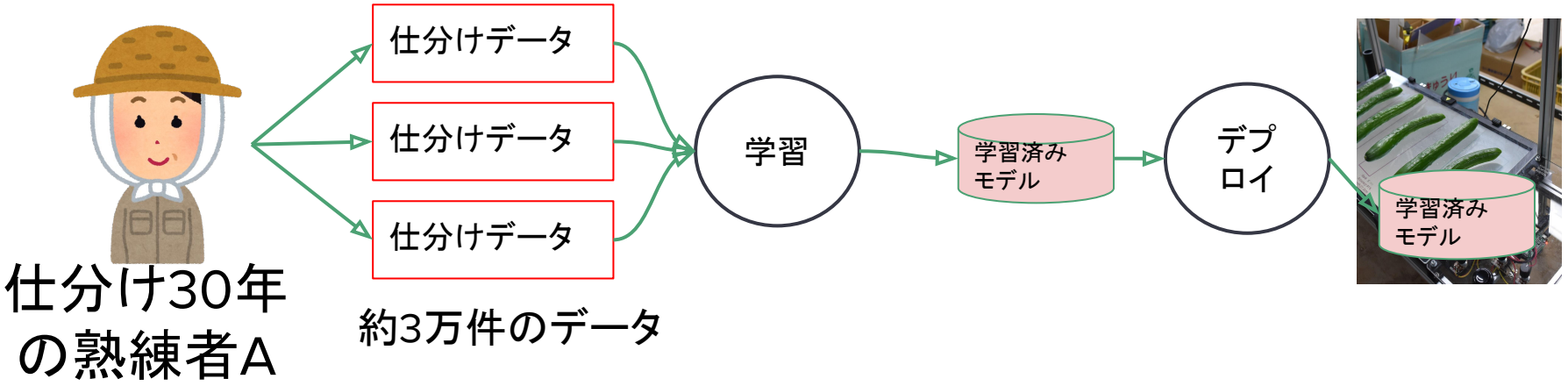
開発者視点

論点：質の良いデータが集まらない，又は，現状のAI技術では困難などで，AIソフトをアップデートできない場合もあることを考慮しておくべきではないか？

- 開発者は安全のために，AIソフトの点検・修理・アップデートという対策だけでなく，AIソフトの外側にフェールセーフシステムを置くなどして，システム全体で安全性を確保するような対策についても考慮すべきではないか.

データの取扱とトレーサビリティ

- キュウリの等級判断で扱ったデータ



- ❏ 仕分けデータは，単なる等級ごとのキュウリ画像ですが，農家30年の熟練者の目利き**ノウハウが詰まったデータ**です．
- ❏ また，このノウハウは熟練者Aさんだけが保有するもので，**熟練者Aさんを特定可能**なデータでもあります．

ノウハウが詰まったデータの取扱

- ノウハウが詰まったデータとは
  - キュウリの仕分け写真
  - イラスト
  - デザインフォント・書
  - ハウスの温度データ
  - ペンの加速度データ

著作物
（著作権で保護）

生データ
（著作権で保護されない）

AI開発で使用して良いデータ

ハウス管理，イラストのノウハウを抽出可能

など

- ノウハウが詰まったデータには，人間がそのノウハウを駆使して作成した著作物やノウハウを駆使して行った活動から生まれたデータ（データからノウハウが見て取れるもの）が存在する
- AI開発では，著作権者の許可なくこれらのデータを自由に使用することができまる※但し，著作権者の利益を不当に害さない範囲で（著作権法30条の4）

ノウハウが詰まったデータの論点

[尊厳・自律の原則]

利用者は，AIシステム又はAIサービスの利活用において，人間の尊厳と個人の自律を尊重する.

利用者・開発者の視点

論点：人間が長い時間を費やして獲得した技術・ノウハウを，著作物やデータから簡単に短時間でAIに掠め取られてしまう社会は『人間中心の健全なAI社会』と言えるのか？

- 現状のAIは，人間が生み出したデータから学習を行うものが多い．学習の元となるデータの源泉（人間の創作・学習意欲）を枯らしてしまわないか
- （人間も他人から技術・ノウハウを学ぶが，それと同じか）

パーソナルデータの取扱

- 個人の特徴を表すAIの取扱

＜熟練者Aの場合＞

「私と同じようにキュウリの等級判断するAI．他人に使われるのはなんだか恥ずかしい．」

特定個人の性質を再現するAI（学習済みモデル）を，本人の許可なく公開してもよいか．
※現状は利益を不当に害していないため可能？

学習済み
モデル

- ビッグデータから学習したAIは，どうしても平均的な振る舞いとなってしまい面白くない．今後は，よりパーソナルデータを用いる事例が増えるのではないか（エンタメでの活用など）
- パーソナルデータの取扱は，データ収集や学習だけでなく，作成した学習済みモデルやそのモデルの運用を含めて，プライバシー侵害がないことをチェックする必要があると考える（フェイクニュースなども）

パーソナルデータの取扱についての論点

[プライバシーの原則]

利用者及びデータ提供者は，AIシステム又はAIサービスの利活用において，他者又は自己のプライバシーが侵害されないよう配慮する．

利用者・開発者の視点

論点：自分の特徴・振る舞いをAIが再現し，自分がそれを不快に思った場合，プライバシーの侵害といえるか？

※例えば，自分の声を再現するボイスチェンジャーAIが作られたら？

データのトレーサビリティ

- AI時代において，データ自体がとても価値を持つようになる
- データを売り買いする業者の増加や市場の拡大なとが予想される
- そうなると，**データの産地・出所を偽装するような業者**も出てくるのではないか

データのトレーサビリティについての論点

[適正学習の原則]

利用者及びデータ提供者は，AIシステムの学習等に用いるデータの質に留意する.

利用者・開発者の視点

論点：学習等に利用するデータが，いつ，どこで，どういう目的で集められたデータなのかを明確に管理しておく必要はないか？（例えば，データセットの管理用にメタデータの共通フォーマットを準備する等）

- AI利活用の発展のためには，良質なデータの流通が必要
- 利用者，データ提供者は，データの質に留意すべきだが，ビッグデータなどの膨大なデータの質をチェックすることは現実的か．エコシステムが必要にならないだろうか．
- AItoAIな学習（例えば蒸留やAIが作るデータからなど）も考慮すべきでないか

アカウンタビリティとリテラシーの論点

[アカウンタビリティの原則]

AI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タビリティを果たすように務める．

利用者の視点

論点：AIを開発・提供する者は，AIを安全・安心に利用してもらうためのアカウンタビリティを果たすよう務めると共に，サービス利用者のAIリテラシーを高める努力をすべきではないか？

- 安心・安全なAI利活用を広げるためにも，まずは最終利用者のAIリテラシーを高めることが必要不可欠であり，AIを開発者・提供する者はAIリテラシーの教育の取り組みなどが期待される

まとめ

- AIサービスを開発・提供する者：
  - 利用者が適切な範囲・方法でAIを使えるように最大限のサポートが期待される
  - AIソフト単体ではなく，必要であればフェールセーフシステムと組み合わせるなどにより，安全性に配慮すべき
  - データは利用者の重要な資産であるという認識持ち，適切に扱う
- AIサービスの利用者：
  - AIリテラシーを高め，適切な範囲・用途でAIを利用する
  - どんなデータで学習が行われたかを把握し，AIの出力の解釈を適切に行う