AIのトラストとガバナンスを支える研究開発

AIやデジタル化によるイノベーションは，従来のトラストやガバナンスの概念を大きく変えようとしている。日立はこれまで，デジタル社会におけるトラストとガバナンスのあり方を検討してきており，世界経済フォーラム，経済産業省と共同でトラスト・ガバナンス・フレームワークを策定している。また，総務省AIネットワーク社会推進会議に参画し，AIの社会・経済に与える影響やリスクを評価するとともに，AIのガバナンスの検討を進めてきている。
本稿では，デジタル社会におけるウェルビーイングの実現に向けて，「トラスト」の構築に関わる研究の取り組み，および，AIガバナンスを支える技術の研究開発について紹介する。

内田 尚和 | Uchida Naokazu
鍛 忠司 | Kaji Tadashi
Nick Blake
間瀬 正啓 | Mase Masayoshi
大橋 洋輝 | Ohashi Hiroki
Dipanjan Ghosh
Chetan Gupta
直野 健 | Naono Ken
髙田 実佳 | Takata Mika

1. はじめに

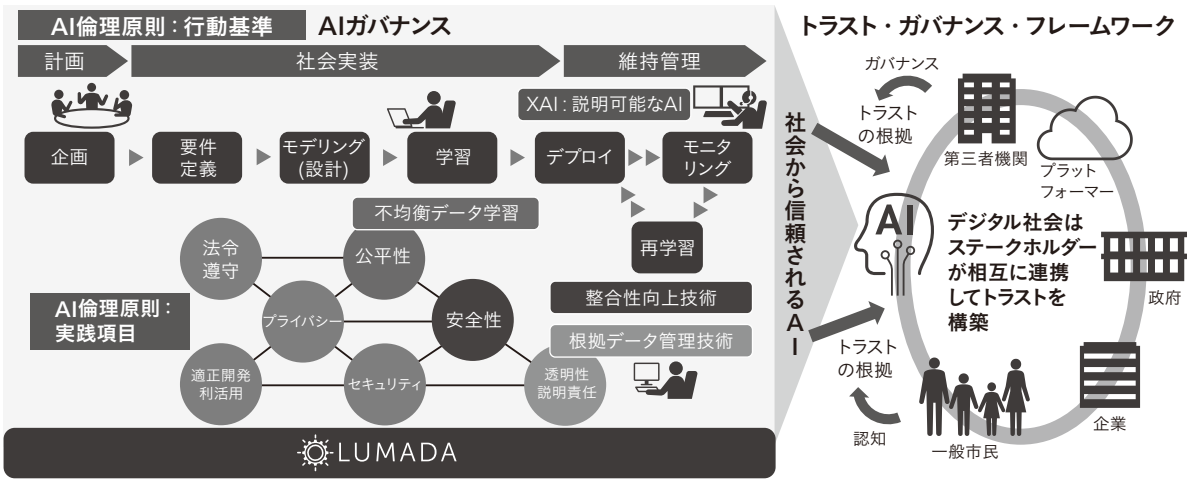
近年，AI（Artificial Intelligence：人工知能）技術は高度に発展し，鉄道，エネルギー，医療，金融など，幅広い分野でのAIの適用が進んでいる。一方で，AIやデジタル化によるイノベーションにより，これまで構築されてきたトラストやガバナンスの概念を変える必要が出てきている。例えば，デジタル社会では，さまざまなサービスが相互に接続して連携することで新たな価値を生み出すが，この構造は問題が発生した際の責任の所在を不明瞭にしている。また，AIによる判断の責任，アルゴリズムやデータの信頼性の担保など，従来の法令などでは対応が難しい課題が出てきている。

デジタル社会ではステークホルダーが相互に連携してトラストを構築する。技術やサービスに対し，信頼に値するという証拠（トラストの根拠）を示し，それがユーザーや市民に認知されることで，その技術やサービスが社会から信頼を得られる。この考え方に基づき，日立では「社会から信頼されるAI」の実現に向けた研究開発の取り組みを進めている。世界経済フォーラム，経済産業省と共同で，デジタル社会のトラストのあり方を提言する白書「Rebuilding Trust and Governance: Towards Data Free Flow with Trust（DFFT）」[2)]を発行し，デジタル社会に必要となるトラストとガバナンスの関係を「トラスト・ガバナンス・フレームワーク」として整理した。また，XAI（Explainable AI：説明可能なAI）をはじめ，さまざまなAIのガバナンスを支える技術を開発し，社会イノベーション事業のエンジンと位置付けるLumadaへの適用を進めている（**図1**参照）。

AIのガバナンスは，AIが社会から信頼されるために極めて重要である。AIのガバナンスにおいては，安全性，セキュリティ，プライバシーなどに加え，公平性，透明性・

図1｜AIガバナンスとトラスト・ガバナンス・フレームワーク

デジタル社会に必要となるトラストとガバナンスの関係を整理した「トラスト・ガバナンス・フレームワーク」と，「社会から信頼されるAI」を実現するAIガバナンスの全体像を示す。



注：略語説明
AI（Artificial Intelligence），XAI（Explainable AI）

説明責任などの観点に配慮する必要がある。

例えば，AIは判断の過程が見えないという問題がある。これに対してXAIは，判断に影響した因子の提示などによって判断の根拠を明らかにでき，AIを適用したシステムの透明性を高めることができる。AIの社会実装フェーズでは，専門家の解釈の下にAIモデルの改善に使用されるほか，AIの維持管理フェーズにおいても，その分析過程がユーザーに見える化されることで，ユーザーからの信頼の醸成につながることが期待される。しかしながら，XAIの分析結果と専門家の知見に不整合が生じることもある。したがって，専門家の知見をうまく取り込み，改善を図る技術の研究開発が必要である。

また，AIはデータからの学習を特徴とするため，学習データに含まれるさまざまなバイアスの影響を受ける。例えば，データ数が少ないカテゴリーはデータ数が多いカテゴリーに比べて分類精度が低く，カテゴリー間の公平性が損なわれる。そこで，バイアスを低減する技術によりこれを解消し，AIの判断が差別や偏見を助長しないようにする。カテゴリー間の分類精度の問題には，データ数だけでなくデータの性質なども影響しているため，バイアスの低減に向けてさまざまなアプローチで研究開発が進められている。

AIの運用時には，環境の変化に対応するために，AIのモデルを再学習する必要が出てくる。AIは学習データにより統計的に振る舞いが決まるため，再学習前と同じように振る舞うことを保証するのが難しい。再学習により，予測精度の著しい低下を招く可能性もある。そのため，再学習前後で，予測の整合性を評価する技術により精度低下を防止する。また，AIシステムの安全性を考慮すると，このように精度だけでなく振る舞いの整合性を維持することが求められる。

AIの開発運用過程においては，どのデータがどのように加工され利用されているのかが追跡できないと，データ利用に対して一般市民へ不安を与えてしまう。そこで，データの来歴を管理する技術により，データ利用の透明性を向上させることが求められる。このようなデータの来歴管理により，AIが学習するデータの信頼性を向上させ，学習結果の活用に至るまでのプロセスをより適正に管理することができる。

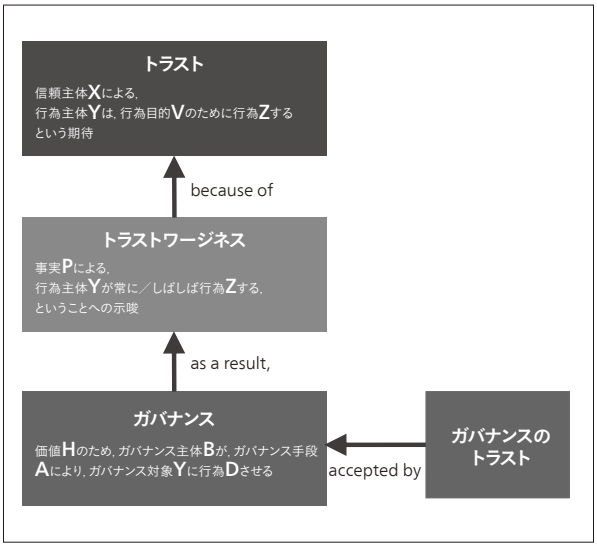
日立は前述した観点を踏まえ，研究開発を推進している。本稿では，2章でトラストの構築に関わる研究の取り組みを説明し，3章でAIガバナンスを支える技術の研究開発について概説する。

2. デジタル社会における トラストとガバナンス

本章では，デジタル社会におけるトラストを確保するためのガバナンスのフレームワークと，ステークホルダーが確かな信頼の下に協創し，データから新たな価値を生み出すことができる仕組みであるデジタルトラストについて説明する。

図2|**トラスト・ガバナンス・フレームワーク**

トラスト，トラストワージネス，ガバナンスを定式化するとともに，これらの関係性を利用して信頼を構造化し，信頼構築の施策検討に活用する。



2.1
トラスト・ガバナンス・フレームワーク

デジタル社会では，さまざまなサービスが相互に接続して連携することで新たな価値を生み出し，人々の幸福度を向上し続けることが期待されている。しかし，新たな価値が社会に提供されるためには社会からの「信頼」の獲得が欠かせない。本来，信頼は主観的なものであり，画一的な信頼獲得方法などは存在しない。対象や環境，内容に応じて信頼獲得に向けた施策を積み重ねていく必要がある。

「トラスト・ガバナンス・フレームワーク」は，社会からの信頼獲得を支援するため，どのようにステークホルダーに働きかければよいのかを整理するためのフレームワークである。特に，複数のステークホルダーが信頼獲得のための検討，議論をする際に共通の視座を与えることができると考えている。

本フレームワークでは，(1) 信頼獲得のためのガバナンスが適切に実施され，(2) ガバナンスの結果として信頼に値するという証拠（トラストワージネス）が蓄積され，(3) トラストワージネスがステークホルダーに信頼の根拠として認知されること，によって構築される信頼を「トラスト」であるとモデル化し，トラスト，トラストワージネス，ガバナンスの関係を整理している（**図2**参照）。

また，トラスト，トラストワージネス，ガバナンスそれぞれを定式化し，その式中の変数（信頼対象，ガバナンス手段など）に，対象や環境に応じた具体的な内容を代入することによってトラスト獲得のための施策検討を効率化する。例えば，「『市民』の『自動運転』は『安全』であるというトラストは，『政府』が『ルール』によって『自動運転車の製造メーカー』をガバナンスし，その結果，『交通事故件数データ』として『自動運転』は『安全』であることを示唆する証拠が蓄積され，『市民』が認知することでトラストが構築される」という具合である。ここで，『』の部分が変数であり，対象や環境，内容に応じて適切な値を検討することになる。

また，デジタル社会では変化のスピードが加速しており，従来型のガバナンスでは，いったん獲得した信頼を維持することは困難である。先ほどの例では，「政府がルールによって自動運転車の製造メーカーをガバナンスする」と述べた。しかし，デジタル社会では，技術革新の速度や，社会環境の変化などに応じて柔軟かつ迅速にルールを更新していく必要がある。そこで，本フレームワークでは，変化に追随するためのアジャイルな信頼構築プロセスについてもモデルを提供している。

2.2
デジタルトラストの構成

デジタル社会のステークホルダーが，確かな信頼の下に協創することを可能にするデジタルトラストには，「デジタルの信頼」と「デジタルによる信頼」という二つの側面がある。

「デジタルの信頼」とは，ステークホルダーがデジタルシステムを十分に受け入れ，採用するために必要な信頼を意味する。これは，サイバー犯罪者からの攻撃や妨害，不正利用に対してシステムとデータが安全であること，セキュアな生体認証によりユーザーの権利と権限が保護されること，また，すべてのステークホルダーが地域およびグローバルのルールと規制を遵守できることをデジタルシステムが保証することで実現される。デジタルシステムに関するルールや規制の例として，欧州委員会によって2018年に施行された一般データ保護規則（GDPR）[3]と，2021年4月に法案が提出され，早ければ2023年に施行される見込みのAI Act[4]が挙げられる。AI Actは，AIの目的と適用範囲を明確化し，人々の安全および基本的人権を効果的に保護する「社会から信頼されるAI」の採用を促すように設計されている。

「デジタルによる信頼」は，企業および行政機関の信頼を高めるためにデジタルシステムを用いることを指す。この信頼を実現するため，すべてのステークホルダーが，サステナビリティや倫理性，安全性に関して合意された基準を相互に遵守するという確信の下でデータの取り引きや共有ができるよう，従来の組織の垣根を越えた完全なトレーサビリティと透明性，電子公証を提供するデジタルシステムが求められる（**図3**参照）。

日立ヨーロッパ社は，欧州自動運転プロジェクト「HumanDrive」において，最先端のAIを用いた自律制御ソ

図3｜デジタルトラストの構成

デジタルトラストとは，デジタル社会のすべてのステークホルダーが，安全性やセキュリティ，プライバシー，倫理性に関して合意されたルールや基準を，透明性をもって相互に集団的に遵守するという確信の下で，協創することを可能にする信頼である。

注：略語説明
OT（Operational Technology）

フトウェアを開発した[5]。この技術の新機軸は，数テラバイトに及ぶ人間の運転データから学習に適したデータを抽出し，学習を行うインテリジェントなデータ管理ツールDRIVBAS（Driving Behaviour Analysis Software）の作成である。このツールを用いることにより，AIモデルのバイアスを取り除くことができ，道路環境を解釈して安全な経路を生成可能とするAIモデルを構築することができた。

3. AIガバナンスを支える技術

AI技術の進歩に伴い，社会インフラに関わるミッションクリティカルなアプリケーションにおいてもAIの利活用が求められている。一方で複雑なAIモデルの挙動は理解が難しく，AIの説明性の向上が課題となっている。個人や企業がAIを受け入れ，継続的に使用していくための公平なAIガバナンスを効果的に行うには，AIの判断の過程を明らかにし，これを専門家が解釈して必要に応じてAIの見直しを行うXAI技術，不均衡データを用いた際のAIモデルの出力バイアスを低減する不均衡データ学習技術，AI長期運用時の整合性向上技術，およびAI分析プロセスの透明性を確保する根拠データ管理技術などが有効である。本章では，これらの研究開発の状況について紹介する。

3.1 XAIと公平性に向けた取り組み

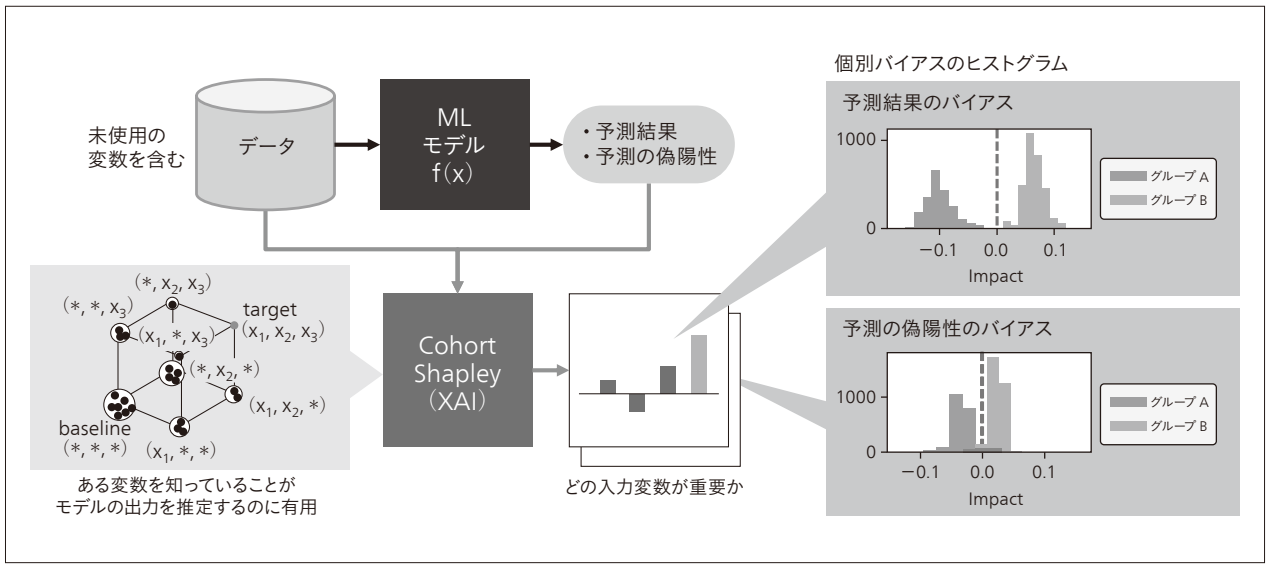
日立は，XAIを活用したモデルや学習データの多角的診断技術を開発し，AI導入・運用支援サービスを提供している。XAIを利用することで，例えば，住宅ローン審査において，ブラックボックス型のAIが予測した審査結果について，判定に影響した因子や事例を判断根拠として提示することができる。

ブラックボックス型のAIモデルの挙動を理解するための基本的なアプローチは，モデルの入力のどの変数が予測に寄与したかを定量化することである。統計学の分野において古くから研究・実践されており，何をもってその変数が重要とするかにはさまざまな考え方がある。多くの手法は「ある変数の値を機械的に変化させるとモデルの出力が変化する」ため重要である，という考え方である。一方で，公平性を考慮する際には，特定のグループを示す変数をAIモデルの入力に使用しなくても，データの相関を基に結果的にそのグループに不利益な結果をもたらしてしまうことが問題となる。この事象を分析するには「ある変数を知っていることがモデルの出力を推定するのに有用」なため重要である，という考え方が求められ，そのような手法として，Cohort Shapleyを開発した（**図4**参照）。

AIの公平性をめぐっては，モデル出力のグループ間の乖離のみならず，予測が正しくない場合の悪影響など，さまざまな指標のトレードオフを考慮した議論が求められる[6]。そこで，データに対する予測結果と正解データの双方から定義されるさまざまな公平性の定義について分析している。本事例では予測結果および予測の偽陽性のグループ間の偏りであるが，Cohort Shapleyを利用することで従来のグループ全体としてのバイアスに加えて，そのヒストグラムからどの個別事例がバイアスに寄与しているかまで分析を行うことができる[7]。今後，このような分析を利用したドメインエキスパートとの議論を通じて，信頼できるAIモデルの開発・運用を支援していく。

図4｜**Cohort Shapley**

予測結果および予測の偽陽性のクラス間の偏りについて，Cohort Shapleyを利用することで従来のグループ全体としてのバイアスに加えて，そのヒストグラムからどの個別事例がバイアスに寄与しているかまで分析を行うことができる。



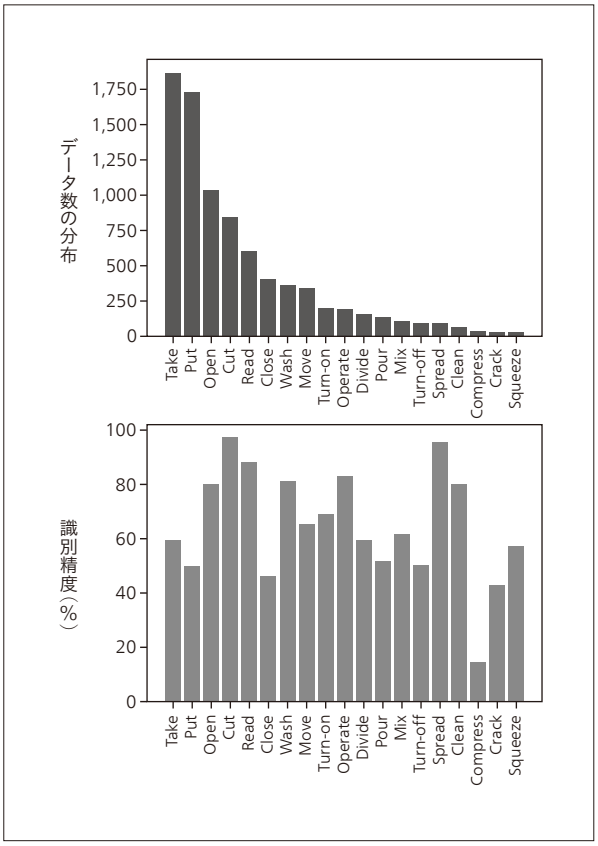
3.2
公平性のための不均衡データ学習技術

近年AIは，アルゴリズムや計算機ハードウェアの発展に加え，誰もがアクセス可能な大規模データセットが整備されてきたことで，大きな発展を遂げてきた。ところが，このような大規模データセットの中には，時に意図せぬバイアスが潜んでいることがあり，それを基にしたAIモデルが社会実装されると，思わぬ不公平を生んでしまうことがある[8]。典型的な例の一つは，データの分布に関するバイアスである。米国の情報工学者であるBuolamwiniらは，被撮影者の地理的な多様性を担保するように配慮して作成され，米国政府によって顔解析技術のベンチマーク用データセットとして使用されているデータセットでさえ，暗い肌の色の女性の画像の数が明るい肌の色の男性の画像の数に比べて極めて少なく，看過できないほどの偏りがデータセットに内在していることを明らかにした[9]。このようなバイアスが存在しているデータセットを基にAIモデルを学習すると，少数派のカテゴリーでAIモデルの性能が著しく低下してしまうことが古くから広く知られている[10]。

日立は，不均衡なデータセットを用いてモデルを学習する際にも，少数カテゴリーに属するデータに対するAIモデルの精度低下を低減し，より公平な出力を行える学習手法の研究開発を進めている。典型的な従来手法では，不均衡データセットを用いて学習を行う際には，各カテゴリーのデータ数に着目し，データ数の少ないカテゴリーを学習する際には重みを大きくし，反対にデータ数の多いカテゴリーを学習する際には重みを小さくすることで，不均衡を改善することを狙っていた。これは，データ数が少なければ少ないほどそのカテゴリーのデータを学習することは難しいという仮定を暗に置いているためであった。しかし，日立はこの仮定は常に正しいとは限らないことを実験的に発見した（**図5**参照）。

図5｜**行動認識データセットのデータ数分布と識別精度**

行動認識データセット（EGTEA dataset[11]）における行動カテゴリー別のデータ数の分布（上）と，当該データセットで行動識別モデルを学習させたときの識別精度（下）を示す。データ数に大きな不均衡が見られるが，データ数が少ないカテゴリーほど識別精度が低いというわけではないことが分かる。



この発見に基づき，モデル学習の過程でカテゴリーごとの学習の難しさを常時モニタリングし，その難しさに合わせた重みづけを行うことで，不均衡の是正を行う手法を提案した。当該手法は，論文発表当時，不均衡なデータセットからの学習において，従来手法を上回る性能を達成した[12]。

3.3
AI長期運用時における整合性向上技術

産業システムにAIが適用される一方で，オペレータは依然として重要な役割を果たしている。それゆえ，信頼性の低いAIシステムは，生産性や安全性の低下，経済的な影響などさまざまな問題を発生させる。例えば，製造工場において信頼性の低いロボットアームは，オペレータに危険を及ぼし，また，信頼性の低い自動運転車は，事故を引き起こす可能性がある。

日立アメリカ社は，これまでの経験から，AIシステムの信頼性向上には，(1) ヒューマン・イン・ザ・ループとルールベースによる保護機構の組み込み，(2) モデル学習方法の改善，の二つが有効であると見ている。XAIを利用した信頼性向上のアプローチもあるが，まだ初期検討の段階である。ヒューマン・イン・ザ・ループとルールベースによる保護機構は確実な方法だが，ドメインやユースケースに依存する。一方，モデル学習方法の改善は，よりロバストで，さまざまなドメインやユースケースに適用可能な基本的なアプローチである。本研究では，このアプローチに焦点を当てた。

信頼性の高い動作とは，整合性のある動作である。AIの観点では，特に予測が正しい場合について，同じ入力に対して同じ出力が得られることを意味する。言い換えれば，一貫して正しい予測ができることである。産業システムやミッションクリティカルなシステムの場合，AIモデルの出力の不整合は，安全上の問題とともに，ビジネスプロセスに悪影響を及ぼす可能性がある。そこで，運用開始後の継続的な再学習におけるAIモデルの挙動の変化，特に再学習によって予測結果が整合しなくなるケースについて検討を行った（**図6**参照）。

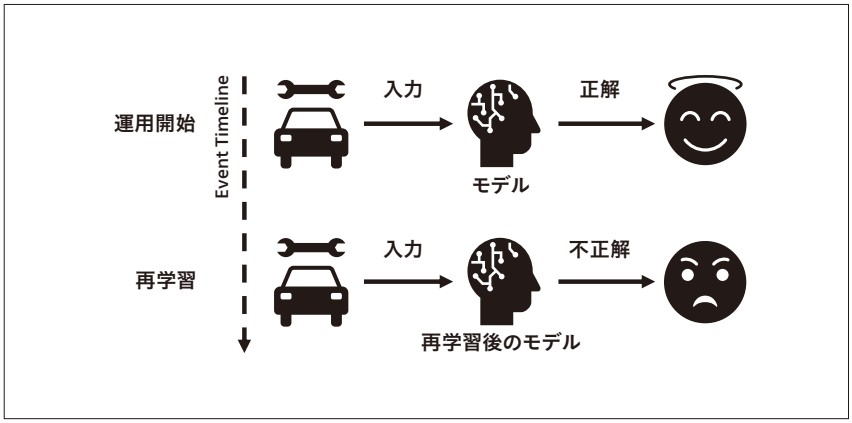
AIモデルの整合性を，再学習を行ってもAIモデルが一貫した予測を行う能力と定義する。これは，予測が正解の場合と不正解の場合の両方に該当するが，望ましくは一貫して正しい予測ができることで，これを正解整合性と定義する。この考えを基に，整合性を定量化する測定基準，および，Dynamic Snapshot Ensemble法[13]と呼ぶ効率的なモデル学習法を開発し，AIモデルの整合性が，理論上，どのような条件で改善されるかを実証した。正解率などの既存の測定基準に加え，提案手法を用いてAIシステムを事前に評価することでAIガバナンスに貢献していく。

3.4
AI分析プロセスの透明性を確保する根拠データ管理技術

AIが社会に浸透していく中，一般市民は，自分のデータがどのように利用されているのか，不安を感じることが多くなっている。例えば，AIが「10年後，あなたは要介護状態になる可能性が45%である」と出力したとする。しかし，45%という数字だけを知らされても，どういう意味を持つのか，どうすればその確率を下げることができるのか，という説明がなければ，単に不安を持つだけになってしまう。この不安を解消するには，まずAIがどのデータを基にどのように処理をして予測したのか，という透明性が確保されなければならない。そのうえでAIの処理プロセスに関して専門家がより深く解釈し，出力結果の意味について判断し，市民に説明する。そうすることで市民は，介護予防のより適切な行動を採りやすくなる。具体的には，要介護になる確率が45%であったとしても，その45%の主要因は何であり，その確率を下げる可能性のある要因は何であるかを，予測の根拠となったデータソースに遡って検索できるようにする。

図6｜再学習によって生じるAIモデルの予測の不整合

AIモデルは，運用開始後も性能の維持管理のために継続的な再学習が必要だが，再学習によってAIモデルの予測に不整合が生じる（予測結果が変わる）ことがある。



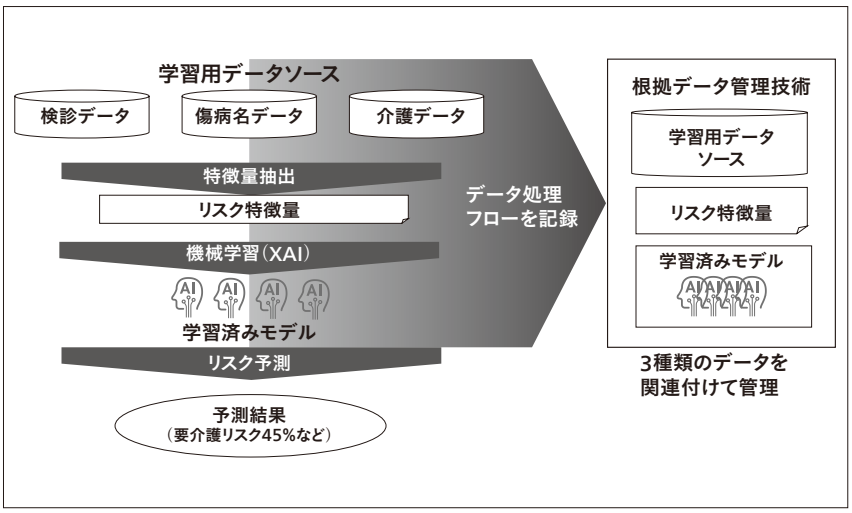


図7｜根拠データ管理技術の概要（医療・介護分野の例）

学習用データソース→リスク特徴量→学習済みモデルというデータ処理フローを記録し，関連付けて管理する技術である。これによって，例えば要介護リスク45％の根拠データとなるデータソースを即座に分析可能になる。

このようにAI予測の「根拠となるデータを管理する」ことで，AIに関する信頼性，安心感が確保されると考えられる。

日立は，この「根拠データ管理技術」を，米国医療機関Partners HealthCareとの協創活動の中で開発した。この協創活動での目的は，医療機関に入院した患者が退院後30日以内に再入院することを予防するため，ある一定以上のリスクの高い退院患者に限って，退院後も適切な連携を行う仕組みを構築することであった。このため，患者個人ごとに再入院確率を予測しつつ，その理由となるデータを提供することが求められた。これを実現するため，「学習用データ」から「リスク特徴量」を生成し，また，そこから「学習済みモデル」を生成する，というAI分析プロセスの一連の流れをすべて記録しておき，これら三つを相互にひも付けて管理するデータ管理方法を開発した（**図7**参照）。これを「根拠データ管理技術」として2018年12月に発表した[14]。

以降，さまざまなプロジェクトにおいて活用し，汎用的なデータ管理技術としてブラッシュアップ中であるが，特に近年，AIにおけるデータの品質管理が重要であるとの指摘が多数されるようになっており，この技術への期待が高まっている。例えば，現在，AI Actを検討する欧州委員会の「完全なデータを利用すること」という規制案に対し，日本のJEITAからの意見書[15]において，現実的な対案として，「データが信頼できることを示すデータ来歴，同意管理技術（根拠データ管理技術）などを適用すれば，データ処理から機械学習，そして学習結果の活用に至るまでのプロセスをより適正に管理することに寄与でき，データの出所の透明性が確保される」として，根拠データ管理技術が提案されている。今後，一般市民が安心してAIの利便性を享受できる世界の実現に向け，AI倫理に適うデータ管理技術として，根拠データ管理技術の普及に努めていきたい。

4. おわりに

本稿では，デジタル社会のトラストとガバナンスの枠組み，および，AIガバナンスを支える技術について紹介した。

日立は，社会イノベーション事業に必要となる「社会から信頼されるAI」を実現すべく，AI倫理原則とLumadaによるAIガバナンスを構築している。AI特有のセキュリティ対策技術やプライバシー保護技術，また，AI品質保証など，本稿では紹介しきれなかった技術の研究開発も多数行っており，これらの技術を組み合わせることで，安全・安心でレジリエントなデジタル社会，人々のウェルビーイングの実現に貢献していく。

謝辞

本稿で述べたトラスト・ガバナンス・フレームワークの開発においては，東京大学 西山圭太客員教授をはじめとする関係各位より多くのご支援を頂いた。また，本稿で述べたCohort Shapley手法はスタンフォード大学Art B. Owen教授，Benjamin B. Seiler氏との共同研究成果である。深く感謝の意を表する次第である。

参考文献など

1）総務省：AIネットワーク社会推進会議，
https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/index.html

2）WEF: Rebuilding Trust and Governance: Towards Data Free Flow with Trust（DFFT）（2021.3），
https://www.weforum.org/whitepapers/rebuilding-trust-and-governance-towards-data-free-flow-with-trust-dfft

3）個人情報保護委員会：GDPR（General Data Protection Regulation：一般データ保護規則），
https://www.ppc.go.jp/enforcement/infoprovision/laws/GDPR/

4）欧州委員会：Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE(ARTIFICIAL INTELLIGENCE ACT）AND AMENDING CERTAIN UNION LEGISLATIVE ACTS,
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

5）HumanDrive project achieves UK's longest and most complex autonomous journey,
https://www.hitachi.eu/en/press/humandrive-project

6）A. Chouldechova: Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big data, 5(2):153–163.（2017）

7）M. Mase, et al.: Cohort Shapley values for algorithmic fairness. Technical report, arXiv: 2105.07168.（2021）

8）A. Caliskan, et al: Semantics derived automatically from language corpora contain human-like biases., Science, 356（6334):183--186（2017）

9）J. Buolamwini, et al: Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification., Proceedings of Machine Learning Research 81:1--15, Conference on Fairness, Accountability, and Transparency（2018）

10）N. Japkowicz, et al: The class imbalance problem: A systematic study., Intelligent Data Analysis, 6（5）: 429–449（2002）

11）Y. Li: In the eye of beholder: Joint learning of gaze and actions in first person video, ECCV（2018）

12）S. Sinha, et al: Class-Wise Difficulty-Balanced Loss for Solving Class-Imbalance, Proceedings of the Asian Conference on Computer Vision（ACCV）（2020）

13）D. Ghosh: Wisdom of the ensemble: Improving consistency of deep learning models,
https://www.hitachi.com/rd/sc/aiblog/035/index.html

14）AIを用いた患者の再入院リスクの予測値とその根拠データを提示することができる情報ダッシュボードを開発,
https://www.hitachi.co.jp/rd/news/topics/2018/1212.html

15）JEITA：欧州委員会「LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE(ARTIFICIAL INTELLIGENCE ACT)AND AMENDING CERTAIN UNION LEGISTLATIVE ACTS(AI規則法案)」へのJEITA意見,
https://home.jeita.or.jp/cgi-bin/topics/detail.cgi?n=2996&ca=13&ca2=749

執筆者紹介

内田 尚和
日立製作所 研究開発グループ 先端AIイノベーションセンタ
メディア知能処理研究部 所属
現在，言語モデルおよび対話システムの研究開発に従事
人工知能学会会員

鍛 忠司
日立製作所 研究開発グループ
社会システムイノベーションセンタ 所属
現在，サイバーセキュリティやデジタルトラストの研究開発に従事
博士（情報科学）
IEEE会員

Nick Blake
Hitachi Europe Ltd., Big Data Laboratory,
European Research and Development Centre 所属
現在，スマートスペースとデジタルトラスト分野における社会イノベーション事業の統括業務に従事

間瀬 正啓
日立製作所 研究開発グループ 先端AIイノベーションセンタ
メディア知能処理研究部 所属
現在，機械学習の説明性・解釈性に関する研究開発に従事
博士（工学）
情報処理学会会員，IEEE Computer Society会員，ACM会員，American Statistical Association会員

大橋 洋輝
日立製作所 研究開発グループ 先端AIイノベーションセンタ
知能ビジョン研究部 所属
現在，機械学習・コンピュータビジョンを用いた製造・保守現場の技能伝承，ミス・事故低減，生産性向上支援の研究開発に従事
情報処理学会会員，人工知能学会会員

Dipanjan Ghosh
Hitachi America Ltd., Research and Development,
Industrial AI Laboratory 所属
現在，AIを活用した故障予兆診断ソリューションの研究開発に従事
PhD.（Mechanical Engineering）

Chetan Gupta
Hitachi America Ltd., Research and Development,
Industrial AI Laboratory 所属
現在，AIを活用した産業ソリューションの研究開発に従事
PhD.（Mathematics and Computer Science）

直野 健
日立製作所 研究開発グループ
デジタルプラットフォームイノベーションセンタ
データマネジメント研究部 所属
現在，医療介護分野のデータ処理技術の研究開発に従事
博士（工学）
日本応用数理学会理事，情報処理学会会員

髙田 実佳
日立製作所 研究開発グループ
デジタルプラットフォームイノベーションセンタ
データマネジメント研究部 所属
現在，データ管理および機械学習管理技術の研究開発に従事
人工知能学会会員，情報処理学会会員