

責任あるAI ガバナンス ガイドブック



責任あるAIとは？

「責任あるAI（レスポンシブルAI）」とは、顧客や社会に対してAIの公平性・透明性を担保する方法論です。この方法論に基づいてAIを設計・構築・展開することで、真に人間中心のAI活用の実現を目指すことが可能となります。

つまり、責任あるAIによって、企業はAIが持つリスクを正確に理解できるようになります。そして、AIが持つ潜在的なリスクへの対策をすることで、人は初めてAIを信頼できるようになり、ひいては自社のビジネスに応用・拡大利用することが可能となるのです。



責任あるAIの ガバナンスガイドブック

本ガイドブックは、お客様自身の組織や機関のニーズに合ったガバナンス戦略を作成できるよう、実践に役立つ知見を提供することを目的としています。確固たる戦略に加え、戦略に整合した適切な方策とシステムを備え、それらを適切に管理できれば、確かな信頼のもとにAIをビジネス全体で活用・運用できるようになります。

さあ、はじめましょう。

本ガイドブックが目指すこと



AI／機械学習を活用する組織において、倫理的で責任あるAIを運用するためのガバナンスフレームワーク（枠組み）がなぜ必要か、論拠を提示します。



AIにより意図しない結果が生じた際にどのように対処すべきか、そしてそれを実現するためにどのように自身の組織を改善すべきかを示したガバナンスフレームワークを提示します。



革新性（イノベーション）や効率性を損なうことなく、AIに対して責任ある形でどうガバナンスを効かせていくのか、具体的なガイダンスを提示します。

本ガイドブックで以下は カバーしていません

- ⊗ AIの基礎知識
- ⊗ AIによる偏見（バイアス）を回避するための基礎知識
- ⊗ 意思決定のための倫理的ガイダンス
- ⊗ 倫理的・非倫理的行動とは何かについての法的なアドバイス
- ⊗ 特定のAI技術の紹介や、当該AIの使用／非使用の推奨

これらの領域についてご相談がございましたら
info.tokyo@accenture.comまでお問い合わせください

関連情報

責任あるAI：AI倫理とガバナンス
<https://www.accenture.com/jp-ja/services/applied-intelligence/ai-ethics-governance>

AIに関する最新の知見や調査レポート
<https://www.accenture.com/jp-ja/insights/artificial-intelligence-summary-index>

AIサービス
<https://www.accenture.com/jp-ja/services/ai-artificial-intelligence-index>

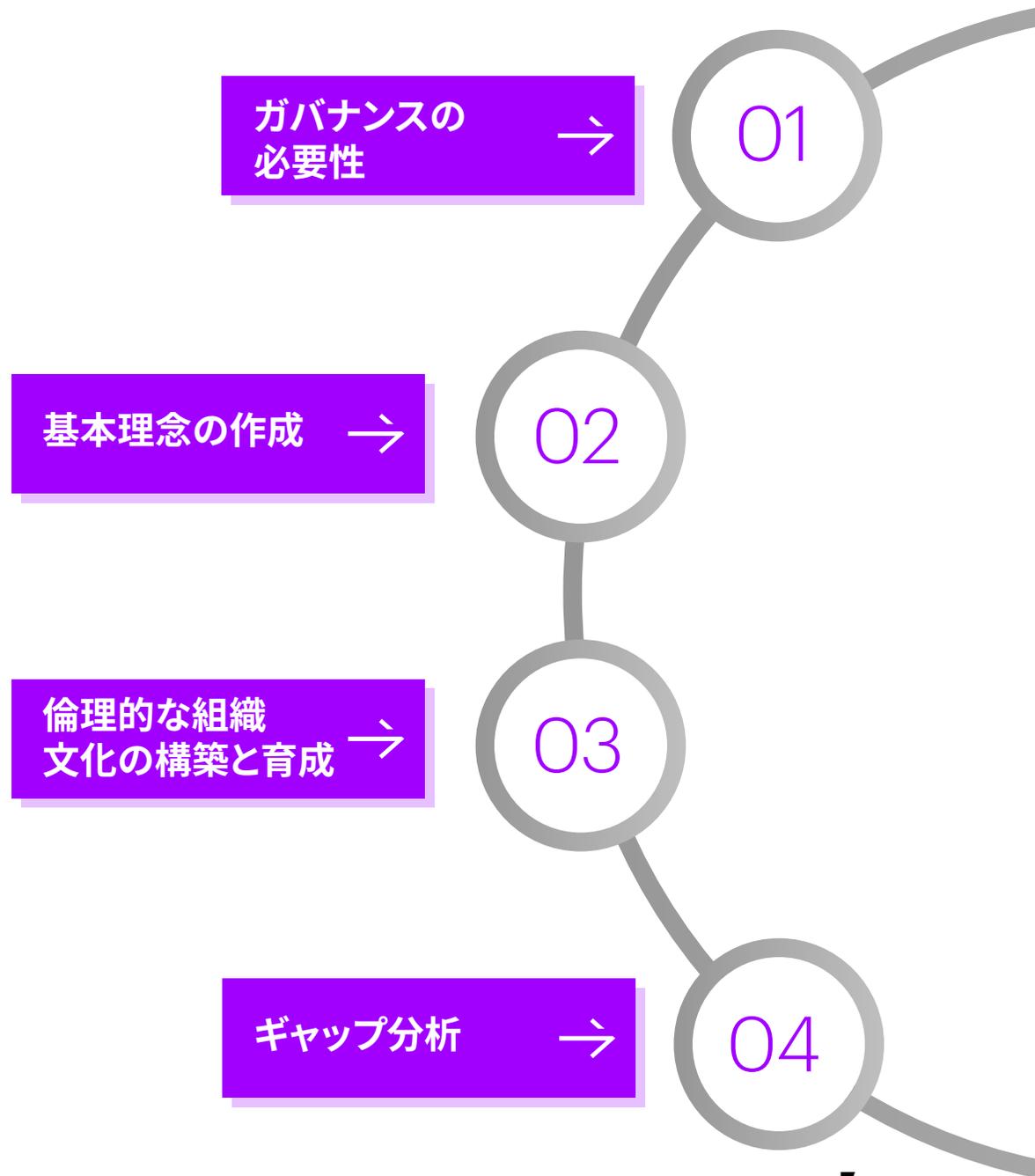
はじめに

倫理的で責任あるAIを設計・開発・保守するために必要な、優れたガバナンスとはどのようなものでしょうか？ AI／機械学習は実社会における複雑な問題を解決するために、多くのステークホルダーを巻き込む形で活用されています。そのため、そのガバナンス戦略は、関連する顧客や企業、社会的なコミュニティなどAIによって影響を受けるであろうありとあらゆるプレイヤーに対し、倫理的な解決策を導き出せるものでなければなりません。

実際のビジネス現場でAIを活用するにあたり、AIが導く結果が意図しないものになるリスクを考慮し、導入前に倫理的観点から致命的な問題がないかを明らかにする必要があります。私たちは皆、個人としてだけでなく、所属する組織や社会の一員として倫理的な責任を負っており、アルゴリズムや最新のテクノロジーの導入においてもその倫理観を適用しなければなりません。私たちに求められるのは、AI／機械学習を社会に実装する際に立ちはだかる困難に立ち向かいながら、関連するステークホルダーが持つ倫理観をAIに適切に反映させることで、人間中心のAI活用を推進することなのです。

AIツールの導入やAI利用を前提とした業務プロセスの改善・再定義を検討中、または導入中にある読者の皆様に向けて、本ガイドブックはまず倫理的なAIに関する3つのテーマを扱い、具体的なユースケースをご紹介した後、ギャップ分析で締めくくります。ギャップ分析は、組織内で責任あるAIのガバナンス戦略を実践するための行動指針として役に立つはずです。

倫理的で責任あるAIの運用のためには、AIが引き起こし得るバイアスについての基礎的な知識が不可欠となります。バイアスによる悪影響については次頁以降をご覧ください。



AIに悪影響を及ぼす様々なバイアス

AI導入に際して本来の意図に反して組み込まれてしまった、繰り返し発生し得るエラー、俗に言うアルゴリズムック・バイアスにより、AIモデルは悪影響を受けます。

アルゴリズムック・バイアスは、次のような悪影響をもたらします：

- 高いリスクを伴う判定／信頼度の低い判定結果の推奨
- 利益の逸失や機会損失
- 特定ユーザーを恣意的に優遇
- 恣意的なグルーピングと不公平な結果

アルゴリズムック・バイアスは、データ、モデル、運用オペレーションを通じてAIシステムに忍び寄ってきます。

アルゴリズムック・バイアス

社会的バイアス

学習モデルに使用されるデータに組み込まれた歴史的・文化的なバイアス

サンプリングバイアス

本来の母集団を表すようなデータ抽出ができていないことによるインプットデータのバイアス

コンセプトの欠陥

ビジネス目的の間違った解釈、モデル構築単位の間違った設定

モデルの不透明度

モデル構築に利用する重要な特徴量の透明性や説明可能性の欠如

モデルバイアス

学習データに含まれるバイアスがモデル構築の過程で増幅されることにより起こるバイアス

検証バイアス

開発者や観察者が真実であると信じている仮説や結果をより重視すること

測定バイアス

モデルの学習や精度検証に使用される入力データ／結果の測定が不十分

オペレーションバイアス

モデル展開後の経年による精度の低下（現実世界の変化、フィードバックバイアス）

不明瞭な結果

モデルのアウトプットが適切でない、または、分かりやすく表現できていないことによりユーザーにとって理解が困難

バイアスを防ぐ仕組みが必要



バイアスによりAIが悪影響を及ぼした事例

消費者へのインパクト



社会的悪影響

歴史的・文化的な固定観念が増強される



経済損失

特定のグループで選択肢が狭まる
(例：クレジットへのアクセス制限)



機会損失

職業、住居、教育など特定の属性を持つグループごとにアクセスした結果が異なる



自由の喪失

監視、人権、言論の自由などに関する懸念

ビジネスへのインパクト



質の悪い顧客体験



不適切なカスタマーサービス



アルゴリズムへの不信感による 厳しい監視の目

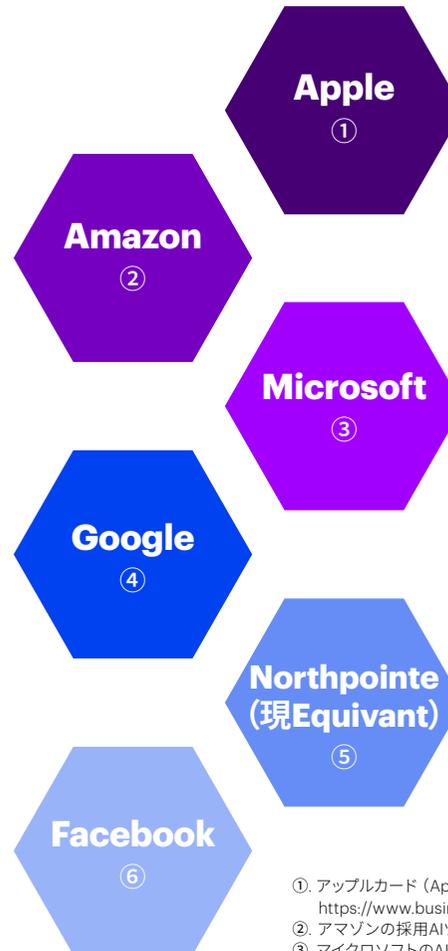


世論／ブランド評価の低下

アマゾン、ジェンダーバイアスの兆候を示したツールをスクラップに

Googleのエンジニアは、画像保存アプリ「Googleフォト」で有色人種に「ゴリラ」のタグ付けをしたことを謝罪

ケンブリッジ・アナリティカはFacebookを使って有権者の心理プロファイルを作成した結果、反発にあった



ゴールドマン・サックスはAppleの偏見疑惑を受け、Appleのクレジットカード限度額を見直し

MicrosoftはTwitterユーザーにより人種差別するよう教えられたチャットボット「Tay」をシャットダウン

未来の犯罪者を予測するソフトウェア「COMPAS」が有色人種に偏見を持った判断を下す

- ①. アップルカード (Apple Card) の性差別問題でゴールドマン・サックスが再審査を受け入れ <https://www.businessinsider.jp/post-202350>
- ②. アマゾンの採用AIツール、女性差別でシャットダウン <https://www.businessinsider.jp/post-177193>
- ③. マイクロソフトのAI、ヘイト発言を「学習」して停止 <https://wired.jp/2016/03/25/tay-tweet-microsoft/>
- ④. 黒人男女の画像、ゴリラと誤認識 グーグル謝罪 <https://www.nikkei.com/article/DGKKZO88907800U5A700C1CR0000/>
- ⑤. Machine Bias - There's software used across the country to predict future criminals. And it's biased against blacks. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- ⑥. FB個人情報流出 不正疑惑の英企業が全業務停止 破産申請へ 米大統領選と英国民投票に利用? <https://www.sankei.com/world/news/180503/wor1805030036-n1.html>

A woman with long dark hair is looking towards the camera in a dark, futuristic environment. She is wearing a light-colored top. The scene is illuminated by vertical neon light tubes in shades of blue and purple. To her left, a hand holds a camera with a monkey mask attached to its lens. The monkey mask has large, white, circular eyes and a dark body. The overall atmosphere is mysterious and high-tech.

01

ガバナンスの 必要性

グローバルのエグゼクティブ調査¹:



60%がデジタル時代におけるビジネスで競争優位性を保つにはAIの導入が必要だと回答しています。



一方、**45%**はAIによる「意図しない結果」について十分に理解できていないと回答しています。

「意図しない結果」とは、様々なコミュニティや利害関係者への潜在的な影響を考慮せずにAIを適用した結果、（そのAIの対象であるかどうかに関わらず）生じる副産物のことです。

単一のテクノロジーではなく多種多様で複雑なテクノロジーの集合体であるAIには、倫理的監督とリーダーシップが求められます。

一方、例えばAIを活用したバックエンドプロセスのような、特定の製品やデータの使用が人間に影響を与えない場合には、AIガバナンスは必要がないかもしれません。次頁以降では、AIが人間に対して「意図しない結果」を与える可能性が少なからずあるケースを想定して説明していきます。

自組織からみたAI倫理 ガバナンスの必要性

AIの倫理的ガバナンスは、自組織のビジネスの変革、所属する業界の変革、さらには社会の変革のために議論が欠かせない重要なアジェンダです。責任を持った形でAIを導入すれば、組織に関わるすべてのプレイヤーに対して広範囲にポジティブな影響を与えることができます。その影響範囲は、業務や働き方、人を支えるテクノロジー、組織のあり方、さらには組織が持つブランドの評判にまで至ります。

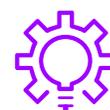
AIテクノロジーを活用したビジネスに対する倫理的ガバナンスの必要性は、あらゆる業界のAIのリーダーたちによって認識されています。アクセンチュアの調査により右記が明らかになっています²。



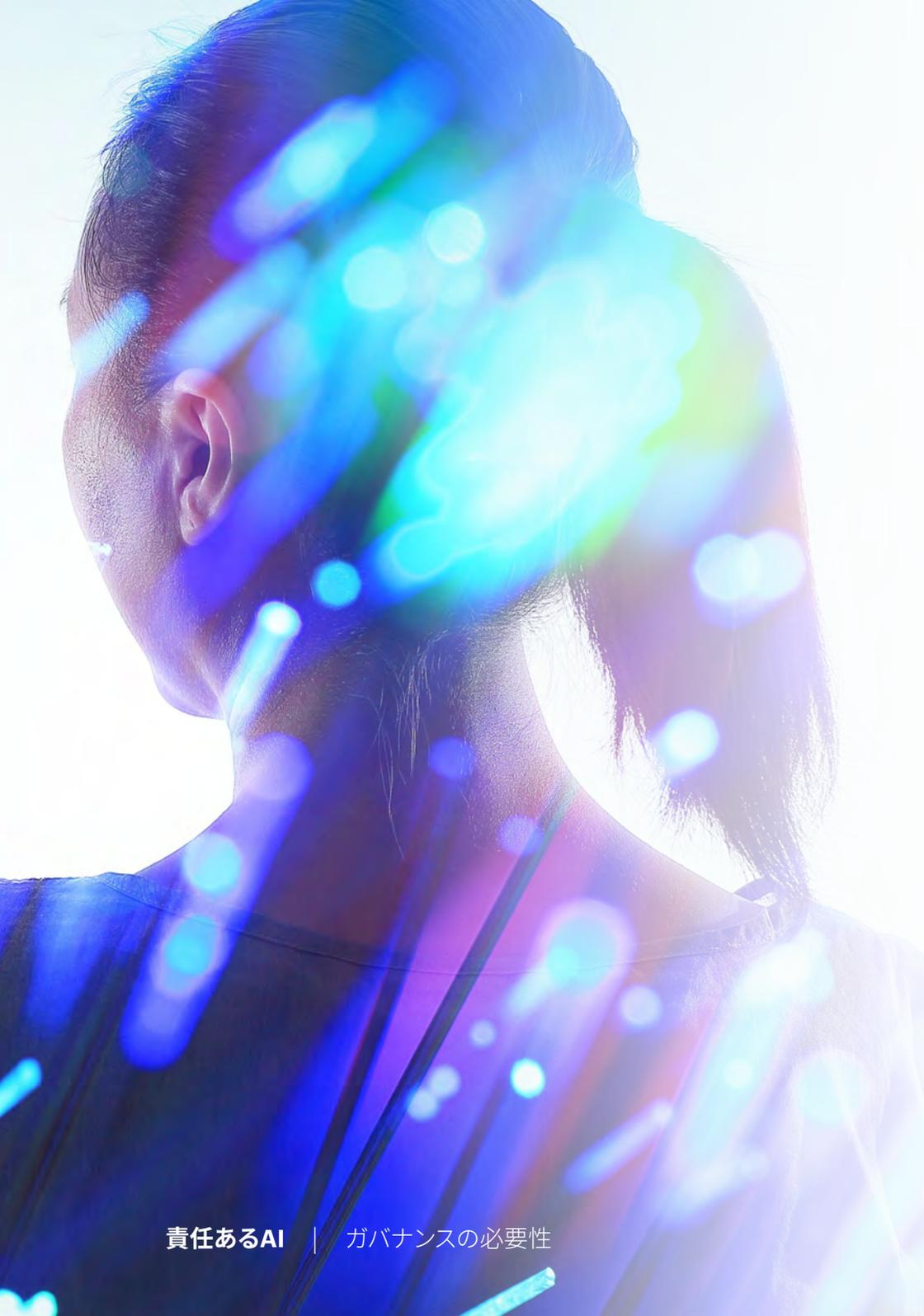
88%がAIを使った意思決定やアウトプットに自信がないと回答。



60%は、少なくとも月に1回はAIシステムが出した結果に対し、人間による手直しが必要と回答。さらに、責任あるAIのシステムに初期から投資しない場合は、後になって高額な修正費用が発生するリスクがあると回答。



24%がAIが導き出す結果に一貫性がない、透明性が欠如している、偏りがある、といった理由により、AIシステムの全面的な改修が必要となったと回答。



AIの倫理的ガバナンスの必要性を認識している組織でも、具体的にどのようにAIの倫理性を担保すればよいか、理解していない場合があります。



63%がAIシステムの監視が重要だと考えているが、ほとんどの回答者はどう監視してよいかわからない、または自身の監視方法が正しいかどうか確信が持てないと回答²。

社会から見たAI倫理ガバナンスの必要性

AI／機械学習が進歩するにつれて、AIのアウトプットから生まれる可能性のある差別に対する、社会全体の不安も増しています。例えば、職業・住居・教育といった機会均等性の崩壊、クレジットカード利用や商品・サービスの価格変動・選択肢の幅など経済面での平等性の崩壊、そしてAIの判断に基づいた監視・逮捕による自由の喪失などがこれに該当します³。組織は責任を持って製品やシステムの設計・開発プロセスを管理し、顕在化している、または今後発覚しうる社会的な負の影響を回避しなければなりません。

データ活用に関する規制は日々厳格化の一途をたどっています。近年、ガバナンスの不足や欠如により消費者の被害が増加していることを受け、個人を保護するための法整備が急がれました。EUの一般データ保護規則

(GDPR) は、ヨーロッパにおける国民一人ひとりのデータプライバシーに関して企業を規制するもので、人々が自らの個人情報を厳しく管理することを可能にすると同時に、そのためのビジネス環境を整えています⁴。

この種の法規制は米国でも支持され、カリフォルニア州では消費者プライバシー法 (CCPA) を可決 (2020年に施行) しました⁵。

上記のような倫理観を考慮せずにAIを導入してしまうと、後追いで社会からの強い要請への対応や法的基準への追従を強いられ、結果、導入済みAIを再設計・大規模改修することになります。今からAIの倫理的ガバナンスに積極的に取り組む方が、組織にとってはるかにメリットがあるのです。

自問してみましょう

自組織にAIガバナンスが必要な状況が存在していませんか？

個人情報（間接的に個人を特定しうるものも含む）または機密情報とみなされるデータを使用するAI／機械学習アルゴリズムを開発していませんか？もしくは、それらの情報を分析した結果として提示するAI／機械学習アルゴリズムを開発していませんか？

個人情報または機密情報とみなされるデータを収集するAI／機械学習アルゴリズムを活用した製品を開発していませんか？

外部ベンダーを使用してAIを開発していませんか？その場合の潜在的なリスク*と影響はどのようなものになるか明示され、関係者の間で共有されていますか？

*リスクの例として、以下が挙げられます

- 過度なベンダーへの依存による、自社の事業継続性への影響リスク
- AIの設計・開発・運用が継続的に連動する状況下で、損害が発生した場合の自社とベンダーとの責任分界点が曖昧になるリスク

組織の倫理原則と統合したガバナンスをAIに求める 仕掛けはありますか？

組織としてのコアバリューの1つである倫理的な行動指針に即し、自らの行動が倫理的であるということを普段からどのように確認していますか？AIにも同様のことを確認していますか？

AIの設計・開発に関するガイドラインは既に存在していますか？その中にAI倫理に関する事項を具体的に織り込むことができますか？

組織の戦略目標はどのようなものですか？その戦略とAIの分析結果が相反しないよう、どのようにAIを組織の中に組み込むことができますか？

法規制やステークホルダーからの要請と整合したガバナンスをAIに求める仕掛けはありますか？

順守すべき国内法規制にはどのようなものがありますか？海外ではどのような法規制の制約を受けますか？その中でAIガバナンスに影響を及ぼす法規制とはどのようなものですか？

消費者やベンダー、顧客企業から、AIを倫理的に運用することに対する要請として、どのようなものがありますか？

競合他社はどのようにAIを運用していますか？業界標準やベストプラクティスはどのようなものですか？

A hand is shown in silhouette, reaching down towards a sunset over a body of water. The sun is low on the horizon, creating a warm orange glow. The water reflects the sunset and the hand, creating a symmetrical image. The background shows a dark, silhouetted landscape under a twilight sky.

02

基本理念の 作成

基本理念は柔軟であると同時に 具体的でなければなりません

AIガバナンスを担う組織は、急速に変化するテクノロジー環境に合わせて進化できるよう汎用性を持ちながら、個々の業界や業務に対応できるよう具体的かつ柔軟な基本理念を確立する必要があります。明確な基本理念は、特定の製品ターゲットや要件に依らず、テクノロジーの進歩と整合し、持続可能な倫理的AIの構築を可能とします。さらに、あらゆる場面で倫理的で責任ある態度と振る舞いを可能にする組織風土を醸成します。

アシモフの「ロボット工学三原則」

50年以上前、SF作家のアイザック・アシモフは、AIが従うべき原則として有名な「ロボット工学三原則」を示しました。アシモフの多くの物語で文学的な仕掛けにもなっている3つの原則は非常にシンプルなもので、「①人間に危害を加えない、②（人間に危害を加えない範囲において）命令に服従する、③（人間に危害を加えず命令に反するおそれのない限りにおいて）自己を防衛する」と定義されています。これらの原則がAIガバナンスのための基礎になると考える人は数多くいますが、この3つの原則には失敗につながる抜け穴があります。例えば、AIが「人間」の識別を誤ったり、これらの原則を正確に実行するために必要な情報を持っていなかったりするケースです。

今日では、ドローンや他の軍事技術のように、より大きな集団を守るために少数の人間に危害を加えることを目的としたロボットもあります。自動運転車は時には他の人を守るために、一部の人間に被害が出ることを想定しなければならないかもしれません（トロッコ問題）。このように、アシモフの三原則は他にも潜在的に起こり得るAIガバナンスのリスクをいくつも包含しています。「ロボット工学三原則」は結果として、多種多様なAIを管理するための基本理念を確立するための課題と必要性を示唆してくれています。

基本理念の作成において考慮すべきこと

高尚で抽象的な目標では、現場で用いられるAIに有効なガバナンスを効かせられない

基本理念は将来を見据えて柔軟性を必要とする一方、「AIは害を及ぼすべきではない」といった高尚で抽象的な目標では、現場で使われるAIにガバナンスを効かせるような具体的方法に落とし込むことが難しくなります。

優れたソリューションは、ルールと理念を両立させる

急速な進化を遂げるテクノロジーに対しても有効なガバナンスを発揮できる基本理念とは、制限事項としてのルールと、途中で変更や修正が可能な理念を両立させたもので、その解釈に明解性と柔軟性の両方を持ち合わせたものです。

厳密な禁止事項・ルールは将来役に立たない

厳密な禁止事項やルールといったものは、テクノロジーのガバナンスや開発に具体的な方針を示すのには有用かもしれませんが、しかし、過度に厳密で技術に特化したルールはテクノロジーの進化の過程で通用しなくなるだけでなく、イノベーションを阻害する可能性さえあります。

アクセントの責任あるAI理念

基本理念の作成にあたっては、各組織が自社独自のコアとなる理念と価値観を反映する必要があります。同時に、将来を見据えて柔軟性を保つために、長期的な戦略目標を考慮する必要があります。それらを考慮した上で、アクセントではTRUSTというAI理念を提唱しています。

- T 信用できる (Trustworthy)**
AIの設計・構築時、安全性を重視し、物事に誠実に向き合い、多様で広い視点を持つ、という実績を1つ1つ積み上げる
- R 信頼できる (Reliable)**
積み上げられた信用から、将来の高度な判断とより良い意思決定への支持を集める
- U 理解できる (Understandable)**
信用を得るためには、AIが透明性を持ち、人によって解釈可能である必要がある
- S 安全が保たれている (Secure)**
信用を得るためには、企業や顧客の情報・データのプライバシーに配慮し、安全性を確保しなければならない
- T 共に学びあう (Teachable)**
このような信用・信頼を勝ち得たAIと人間とが情報交換し、共創し、相互教育をする世界を実現する、人間中心のデザインを目指していく

ガイドライン策定の歴史

1950

SF作家アイザック・アシモフが「ロボット工学三原則 (Three Laws of Robotics)」 (人間に危害を加えない、命令へ服従する、自己を防衛する) を示す

2017.2

日本の人工知能学会 (JSAI: Japanese Society for Artificial Intelligence) が倫理指針を発表

2017.12

UNI Global Unionが「Top 10 Principles for Ethical AI (倫理的なAIのための10大原則)」を発表

2017.1

FLI (Future of Life Institute) が、AIの研究課題、倫理と価値、長期的な課題などを討議した結果をまとめたガイドライン「アシロマAI原則 (Asilomar AI Principles)」を公開

2017.12

IEEE (Institute of Electrical and Electronics Engineers) が「Ethically Aligned Design ver2 (インテリジェント自律システムの倫理に関する一般理念)」を発表

2018.5

トロント宣言「機械学習における平等と差別されない権利の保障」が採択される

2019.3

日本において、AI戦略実行会議・内閣府が「人間中心のAI社会原則」を発表

2019.8

日本において、総務省・AIネットワーク社会推進会議が「AI活用ガイドライン」を策定・発表

2018.4

英国上院のAIに関する特別委員会が「Five core principles to keep AI ethical (AIの倫理的運用のための5大原則)」を策定

2019.4

欧州委員会のAI高等専門家グループが「AI Ethics Guidelines for Trustworthy AI (信頼できるAIに関する倫理ガイドライン)」を発表

自問してみましよう

自組織の中核を担う価値観についてあらためて考えてみましょう。その価値観をどのようにAIに反映していますか？

AIについて、自組織が直面している最も大きな倫理的課題は何ですか？その課題を解決できるよう基本理念を作成することはできますか？

AI活用に関する長期的な戦略目標は何ですか？戦略目標と整合した基本理念を作成することはできますか？

テクノロジーの進化や、AIに対する世論・規制が変化することを想定し、自組織の基本理念を実行可能なものにするには、どのように改定すればよいでしょうか？

基本理念の見直し・再検討する頻度を設定しましょう。基本理念の改定プロセスを既存のプロセスにどのように組み込むと運用可能なものになるでしょうか？

03

倫理的な組織文化の 構築と育成



AIガバナンスの実践には、経営層から現場まで全員が協調しながら建設的に意見を出し合える、倫理的エコシステムが必要となります

AIガバナンスに取り組む先進的な組織の多くがコンプライアンス・プログラムを実施しています。政府が義務づけたコンプライアンス・プログラムに準拠した基準は、倫理基盤としての役割を果たしますが、組織を横断して倫理的な文化を醸成するためには、既存のインフラや倫理概念をどのように拡張していけばよいのでしょうか。

フェーズ1

倫理委員会

倫理委員会とは、倫理や関連分野に精通した専門家の集まりであり、プロジェクト、プログラム、組織のために必要な助言または承認を行います。

フェーズ2

経営トップのコミットメント

経営層は、基本理念を理解、遵守するとともに積極的に推進し、報告体制を整えることで確実に実践できるようにしましょう。

フェーズ3

研修&コミュニケーション

社内の主要なステークホルダー、従業員および契約社員、その他の第三者に対して、基本理念と関連する管理基準について教育を行いましょう。

加えて、影響評価、リスク評価、テストを実施するための研修を行いましょう。

フェーズ4

レッドチームと“消防隊員”

倫理的な問題の発生を未然に防ぐために「レッドチーム」を組織しましょう。

フェーズ5

ポジティブな影響をもたらす倫理指標

ビジネス価値の創出が社会へのポジティブな影響となるような倫理指標を導入しましょう。

フェーズ6

問題提起できる環境

特定の視点だけに凝り固まった画一的な文化を避けつつ、建設的な反対意見を取り入れる仕組みを整備しましょう。

フェーズ1：倫理委員会

“研究とイノベーションのプロセスにできるだけ多くのステークホルダーを巻き込んでこそ、私たちが直面している課題に対する正解を見つけることができます。研究とイノベーションは、社会のニーズと期待に応え、価値観を反映し、責任あるものでなければなりません”

– モイラ・ゲーガン・クイン (Maire Geoghegan-Quinn)
元欧州研究・イノベーション担当委員⁶

倫理委員会とは、倫理や関連分野に精通した専門家の集まりであり、倫理的なAIガバナンスを実現するためのあらゆるフレームワークの基本要素と考えられています。倫理委員会は、プロジェクト、プログラム、組織に対して必要な助言または承認を行います。倫理委員会が組織に関与する度合いは、組織の規模と目標によって大きく異なります。

多様な倫理委員会の在り方を理解するために様々な業種から9つの機関を対象に分析を行いました。倫理委員会を設置・運営するにあたり9つの各機関がとったアプローチについて、①委員会設置の目的・使命、②委員会の権限、③委員構成、④中央集権的ガバナンス、⑤文書化された手続き、⑥ミーティングの運営、の6つの軸で整理したところ、「倫理審査委員会」「倫理諮問委員会」「データ研究審査委員会」「製品諮問委員会」の4種類のタイプに分けることができました。

倫理委員会の定義

「データ研究審査委員会」と「製品諮問委員会」は世間一般での公式名称ではありませんが、共通の特徴を持つ委員会の種類としてアクセンチュアが独自に定義しました。セキュリティと透明性を重視するデータ研究審査委員会は、データ使用における適切な状況と倫理的な公正さを担保します。製品諮問委員会は、製品へのAIの倫理的かつ適切な適用を担保します。



各委員会の特性を示す“6つの軸”

委員会設置の目的・使命

倫理委員会の存在意義を示した声明。委員会の価値観や重視すること、またはその能力を伝えるための文章。

委員会の権限

倫理委員会の権限と管理・監督する範囲についての説明。

委員構成

倫理委員会に必要とされる委員の数。特に、メンバーの多様性の確保が重要（例：ジェンダーの多様性、複数の専門性、科学関係・非科学関係メンバーの混合、系列外のメンバーの加入など）。

中央集権的ガバナンス

倫理委員会の規模と仲介機関の存在の有無。

文書化された手続き

倫理委員会のレビュープロセスと手続きの指針となる重要な規制上の定義や運用上の詳細（例：委員会に提出する書類リスト、書類受領のタイムライン、委員会によるレビューの対象となる研究かどうかの判断など）。

ミーティングの運営

委員の出席状況、開催場所・方法（例：対面、ビデオ会議、電話など）、メンバーの補充に関する手続き。

	倫理審査委員会	倫理諮問委員会	データ研究審査委員会	製品諮問委員会
委員会設置の目的・使命	研究対象者の権利、安全、福祉を保護するために適切な措置が取られていることを、事前に、また定期的な見直しによって保証する。	倫理的なジレンマが発生した場合にプロジェクトに指針を提示することで、研究の質を高めるとともに、社会的影響力を向上させ、一貫性を保つことを目指す。	安全性と透明性を重視し、データ使用における適切な状況と倫理的な公正さを確保する。	製品の倫理的かつ適切な適用を確実にする。
委員会の権限	米国では、FDA（Food and Drug Administration：米国食品医薬局）がヒトを対象とした研究に対して倫理審査委員会への登録を要求できる。倫理審査委員会は、 研究を承認または否認する権限 を持ち、研究対象者の権利、安全、福祉を保護する上で重要な役割を果たす。	プロジェクトにおける 倫理的なジレンマに対する指針を提示 。監督機能を持ち、プロジェクトの特定トピックの 倫理基準が遵守されているかどうかを定期的にチェック することができる。	研究に重点を置き、 社内ポリシーに準じて倫理的実践の指導を行う 。委員会は、重大な倫理違反が確認された場合に プロジェクトを終了させる権限 を持つべき。	製品に焦点を当て、委員会の中核を成す 価値観または理念に則った倫理的実践の指導 を行う。委員会は、重大な倫理違反が確認された場合には、 プロジェクトを終了させる権限 を持つべき。
委員構成	少なくとも5人の委員が必要 。委員は 組織内の各部署のメンバーで構成 され、公開されなければならない ⁷ 。	少なくとも5人の委員が必要 で、法律、データ保護/プライバシー、研究倫理における 組織内外の専門家 で構成されなければならない。同様に、プロジェクトの特定トピック領域における倫理問題の評価に実務経験を持つ 専門家 で構成されている必要がある ⁸ 。	組織によって様々。 組織の内部メンバーで構成 されており、公表しない場合もあり得る。	組織によって様々。ほとんどの場合公開されており、 組織内外のメンバーを含めた多様な専門家 で構成される。
中央集権的ガバナンス	組織外のプロジェクトを審査することが認められているが、多様で専門的な議論に応えられる 専門家の参加 が担保されている必要がある。	ヨーロッパにおいては、欧州委員会から資金提供を受けているプロジェクトのみを審査。	(組織の規模、構成、および既存のガバナンスの枠組みによって大きく異なる)	(組織の規模、構成、および既存のガバナンスの枠組みによって大きく異なる)
文書化された手続き	毎月開催。委員会の審査に先立ち、遅くとも直前の会議に関連資料を提供する必要がある。	組織によって様々。年に複数回開催。	(組織の規模、構成、および既存のガバナンスの枠組みによって大きく異なる) 組織内における倫理審査の対象を特定するため、ギャップ分析を行う。	(組織の規模、構成、および既存のガバナンスの枠組みによって大きく異なる) 組織内における倫理審査の対象を特定するため、ギャップ分析を行う。
ミーティングの運営	企業によって様々。事前の議題共有と議事録の作成が必要。	組織によって様々。対面の会議が推奨される。	(組織の規模、構成、および既存のガバナンスの枠組みによって大きく異なる)	(組織の規模、構成、および既存のガバナンスの枠組みによって大きく異なる)

フェーズ2：経営トップのコミットメント

経営層は、組織内の既存のガバナンスの仕組みと組織文化を考慮した上でコンプライアンスの枠組みを築くべきです。既に組織内に強固なコンプライアンスの仕組みがある場合は、AIの倫理的なガバナンスの構築も容易になります。一個人に発生したコンプライアンス上の問題について、（組織からの「報復」を恐れることなく）自信を持って報告できると従業員が感じられる企業文化であれば、同様に、AIの中に存在する倫理に反した問題を提起することを厭わないでしょう。組織内の既存の構造を活用することで、責任あるAIを中心とした強力で倫理的な組織文化を作ることができるのです。

基本理念の実装

基本理念を定義すれば、それを基に倫理的コンプライアンスの枠組みを拡大する方針を作ることができます。その後、基本理念を浸透させるため、従業員を対象にした研修などの教育プログラムを計画しましょう。

報告体制

既存のガバナンス構造の中に、倫理に反する問題について報告できる窓口・プロセスをどのように設けられるでしょうか。既存のコンプライアンスの枠組みを拡張する形で実現できないかどうかから考えてみましょう。既存のもので不十分な場合は、新たな報告窓口・プロセスの確立を検討しましょう。さらに、指揮系統が機能しない場合に備えて、反対意見を表明できる窓口を用意しましょう。

基準の履行

組織に倫理的な文化を根付かせるためには、倫理委員会によって定められたリサーチ、設計、開発、テストの基準を経営層が積極的に現場に履行させるように動くことが大切です。これはまた、経営層が各チームを成功させるために十分な研修、基準、ツールを整備しなければならないことを意味しています。

フェーズ3：研修&コミュニケーション

責任あるAIに関する研修は、経営層が策定した方針やプロセスに関する研修、リスク評価や影響評価ツールの学習、AI／機械学習に親しむことなどを組み合わせたものになります。研修に組み込むべき、自組織のAIに対する倫理ニーズを理解するために、ギャップ分析を行いましょう。

詳しくは4章の「ギャップ分析」をご覧ください。



フェーズ4：レッドチームと “消防隊員”

アルゴリズムック・バイアスや、倫理に反するAI／機械学習が適用されることで発生する、自社製品に及ぼされる影響を“火事”と想像してみてください。倫理的な問題が火事であるならば、火事が起こるのを未然に防ぐよう、もしくは起きたとしても初期消火により火が燃え広がらないよう、“消火活動”に似た対策を準備する必要があります。この場合、望ましいアプローチは火事が起こる前、もしくは初期段階で注意すべき問題をエスカレーションするというスキルをチーム内に組み込むことです。これにより、“消防隊員”が早期に現場の安全確保に必要なアクションを取る、もしくは現場に取らせるよう指導することができます⁹。

“消防隊員”であり、
分析×法規制の専門知識を有する分析官

消防隊員

“消防隊員”になるための適切なトレーニングを受け、影響評価およびリスク評価を行うツールを準備すれば、“消防隊員”はアルゴリズムに潜むバイアスやAI／機械学習の倫理に反する適用を特定することができるようになります。彼らの主な仕事は現場の手に負えなくなった“火事”を消火し、後始末をすることではありません。警報プロセスに従い、“火事”を未然に防ぐ、または初期消火を行うことにあるのです。

分析×法規制の専門知識

一方、開発チームはデータサイエンスとコンプライアンスの双方について十分な知識を持つ必要があります。これらの理解の元、あらゆる角度から結果を分析し、そこに潜むリスクについての理解を深める必要があります。

誤報をおそれないこと

報告が遅れ後々制御不能な事態に陥るよりも、早期に警報を発した方がたとえそれが誤報に終わったとしてもはるかにリスクを低減できます。開発チームは警鐘を鳴らすことへの恐怖心をなくすよう奨励されるべきです。経験を積んでこそチームはリスクをより正確に予測できるようになり、責任あるAIがもたらす価値を真に理解することができるのです。

レッドチーム

レッドチームとは、テクノロジー業界ではサイバーセキュリティに広く適用されている軍事上の概念です。組織の改善を目的に潜在的な脆弱性や見過ごされていたリスクを特定するために設置され、従業員や競合、その他の人々が自社のAIサービスに持つであろう倫理的な問題意識を予見し、計画に反映させ、ビジネスコンセプトやその検証方法について公正で偏りのない目線でレビューや分析を行います¹⁰。レッドチームはAI／機械学習のガバナンスの文脈でも有用で、AI技術の進化に伴う課題に取り組むことができます。AI／機械学習の技術は今後も飛躍的に進化していくと考えられていることから、その進化したAIから生じる意図しない分析結果が孕むリスクも同様に拡大の一途をたどることが予想されます。従って、ソリューションが設計、開発、そして実装される前に、倫理的な問題を特定する専門のレッドチームが必要になります。

フェーズ5：ポジティブな影響をもたらす倫理指標

国連が2016年に発表した17の持続可能な開発目標（SDGs）は、政府、企業、民間団体が協力し、人々と地球のために包括的で持続可能かつ回復力のある未来を構築するという共通の使命を果たすことを目的としています。SDGsが掲げる17の目標は、社会にポジティブな影響を与えるためのブループリントとして提供されています¹¹。

消費者はこれまで以上に、自らと同じ価値観を持つブランドを選択するようになってきています。世界的な消費者調査では、回答者の62%が企業に「自分が強く関心を持つ課題について行動を起こしてほしい」と望んでいます。

組織は社会に与える倫理的な影響力と、消費者が関心を持つ社会課題に対する組織のブランドの影響力について、より深く考えていく必要があるのです¹²。

倫理的な姿勢が組織文化として根付いている組織のブランドは、消費者からの信用と信頼を勝ち得ています。倫理指標をブランドに反映させるには、組織文化の中で倫理性が優先されている必要があります。

一方、倫理性の問題が過大に評価されていないかにも注意を払うべきです。倫理指標として明確な基準を用意し、それをもとに判断を実施すべきです。

フェーズ6：問題提起できる環境

従業員が気軽に問題を提起できる窓口は、組織のガバナンスフレームワークの中で重要な役割を果たします。コミュニケーションの断絶やリーダーシップの欠如、リスク評価や影響評価が不十分な場合であっても従業員が自信を持って懸念を表明するには、普段のコミュニケーション以外の方法で問題を提起できる方法が必要不可欠となります。さらに、建設的な反対意見は重要視されるべきです。もちろん、反対意見が極端に高い頻度で提起された場合、長期的には生産性を阻害する可能性があります。

画一的な視点の回避

すべての組織は、属性や考え方、意思決定の方法などで特定の視点だけに根差した画一的な文化を形成してしまうリスクがあります。画一的ではなく多様な視点を取り込み、設計、開発、研究に関して皆が共有できる方法を確立するため、反対意見の表明を促進、奨励することが大切です。

インクルーシブデザイン

俊敏性を備えた組織を実現するため、ソフトウェア開発のライフサイクルにデザインシンキングを組み込み、反対意見を奨励する場として活用しましょう。

建設的な反対意見の推奨

建設的な反対意見を表明しやすい組織文化はどうしたら形成することができるでしょうか。問題に迅速に対応するため、倫理委員会の連絡先のように、反対意見の窓口も連絡先を明示し、容易にアクセスできるようにすべきです。

ユースケース

責任あるAIのガバナンスフレームワーク実装について、架空のソフトウェア会社のケースで具体的に見ていきましょう。

ABCソフトウェア社は、様々な小売業者向けに消費者が商品を購入前に試すことができる画像認識プラットフォーム「スマートミラー」を開発しました。顔認証によるユーザー認証を経て使われたスマートミラーの使用履歴は、各自の“好み”の傾向として対象となるすべての小売店で蓄積されます。スマートミラーは顧客の顔の形や体型、肌の色などに基づき、お似合いのスタイルを提案してくれます。

スマートミラー・ユーザーの1人がスマートミラーのデータを分析したところ、女性と識別されたユーザーは男性と識別されたユーザーよりも25%高い製品をおすすめされていることを突き止めました。この結果がオンラインで公開されるとユーザーからの批判が噴出し、スマートミラーのボイコット運動に発展。ABCソフトウェア社の信頼は著しく損なわれました。結果として、小売店におけるスマートミラーの利用率は激減し、ABCソフトウェア社の売上も大きく下がることとなりました。

これを受けABCソフトウェア社は、消費者と小売業者からの信頼を回復し、倫理観を持ったブランド・アイデンティティを再構築する重要性を認識することとなりました。

消費者や小売業者の信頼を失ったABCソフトウェア社は、ガバナンスフレームワークの構築を通じて倫理的な組織文化を全社的に育むために動き出しました。ABCソフトウェア社が行った数々の取り組みは、同社の製品の設計、開発、保守といった活動の中核に責任あるAIがあることを社内外の人々に印象付ける結果となります。TRUSTの理念（P22参照）を用いて、**ABCソフトウェア社は次の手順を踏みました。**

1. 多様なテーマに関する専門知識を持つステークホルダー（特に倫理および法律に詳しい専門家）からなる倫理委員会を設置。
2. 倫理委員会は、必要となる研修、基準、リスク評価、プログラムを特定し作成。さらに、従業員と消費者に良い影響を与えるための評価指標を策定。
3. より倫理的なソリューションを設計・開発できるよう、また、倫理的な問題が発生した場合にリスク評価ツールを用いて適切に対応できるよう、経営層や従業員を教育。問題に実際に対応する“消防隊員”も任命。
4. 倫理的な懸念を提起するための適切な窓口を設置し、従業員に周知。

さらなる改善を目指して

1. より複雑な問題が発生したときに相談できるよう、責任あるAIの取り組みにコミットしている外部グループと提携。
2. レッドチームを結成して、脆弱性とバイアスのあるアルゴリズムを特定。また、倫理委員会で潜在的な解決策を議論するための仮想シナリオを特定。

事故を未然に防ぐユースケース (倫理的なAIガバナンスが既にあったなら)

ABCソフトウェア社が責任あるAIのガバナンス導入を既に行っていたとしたならば、この性別による価格設定問題にどのように対処したのでしょうか？きっと、ガバナンスフレームワークによってユーザーよりも先に従業員が問題を見つけ、経営層に提起していたはずです。

ガバナンスフレームワークを備えたABCソフトウェア社では、従業員がスマートミラーのデータを見直し、女性として識別したユーザーに男性として識別したユーザーよりも25%以上高い製品をおすすめしていることをユーザーより先に発見します。更なる調査で、AIから導かれた倫理に反する結果は、学習時に使用されたコスト最適化を目的とした過去の価格データに依るものであることを突き止めます。

このように、性別による価格設定問題の原因は、データサイエンティスト（すなわち“消防隊員”）によって特定されました。そのデータサイエンティストは、責任あるAIの研修であからさまな価格差は会社の基本理念の1つである信頼性のポリシーに反することを学んでいたため、グループ内の責任あるAIに関する問題の窓口となっているチームリーダーに提起し、リーダーは倫理委員会のメンバーに問題をエスカレーションしました。

さらに、インクルーシブデザインのセッションでは潜在的な課題も見つかりました。参加者グループが、あるジェンダーコミュニティが体系的に排除されてしまう可能性を指摘したのです。加えて、ユーザーからのフィードバックを得て改善するソリューションも提案され、製品を改善することができました。

さらに、倫理委員会は今回学んだ教訓を議論し、ポリシーやプログラム、研修、レッドチームのプロセスを改定するために行動を起こします。



課題に対処するために取られた行動

1. 開発を凍結し、アルゴリズムを所有するチームに関連するツールやフレームワークを使用して問題点を調べるよう指示する。
2. 倫理委員会はその結果を精査し、同様のケースに対応した前例があるかどうかを調べ、短期的・長期的両面での解決に向けた戦略を議論する。必要に応じて、外部の専門家に相談する。
3. 分析結果の透明性に関する自社の基本理念に基づき、取引のある小売業者にサービスについての問題を自主的に伝え、解決に向けた行動をとっていることを公表する。
4. 短期的対応として、性別によりあからさまな価格差が起り得る製品を候補から除外した状態で、おすすめ機能をリリースする。

5. 長期的対応として、価格設定に問題があるアルゴリズムを改善するための社内プログラムを作成し、小売業者と協力して価格設定の差を助長しうるデータを排除していく。
6. ユーザーが自身の意思に基づいて性別やスタイルを選択しおすすめ機能に入力できるような、より人間中心のインターフェースを新たに作成する。

このように、もしABCソフトウェア社が責任あるAIのためのガバナンスフレームワークを備えていたならば、ユーザーが指摘する前に従業員が性別による価格設定の問題に気づき、適切なコミュニケーションチャンネルを通じて問題を提起し、前のシナリオで経験したような負の結果を抑制・軽減できたことでしょう。

自問してみましょう

組織におけるAIガバナンスは、どのように構築されるべきでしょうか？

組織のコア業務は何ですか？

外部の専門家に加わってもらうことで大きな恩恵を受けられそうでしょうか？

ターゲットとする顧客は誰ですか？顧客の生活をより良くするためにAIで何を実現しますか？

独立した外部の倫理委員会を設置することで大きな恩恵を受けられそうでしょうか？

AIの倫理的利用に取り組んでいるのは誰ですか？倫理的な問題にどのように対処していますか？

誰が倫理委員会の委員を選定しますか？

取締役会の役割は何でしょうか？取締役会に対し、どのようなことが期待されていますか？

組織にとって倫理委員会の権限はどの程度が最も適していますか？（例えば、すべてのプロジェクトをチェックし承認行為を行う、という強い権限を有するのか、はたまた、各プロジェクトへの助言にとどめる、という弱い権限を有するのか）

誰の価値観が取締役に反映されますか？誰の声が代表され、誰の声がかき消されていますか？

倫理委員会は組織構造のどこに位置付けますか？

倫理的なAIガバナンスに関して、多様性と専門性がどの程度確保されていますか？

倫理委員会の会議の適切な頻度は？
(毎月、または、年に数回)

倫理的な組織文化を醸成するには何をすべきですか？

従業員がAIの倫理的問題について安心して質問ができる環境をどのように作っていますか？

AI／機械学習のアルゴリズムやそれを使用する製品の倫理的問題を特定（およびその深刻度を評価）するために、従業員はどのような教育を受けていますか？

従業員が発見した問題を伝えるためにどのような窓口がありますか？それは周知されていますか？

従業員が建設的な反対意見を提起できる窓口はありますか？

サービスローンチ前に、製品の倫理的問題を検知するための手順には、どのようなものがありますか？

サービスローンチ前に、製品の倫理的問題を緩和するための手順には、どのようなものがありますか？

倫理的なAI／機械学習を実践するにあたり、場合によってはビジネス上の妥協をする覚悟ができていますか？より倫理的なビジネス活動に役立つ指標はどのようなものですか？

04

ギャップ 分析



所属組織を分析 してみましょ

ビジネスにおいてAIを継続的に改善し進歩させることは、組織の成功に必要不可欠です。製品・サービスの改良を進めるのと同時に、AIに関する倫理的な問題を解決し、新たな問題を予見するポリシーやプロセスを準備する必要があります。



次に紹介するギャップ分析を行う際には、まず自組織の現状を整理した上で、倫理的なAIガバナンスのあるべき将来像を考えてみましょう。現状とあるべき将来像を把握したら、ギャップ分析を行い、次のステップに進むための行動を起こしましょう。

【事例】 裁判前リスク評価ツール

裁判前リスク評価ツールは、被告人が裁判に出頭しなかったり、裁判を待っている間に再犯を起こすリスクレベルを測定するためのツールです。これらのツールは、米国のさまざまな管轄区域をまたいで、犯罪歴、薬物使用歴、学歴、年齢などのリスク要因を用いて公判前リスクを評価しています¹³。

2019年4月、AIの研究団体「パートナーシップ・オン・AI」は、裁判前リスク評価ツールがアルゴリズム上、倫理

的な欠点を抱えていることを指摘する報告書を発表しました。加えて、政府がこれらのツールを活用する前に対処しなければならない追加要件も併せて提案しました¹⁴。

これらのツールのガバナンスを総合的に分析し将来のあるべき姿を描きなおすことで、裁判前リスク評価ツールを提供している組織に対し、倫理的なAIガバナンスを実現するためのロードマップを示すことができるのです。

倫理委員会

現在

裁判前リスク評価ツールを開発している企業に対する倫理的評価と監視状況については、一般に公開されていない。2018年には、120の市民権団体がツールの開発と展開に適切な責任構造がないことが適切な保釈の促進を妨げると指摘し、一般市民が監視できるよう透明性を確保し公開することを求めた¹⁵。

あるべき姿

裁判前リスク評価ツールを開発する企業は、透明性の高いガバナンス体制を持ち、AIの可視化と説明責任の向上に努めるべきである。さらに、製品開発には多様な視点を取り入れ、コミュニティの中で「リスク許容度」を評価指標として設定することでステークホルダーを巻き込むべきである。

経営トップのコミットメント

現在

裁判前リスク評価ツールを開発する企業における基本理念とその実践状況は、一般に公開されていない。

あるべき姿

経営層は、AIの設計・開発に関する方向性を定義し、その実践を促すべきである。裁判前リスク評価ツールが司法制度に活用されることを考えると、外部のステークホルダーとして政府や一般市民と協調してガイドラインを策定すべきである。

研修&コミュニケーション

現在

裁判前リスク評価ツールを開発する企業に対する責任あるAIの設計、開発、モニタリングに関する研修状況は、一般に公開されていない。

あるべき姿

企業は、責任あるAIの設計、開発、モニタリングについて従業員に教育すべきである。例えば、データサイエンティストはデータセット内のバイアスを識別し、評価し、それを軽減できるよう研修を受けるべきである。経営層によって定義された基本理念とプロセスは、組織内での研修実施・受講の優先度を上げるために役立つ。

レッドチームと“消防隊員”

現在

裁判前リスク評価ツールを開発している企業の中には、アルゴリズムが適正であると主張するばかりで、モデルの監査とモニタリングについて情報を開示しない企業がある。

あるべき姿

企業は、アルゴリズムの監査と監視のための標準的プロセスを持つべきである。さらに、問題と脆弱性を指摘するレッドチームを設置すべきである。

ポジティブな影響をもたらす倫理指標

現在

裁判前リスク評価ツールを開発している企業の倫理指標は、一般に公開されていない。

あるべき姿

企業は自社の製品を最適化するために使用している倫理指標を具体的に示すべきである。例えば、裁判前リスク評価ツールの精度目標を100%にすることを掲げてしまうと、ツールは保守的な方向に過剰に最適化されてしまう。結果、裁判前に逃亡するリスクの低い被告人や、再犯を起こすリスクの低い被告人まで不必要に拘留してしまう可能性が出てきてしまう。

問題提起できる環境

現在

裁判前リスク評価ツールを開発する企業における反対意見を提起するための窓口は、一般に公開されていない。

あるべき姿

企業は従業員が倫理的問題を提起するための窓口を整備すべきである。

現状を把握しましょう

- 1** 現在のガバナンス構造やフレームワークはどのようなものですか？
- 2** AI倫理に関するポリシーやプログラムを作成・改定し、製品やデータを研究するチームに助言できる様々な分野の専門家がいますか？
- 3** 組織が重要視するビジネス上の価値観は何ですか？
- 4** AI／機械学習を活用するにあたって、長期の戦略目標は何ですか？
- 5** 現在のガバナンス構造の中で、AI倫理に関する新たなポリシーやプログラムを作成するためのプロセスはどのようなものがありますか？
- 6** 既存の報告プロセスにおいて、従業員はどのようにして倫理に反する行為を報告することができますか？
- 7** 組織内で倫理に反する行為についての報告を受けたとき、解決のためにどのようなプロセスが確立されていますか？

- 8** 従業員に倫理的行動を促すために、どのような研修や教育プログラムを準備・活用していますか？
- 9** ①新製品の設計・開発前に、②サードパーティ製のツールを業務／組織内ツール／顧客向け製品に統合する前に、③データ調査を実施する際に、現在どのようなリスク評価や影響評価を行っていますか？
- 10** 組織内の倫理指標を実践する責任者は誰ですか？
- 11** 設計、開発、またはデータ調査の際に、現在どのような監査の仕組みを活用していますか？
- 12** 建設的な反対意見を表明するにはどのような手段がありますか？

組織のあるべき将来像を構想しましょう

- 1** AIを倫理的に運用するためのガバナンスを既存のフレームワークに組み込むには、どのようにすべきですか？
- 2** 組織に最も適した倫理委員会のタイプは、4つ内どれでしょうか（P29参照）？外部から専門家を招聘する、外部に独立した倫理委員会を設置することで、大きな恩恵は受けられそうですか？
- 3** ポリシーやプログラムの作成・改定、製品の設計・開発、データ調査についての助言を受けるには、どのような分野に詳しい専門家が必要でしょうか？
- 4** 基本理念には何を含めるべきですか？基本理念は組織の重要な価値観と整合していますか？
- 5** 基本理念は、AI技術に関する長期の戦略目標を実現するために十分な柔軟性を備えていますか？
- 6** 既存のガバナンス構造の中で、倫理的なAIガバナンスを実現する新たなポリシーやプログラムをどのように策定することができますか？
- 7** 従業員が倫理に反したAI運用を気軽に報告できるよう、どのような報告窓口を準備すべきですか？

- 8** 倫理に反したAI運用の報告を受けたとき、解決に向けて従うべきプロセスが存在しますか？
- 9** 倫理的なAIに関する組織内ポリシーについて従業員に教育するために、どのような研修や教育プログラムを準備し、受講を推進していくべきでしょうか？
- 10** アルゴリズムック・バイアスを特定し、回避策を打てるよう、どのようなツールを従業員に提供すべきでしょうか？
- 11** ①新製品の設計・開発前に、②サードパーティ製のツールを業務／組織内ツール／顧客向け製品に統合する前に、③データ調査を実施する際に、どのようなリスク評価や影響評価を行うべきでしょうか？
- 12** 倫理的なAI設計、開発、データ研究に関して、どのようなベンチマークを参照すべきでしょうか？
- 13** 組織内で倫理的なAIの設計、開発、またはデータ研究の倫理指標を実践すべき責任者は誰ですか？
- 14** 既存の監査の仕組みの中で、倫理的なAI運用を実現するためのレッドチームはどうあるべきでしょうか？
- 15** 建設的な反対意見を表明するにはどのような手段があるべきですか？

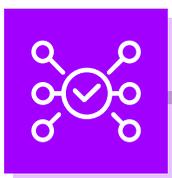
責任あるAIの啓蒙

本ガイドブックが、ここまでお読みいただいた皆様一人ひとりにとって、AIガバナンスの重要性について改めて考え、より自信をもってAIガバナンスを構築・実践するための理解を深める一助となりましたら幸いです。

責任あるAIを所属組織で真に実践するためには、ビジネスのあらゆる側面を考慮する必要があります。ここまでお読みになった方は既に、ガバナンスなど運用において考慮すべき重要な事項について所属組織の現状に照らし合わせて考えはじめていることでしょう。

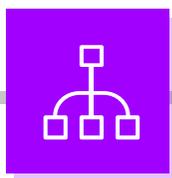
新たなガバナンス戦略の策定や、テクノロジー／組織／ブランドなどに責任あるAIを組み込むにあたって支援が必要な場合は、お気軽にお問い合わせください。

アクセンチュアの責任あるAIへのアプローチ



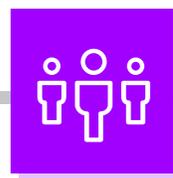
ガバナンス戦略

責任あるAIの基本理念を定義し、その主要な目的およびビジネス成果を構想し、AIの最大活用を可能にする（制度、評価基準、管理を含む）ガバナンス戦略を確立するお手伝いをします。



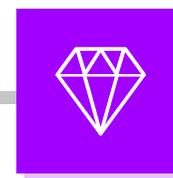
テクノロジー

信頼性が高く説明可能なAIモデル、システム、プラットフォームを設計、構築し、実ビジネスに適用します。



組織設計

人間+マシンの協働を確立するため、新しく必要になる役割を定義するとともに、この新しい働き方に対応するための再教育の仕組みをご提案します。また、追加採用が必要な領域についても特定し、貴組織の採用戦略のインプットを提供します。



ブランド

貴組織の重要な価値観、倫理規範、説明責任に根差したブランディングと、公共の場におけるレピュテーションリスク等を評価し、ビジネスにおける責任あるミッションを明確にします。

執筆者



保科 学世

ビジネス コンサルティング本部 AIグループ日本統括 AIセンター長
アクセントチュア・イノベーション・ハブ東京共同統括
マネジング・ディレクター 博士（理学）

 www.linkedin.com/in/gakusehoshina/



堺 勝信

ビジネス
コンサルティング本部
AIグループ
マネジング・ディレクター



佐々木 三泰

ビジネス
コンサルティング本部
AIグループ
アソシエイト・ディレクター



中畑 良介

ビジネス
コンサルティング本部
AIグループ
シニア・マネジャー



鈴木 博和

ビジネス
コンサルティング本部
AIグループ
シニア・マネジャー

お問い合わせ

責任あるAIに関するご相談やお問い合わせはこちらまでご連絡ください。

アクセントチュア株式会社
〒107-8672
東京都港区赤坂1-8-1
赤坂インターシティAIR
Tel: 03-3588-3000（代）
Fax: 03-3588-3001（代）
Mail: info.tokyo@accenture.com

関連情報紹介

責任あるAI：AI倫理とガバナンス
<https://www.accenture.com/jp-ja/services/applied-intelligence/ai-ethics-governance>

AIに関する最新の知見や調査レポート
<https://www.accenture.com/jp-ja/insights/artificial-intelligence-summary-index>

AIサービス
<https://www.accenture.com/jp-ja/services/ai-artificial-intelligence-index>

出典

- 1 <https://www.accenture.com/us-en/insights/digital/critical-mass-managing-ai-unstoppable-progress>. Accessed 7 December 2018
- 2 <https://www.accenture.com/us-en/insights/digital/critical-mass-managing-ai-unstoppable-progress>. Accessed 7 December 2018
- 3 <https://fpf.org/2017/12/11/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/>. Accessed 21 March 2019
- 4 https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. Accessed 7 December 2018
- 5 <https://www.caprivacy.org/>. Accessed 7 December 2018
- 6 Roles and Functions of Ethics Advisors/Ethics Advisory Boards in EC-funded Projects. (2012, December). Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/ethics-guide-advisors_en.pdf. Máire Geoghegan-Quinn, European Commissioner for Research, Innovation and Science Message delivered at the conference "Science in Dialogue - Towards a European Model for Responsible Research and Innovation" Odense, Denmark, 23-25 April 2012
- 7 Quoted entirely from and referenced in "IDE Institutional Review Board." <https://www.fda.gov/medicaldevices/deviceregulationandguidance/howtomarketyourdevice/investigationaldeviceexemptionide/ucm046745.htm>. Accessed 7 December 2018.
- 8 Roles and Functions of Ethics Advisors/Ethics Advisory Boards in EC-funded Projects. (2012, December). Retrieved December 7, 2018, from http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/ethics-guide-advisors_en.pdf
- 9 <https://www.accenture.com/us-en/insights/artificial-intelligence/ai-governance-support-innovation>. Accessed May 21, 2019
- 10 Benson, K. (2007, November 14). Military Adaptation of Red Teaming. Retrieved December 7, 2018, from http://web.mit.edu/SSP/seminars/wed_archives07fall/benson.htm
- 11 <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>. Accessed May 27, 2019
- 12 https://www.accenture.com/us-en/insights/strategy/Brand-purpose?c=strat_competitiveagilnovalue_10437227&n=mrl_1118. Accessed May 27, 2019
- 13 Desmarais, S. L.; Lowder, E. M. (2019, February). Pretrial Risk Assessment Tools: A Primer for Judges, Prosecutors, and Defense Attorneys. Retrieved May 1, 2019, from <http://www.safetvandjusticechallenge.org/wp-content/uploads/2019/02/Pretrial-Risk-Assessment-Primer-February-2019.pdf>
- 14 Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System. (2019, April 26). Retrieved from <https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/>
- 15 Leadership Conference on Civil and Human Rights et al., The Use of Pretrial "Risk Assessment" Instruments: A Shared Statement of Civil Rights Concerns (2018), <http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf>

AIグループについて

アクセンチュアのAIグループ（アプライド・インテリジェンス）は、データサイエンスとインテリジェントテクノロジーをあらゆる機能に適用し、ビジネスを変革し、成果を迅速かつ大規模に達成できるよう支援します。私たちは業界アナリストからリーダーとして評価されており、お客様企業のインテリジェンス創出のために、インサイト・プラットフォームを活用したAI、機械学習、独自のアルゴリズムやアプリベースのソリューションを提供しています。強力なアライアンスおよびデリバリーネットワークと連携し、あらゆる市場や業界での運用を可能にすることで、お客様企業の価値創出をお手伝いします。アクセンチュアは、あらゆる業界、アナリティクス、テクノロジー、デザインの各分野の専門知識を結集し、新たなビジネス成果を正確かつ大規模に推進します。詳細はwww.accenture.com/jp-ja/services/ai-artificial-intelligence-indexをご覧ください。

本ガイドブックは一般的な情報提供を目的にアクセンチュアのコンサルタントが作成したもので、特定の読者の固有のニーズを考慮したものではありません。本ガイドブックに記載されている事項についてアドバイスや詳細な情報をお求めの場合は、アクセンチュアの担当者にお問い合わせください。

アクセンチュアについて

アクセンチュアは、デジタル、クラウドおよびセキュリティ領域において卓越した能力で世界をリードするプロフェッショナルサービス企業です。40を超える業界の比類なき知見、経験と専門スキルを組み合わせ、ストラテジー&コンサルティング、インタラクティブ、テクノロジー、オペレーションズサービスを、世界最大の先端テクノロジーセンターとインテリジェントオペレーションセンターのネットワークを活用して提供しています。アクセンチュアは53万7,000人の社員が、世界120カ国以上のお客様に対してサービスを提供しています。アクセンチュアは、変化がもたらす力を受け入れ、お客様、社員、株主、パートナー企業や社会へのさらなる価値を創出します。

アクセンチュアの詳細はwww.accenture.comを、
アクセンチュア株式会社の詳細はwww.accenture.com/jp
をご覧ください。