AI 原則実践のための
ガバナンス・ガイドライン

ver. 1.0

令和３年７月９日

AI 原則の実践の在り方に関する検討会

AI ガバナンス・ガイドライン WG

A. はじめに

1．AI ガバナンス・ガイドラインの狙い

我が国は、2019 年 3 月、OECD の AI 勧告案策定に貢献した、統合イノベーション戦略推進会議が決定した「人間中心の AI 社会原則」を公表した。これには、社会全体が主体となり実現すべき AI 社会原則が定められるとともに、この原則を踏まえて、AI の開発・運用等の当事者となる事業者が、各自の AI の開発・運用等の目的や方法等に応じ、実施すべき目標（AI 開発利用原則）を自ら定め、遵守すべきであることが示されている。

AI 社会原則は、①人間中心の原則、②教育・リテラシーの原則、③プライバシー確保の原則、④セキュリティ確保の原則、⑤公正競争確保の原則、⑥公平性、説明責任及び透明性の原則、⑦イノベーションの原則の 7 つの原則から構成される。この AI 原則実践のための企業ガバナンス・ガイドライン（略して「AI ガバナンス・ガイドライン」という。）では、**AI の社会実装の促進**に必要な **AI 原則の実践を支援すべく**、AI 事業者が**実施すべき行動目標**を提示するとともに、それぞれの**行動目標に対応する仮想的な実践例**や **AI ガバナンス・ゴールとの乖離を評価するための実務的な対応例**（以下「乖離評価例」という。）も例示している。ただし、実践例や乖離評価例は参考例であり、網羅的とすることは意図していない。

2．本ガイドラインの法的性格

本ガイドライン自体には、**法的拘束力はない**。このガイドラインは、実施すべき行動目標、実践例、乖離評価例等からなるが、いずれも社会で一定程度共有されている標準的な目標や実践例をまとめたものであるが、同じく法的拘束力のない「人間中心の AI 社会原則」がその普遍的な内容から社会で尊重されているように、本ガイドラインにも **AI システムの開発・運用等に関わる事業者の取引等で広く参照されることや、AI 原則の実践に関するステークホルダーの共通認識の形成を通じて、各社の自主的な取り組みを後押しすることが期待される**。なお、このガイドラインにしたがって体制等を整備しても、必ずしも関連する法令を順守することにはならないことから、関連法令の遵守についても留意されたい。

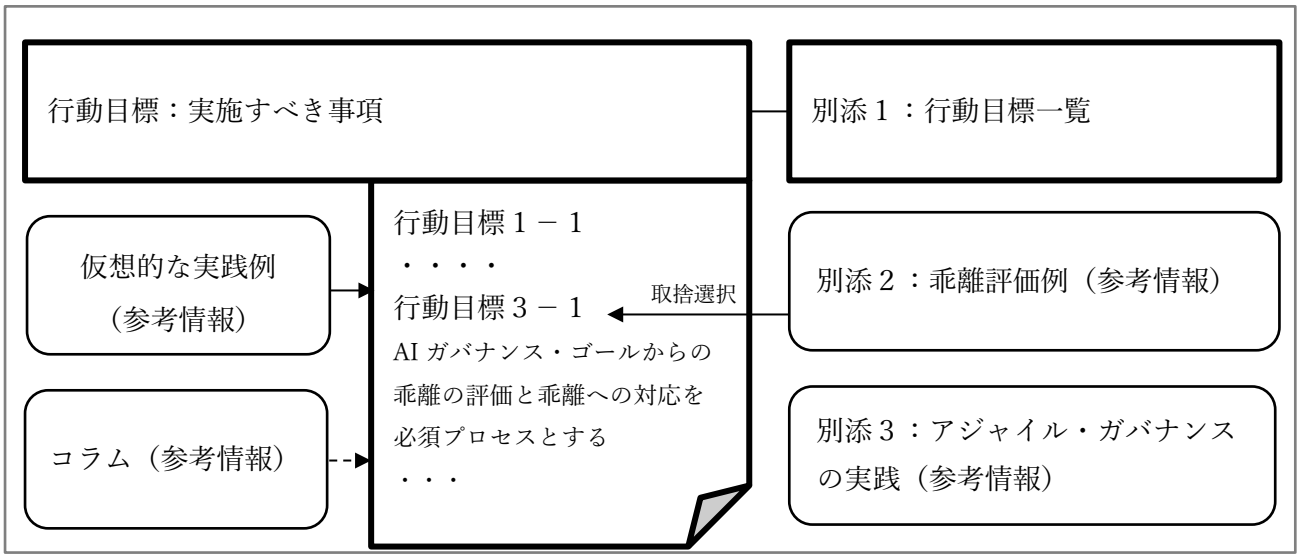
3．他のガイドライン等との関係

本ガイドラインは、新たな要素を加えつつも、国内外で公表されている様々なコード、ガイドライン、アセスメントリスト等のエッセンス同士の接続を重視し、関連文書を統合する**ガイドライン（ガイドラインのガイドライン）を目指したもの**であり、他の関連文書を相互に参照することで、全体として包括的なガイドラインとして機能するように工夫されている。

４．AI ガバナンス・ガイドラインの使い方

本ガイドラインは、本編（行動目標、実践例、コラム）と別添（行動目標一覧、乖離評価例（行動目標３－１の具体化)、アジャイル・ガバナンスの実践）からなる。このうち、**行動目標については、一般的かつ客観的な目標であり、社会に対して一定の負のインパクトを与えうる AI システムの開発・運用等に関わる全ての AI 事業者が実施すべきものである**（行動目標同士の関係は、7~8 頁で後述)。他方で、実践例や乖離評価例は、各 AI 事業者が置かれた個別具体的な状況までは考慮されておらず、特に乖離評価例については、AI システムの開発・運用等の目的、方法、評価の対象によっては不十分であることも過剰であることもありうる。そのため、**実践例や乖離評価例の採否は AI 事業者の任意に委ねられることはもちろん、採用する場合であっても各自の事情に応じた修正や取捨選択を検討する必要がある**。

本ガイドラインの構成の図説



５．Living Document

AI 技術は発展途上にある。また、AI システムの提供によって生じる負のインパクトをステークホルダーにとって受容可能な水準で管理しつつ、そこからもたらされる正のインパクトを最大化するための知恵が社会で蓄積されていくことが予想される。このため、本ガイドラインが改訂なしに将来にわたって適切に機能するとは考えられない。今後も、AI ガバナンスの改善に向け、アジャイル・ガバナンスの設計思想を参考にしながら、マルチステークホルダーの関与の下で、**AI ガバナンス及び本ガイドラインの在り方の検討を継続し、必要に応じて改訂を行うことが不可欠である**。

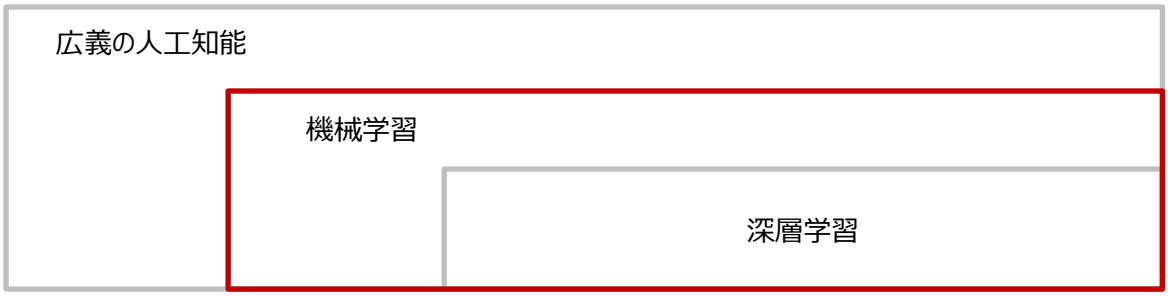
B. 定義

本ガイドラインで使用する用語を以下のとおり定義する。

本ガイドラインは、以下に示すように、**少なくともその一部がデータを用いて帰納的に作成される、機械学習アプローチを用いた AI システムを対象**としている。ただし、**人間の判断を代替しうるものであって、利用者から判断過程が見えにくいソフトウェア等については、機械学習アプローチを用いていない場合であっても、必要に応じて本ガイドラインを参照することが期待される**[1]。

AI システム：深層学習を含む様々な方法からなる、教師あり、教師なし、強化学習を含む機械学習アプローチを用いたシステムであって、人間が定義した特定の目的のために、現実又は仮想環境に影響を与えるような予測、助言、決定を行う性能を有するシステム。この AI システムは設計次第で様々な自律の程度で動作する。この AI システムには、ソフトウェアだけではなく、ソフトウェアを要素として含む機械も含まれる[2]。

赤枠が本ガイドラインの対象範囲



本ガイドラインは、**AI 事業者**（AI システム開発者（＝AI システムを開発する企業）、AI システム運用者（＝AI システムを運用する企業）、データ事業者）を対象としている。

AI システム開発者：自身で運用／他者に提供するための AI システムを開発する者（AI システムの性能維持等のために再学習を行う者も含む。）。

AI システム運用者：自身で利用／他者に利用させるために AI システムを運用する者（一例として、AI システムの開発を依頼せず、AI システムを単に調達して運用する者を含む。）であって、AI システムの運用や性能維持等に一定の責任を負う者。AI システムの

[1] 現時点で人工知能の明確な定義はなく（統合イノベーション戦略推進会議決定『人間中心の AI 社会原則』（2019 年 3 月 29 日）、広義の人工知能の外延を厳密に定義することは適当ではない。
[2] OECD の AI システムに対する定義を参考にしている。欧州委員会の AI 規則案では、AI システムはソフトウェアだけを指している。OECD の AI システムの定義はソフトウェアに限定されていない。

法的な権利者と必ず一致するわけではないが、そのような場合が多いと考えられる。

AI システム利用者：他の AI システム開発者が開発した AI システムや他の AI システム運用者が提供する AI システムを単に利用する者であって、AI システムの運用や性能維持等に責任を負わない者。なお、**AI システム利用者には、ビジネスで利用する者やビジネス以外で利用する消費者等が含まれるが、**本ガイドラインは、これらを明示的に定義するよりも、**リテラシー水準に応じて柔軟に利用者を把握する方が望ましい**と考えている[3]。

データ事業者：AI システムの学習等のために、不特定多数からの収集したデータ、特定の者から取得したデータ、事業者が自ら用意したデータのいずれか又はそれらを組み合わせたデータ、あるいはそれらに加工を施したデータを他者に提供する者。

＊1 つの AI 事業者が同時に複数の役割主体に区分されることもある。たとえば、AI システムの開発も運用も同一の企業が行う場合には、この企業は AI システム開発者でもあり、AI システム運用者でもある。

本ガイドラインは、企業内の実施主体を以下の二層を想定しながら整理している。なお、両者の区分が明瞭でない小規模企業等は、特定の者や集合が両方の責務を負っているものと理解して差し支えない。

経営層：株主や株主以外の様々なステークホルダーに対する適切な情報開示やステークホルダーとの協働を確保するとともに、健全な事業活動倫理を尊重するためのマネジメントシステムの確立のために運営層に大きな方向性を示すことを責務とする者又はその集合

運営層：経営層が示した方向性にしたがって健全な事業活動倫理を尊重するためのマネジメントシステムを設計し、運用し、環境・リスクの分析、ゴールの設定、マネジメントシステムの評価を実施ないし支援する者又はその集合。

[3] 利用者のリテラシーについて一定の目安を提示するものに、OECD, OECD FRAMEWORK FOR THE CLASSIFICATION OF AI SYSTEMS – PUBLIC CONSULTATION ON PRELIMINARY FINDINGS (May 2020)がある。ここでは、研修を全く受けていないアマチュア、特定のシステムについて研修を受けた者、AI に関する知識を有する専門家に区分している。https://aipo-api.buddyweb.fr/app/uploads/2021/05/Report-for-consultation_OECD.AI_Classification_final.pdf.

C. AI ガバナンス・ガイドライン

AI システムは、人材不足の解消、生産性の向上、高付加価値事業の開発など、ビジネスにとって正のインパクトをもたらしうる一方で、AI システムの開発や運用には、意図せずして、公平性を損なってしまったり、安全性の問題が生じたりするなど、AI 特有のリスクも伴うことから、AI システムを開発・運用する企業は、自分事として AI システムに関わる環境や価値提供モデル全体を理解すべきである。各企業には、**行動目標を杓子定規に実施するのではなく**、**このガイドラインを伴走相手として各行動目標の意義を理解**した上で、活用してもらうことを期待している。**そして、行動目標の意義を理解すれば、それらが実施すべき標準的な事項でありながら実施にあたっての柔軟性も同時に備えていることがわかる**だろう。

このガイドラインは**環境・リスク分析**から始まる。企業単位（場合によっては事業部単位）の方針を決めるにあたっては、**AI システムがもたらしうる正負のインパクト**、**AI システムの開発や運用に関する社会的受容、**そして自社の事業範囲等に照らして負のインパクトが軽微ではないと判断した場合には、**自社の AI 習熟度（AI システムの開発・運用時に求められる準備がどれだけできているのか）**を考慮すべきである。

これらを踏まえ、たとえば、潜在的な負のインパクトの大きさを考慮しつつ AI 分野における自社の経験不足や現在の社会的受容に照らしてそもそも AI システムを開発・運用しないという方針、潜在的な負のインパクトが軽微な分野に限定して、AI システムを開発・運用していくという方針、潜在的な負のインパクトを管理しつつ AI システムを開発・運用していくなどの方針を策定することになるだろう。そして、AI システムを開発・運用する場合には、潜在的な負のインパクトの性質や大きさを考慮しながらステークホルダーにとって受容可能な水準に管理する際の羅針盤となる企業単位（場合によっては事業部単位）の **AI ガバナンス・ゴール（たとえば AI ポリシー）を設定するか否かについて検討すべき**であり、潜在的な負のインパクトが軽微であることを理由に **AI ガバナンス・ゴールを設定しない場合には、その理由等をステークホルダーに説明できるようにしておくべき**である。

次に、AI ガバナンス・ゴールを達成するための **AI マネジメントシステムの設計**が求められる。具体的には、**AI ガバナンス・ゴールからの乖離の評価と乖離への対応**、**AI マネジメントシステムを担う人材のリテラシー向上**、適切な情報共有などの**事業者間・部門間協力による AI マネジメントの強化**、インシデントの予防や早期対応を通じた**インシデントに関わる AI システム利用者の負担軽減**を挙げることができる。

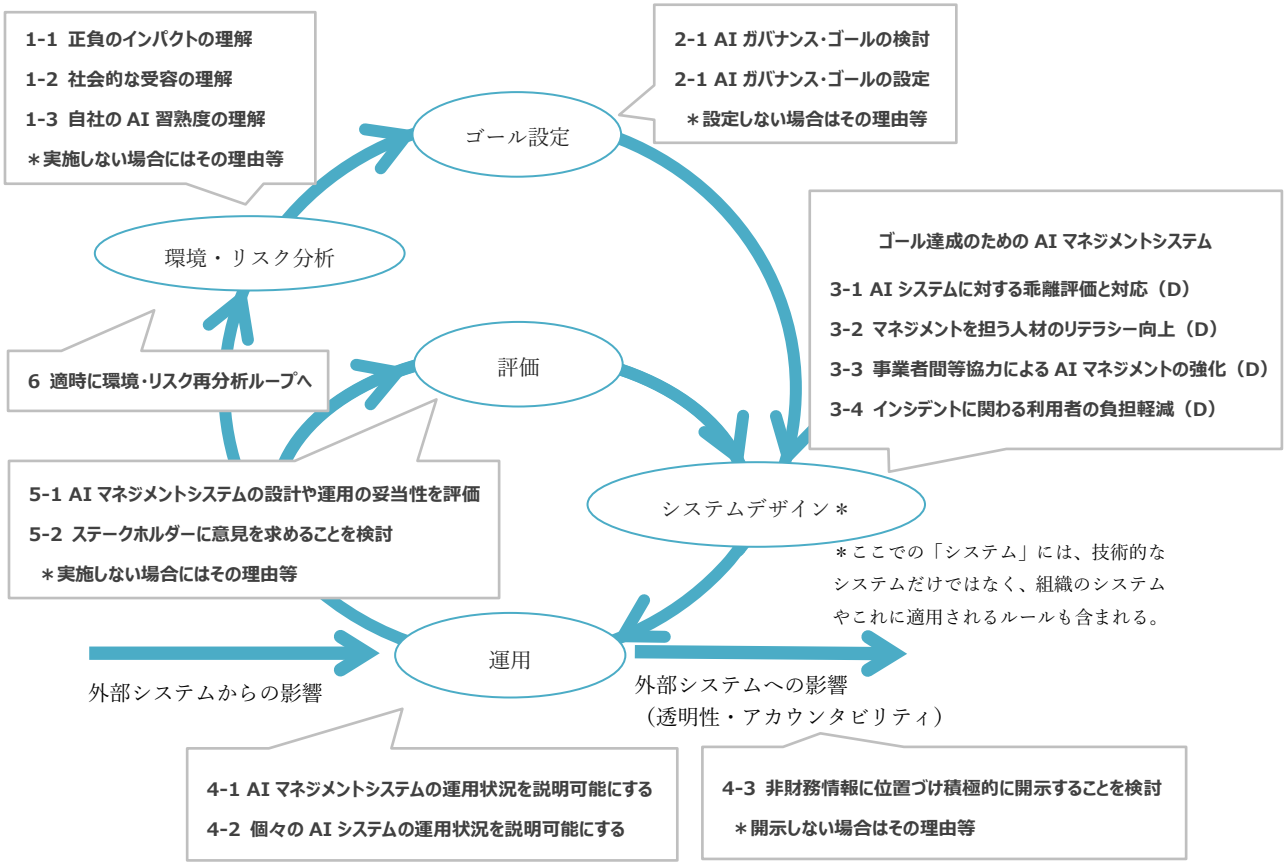
AI 技術の発展にアジャイルに適合していくためには、AI ガバナンス・ゴールや AI マネジメントシステムを**継続的に評価**する必要がある。まず、**AI マネジメントシステム及び個々**

の AI システムの運用状況について説明可能な状態を確保すること。また、ステークホルダーとの一層円滑なコミュニケーションのために、これらの情報をコーポレートガバナンス・コードの非財務情報に位置づけ、積極的に開示することを検討すべきであり、開示しない場合には、その理由等を説明できるようにしておくべきである。

次に、AI マネジメントシステムの設計や運用から独立した者に、その設計や運用の妥当性を評価させるべきである。上述の運用状況に関する情報を用いながら社内で妥当性の評価を実施すべきことはもちろんのこと、株主だけではなく、ビジネスパートナー、消費者、AI システムの適切な運用をめぐる動向に詳しい有識者などのステークホルダーに意見を求めることを検討すべきであり、必要に応じてそのような機会を積極的に設けることもありうる。

さらに、AI ガバナンス・ゴールの設定自体の妥当性を検証するために、AI システムがもたらしうる正負のインパクト、AI システムの開発や運用に関する社会的受容、自社の AI 習熟度からなる環境・リスクの再分析を適時に実施すべきである。

AI システム開発者・運用者のアジャイル・ガバナンス（(D) ではデータ事業者への言及あり）



1．環境・リスク分析

(1) AI システムがもたらしうる正負のインパクトを理解する

> 行動目標１－１：AI システムを開発･運用する企業は、経営層のリーダーシップの下、AI システムから得られる正のインパクトだけではなく意図せざるリスク等の負のインパクトがあることも理解し、これらを経営層に報告し、経営層で共有し、適時に理解を更新すべきである。

【実践例１】

AI システムは、新規ビジネスを生み出したり、既存ビジネスの付加価値を高めたり、生産性を向上させたりするなどの正のインパクトをもたらしてくれるが、反面、負のインパクトがあることも忘れてはならない。これらの正のインパクトを最大限享受するためには、負のインパクトや意図せざるリスクについても理解し、正のインパクトとのバランスを検討する必要がある。そのため、**AI システムを開発・運用する企業は、経営層のリーダーシップの下（すなわち、組織内で気運を作り出し、運営層の活動を支えることを通じて）、正のインパクトだけでなく負のインパクトについても検討し、その検討結果を経営層で共有するとともに、適時に理解を更新する必要がある**。

AI システムを開発・運用しようとする企業には正のインパクトは既知と思われるが、当社は、独立行政法人情報処理推進機構（IPA）がまとめている AI 白書のような包括的・網羅的な解説書等を用いて**AI 技術がもたらしうる正のインパクトを改めて整理した**。

また当社は、**これから開発・運用しようとしている AI システムと同じ又は類似の機能や分野においてインシデントが過去に起きていないか、あるいは、過去に起きていないとしてもインシデントが起きる具体的な可能性が指摘されていないかについて調査した**。インシデント情報は様々な文書やインターネットから入手することができる。当社は、日本での開発・運用のみを予定しているため、日本で共有されている情報の収集から始めた。その際、**消費者庁の「AI 利活用ハンドブック～AI をかしこく使いこなすために～」は良い出発点となりうる**[4]。たとえば、消費者のチェックポイントに「AI が音声を誤認識してしまい、間違えた指示をしたり、普段の会話情報が収集されてしまう可能性があります。」が挙げられているが、これは潜在的なインシデントを消費者の視点から表現したものである。また、インシデント

[4] 消費者庁「AI 利活用ハンドブック～AI をかしこく使いこなすために～」(2020 年 7 月発行)、https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/review_meeting_004/ai_handbook.html.

や将来起きうることに言及した**AI に関する書籍も充実**している。日本ディープラーニング協会（JDLA）の G 検定は倫理的事項も対象としており、この検定の一環としてインシデントに関する情報を得ることができる。「プロファイリングに関する提言案付属 中間報告書」では、いくつかのケースがわかりやすく解説されている[5]。さらに、当社では、**AI システムに対する社会的受容が国・地域ごとに異なりうることを認識しつつも、後述する「コラム：インシデントの共有」で挙げられているインシデントデータベースも参考にした**。これまでの分析では、個人情報の取扱い、公平性、安全性に関するものが多いことがわかっている。なお、個々の具体的な AI システムのインパクト分析は、行動目標３－１の乖離評価時に行う予定である。

【実践例２】

当社は、開発・運用している AI システムの分野が多様であることから、実践例 1 に加え、**社会的に負のインパクトを及ぼしたインシデントや負のインパクトを及ぼす可能性が指摘されている将来的課題について、その全体像を把握するために、一般的なフレームワークに照らしながら、大まかに整理している**。当社では、独自のフレームワークを用いているが、最近は OECD の分類フレームワークの議論の進展に注目している[6]。正式版が 2021 年 9 月から 10 月にかけて公表されるようである。環境・リスク分析に概ね対応する CONTEXT の章では、OECD の AI 原則と産業分野の関係性、ビジネス用途、影響を受けうるステークホルダー、影響の範囲などの視点から一般的なフレームワークが提示されている。これらの分類は負のインパクトを大まかに理解するための補助的ツールにすぎないことに留意しつつ、現在、自社のフレームワークへの反映を検討している。なお、個々の具体的な AI システムのインパクト分析は、行動目標３－１の乖離評価時に行う予定である。

【実践例３】

当社は、AI システムの開発・運用の範囲が広く、インシデントが生じると社会に大きな影響を与えることを自覚している。そのため、AI の正と負のインパクトについては、**自社が直接関与した経験と同業他社や場合によっては他の業界の経験から得られた情報を組み合わせることでより有用性の高い分析が可能**であると考え、文理横断的な社内勉強会等で分析

[5] パーソナルデータ＋α研究会『プロファイリングに関する提言案付属 中間報告書』（2018 年 12 月 19 日）
[6] OECD, OECD FRAMEWORK FOR THE CLASSIFICATION OF AI SYSTEMS – PUBLIC CONSULTATION ON PRELIMINARY FINDINGS (May 2020), https://aipo-api.buddyweb.fr/app/uploads/2021/05/Report-for-consultation_OECD.AI_Classification_final.pdf.

を行っている。そして、この分析を一定の頻度で継続することで、**インシデントが起きる前であっても、適時に AI ガバナンス・ゴールの見直しについて検討できるようにしている**。

コラム：インシデントの共有

AI システムの開発や運用に伴う負のインパクトについては過去のインシデントに学ぶしかないという指摘が多い。AI システムはデータセットに基づいて帰納的に構築され、その負のインパクトには意図していないものも多いことから、負のインパクトを低減するためには過去のインシデントを理解することが有効である。これらのインシデント事例は、一般的にはニュースや論文等の公開情報から得られることになるが、必要な情報にアクセスすることは簡単なことではない。

このアクセス性の課題に対応するために、Partnership on AI は、2020 年 11 月に The AI Incident Database（AIID）をリリースした[7]。AIID には 1000 以上のインシデントが URL リンク付きで掲載されており、検索用のアプリも提供されている。Partnership on AI 以外にも、GitHub 上で AI Incident Tracker が公開されている[8]。

他方で、このようなデータベースを持続的に整備することは課題であるようだ。AIID のインシデント事例は学者らが提供した初期リストがほとんどであるという。AI の提供が進み情報が増加する中、重要な情報を中心に蓄積していくことも課題であるという。また、公開情報になっていない各社の「ヒヤリ・ハット」はそれ自体が重要な経験であって各社の知的財産と評価しうる場合もあることから、インシデント事例を積極的に集めて「共有財産」とすることは容易ではないという指摘もある。

[7] Sean McGregor, When AI Systems Fail: Introducing the AI Incident Database (November 18, 2020), https://www.partnershiponai.org/aiincidentdatabase/.
[8] jphall663, awesome-machine-learning-interpretability, https://github.com/jphall663/awesome-machine-learning-interpretability/blob/master/README.md#ai-incident-tracker.

(2) AI システムの開発や運用に関する社会的受容を理解する

> 行動目標１－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、本格的な AI の提供に先立ち、直接的なステークホルダーだけではなく潜在的なステークホルダーの意見に基づいて、社会的な受容の現状を理解すべきである。また、本格的な AI システムの運用後も、適時にステークホルダーの意見を再確認するとともに、新しい視点を更新すべきである。

【実践例１】

AI システムを開発・運用する企業は、経営層のリーダーシップの下、潜在的なステークホルダーの意見に基づいて社会的な受容の現状を理解すべきである。**AI は比較的新しい技術であるため、AI システムを開発・運用する企業と利用者との間の AI への理解度に差が生まれやすい**ことを強く意識すべきである。

当社では、政府、公的機関、シンクタンク等が公表している**消費者アンケートを最初の手がかりとした**。たとえば、**消費者庁は、**「消費者のデジタル化への対応に関する検討会ＡＩワーキンググループ」において、①消費者の AI に関する理解の状況、②消費者による AI への期待と課題、利用意向、③消費者が利用している AI 提供サービス（どのようなリスクを抱えているか）、④AI のサービスに係るリスクについて、どの程度認識・理解して使用しているかについて、**アンケート調査を実施し、その結果を公表している**[9]。当社は、**国際的な展開も考えていることから、海外の消費者のアンケート調査も参考にした**[10]。さらには、**AI システムに対する市民団体の意見も参考にした**。

ここで得られた社会的な受容に関する情報は、AI ガバナンスの全体的な設計の際に用いられることになるため、経営層が意思決定できるように、枝葉をそぎ落として、幹となる情報を抽出することが求められる。当社では、行動目標１－１で得られた情報や分析を活用しながら、**様々な AI システムを、**いかなる説明をしても社会的に理解が得られる水準に達していない可能性が高い用途、積極的かつ十分に説明することで社会的に理解が得られる可能性が高い用途、必要に応じて説明することで社会的に理解が得られる可能性が高い用途、消

[9] 消費者庁は「消費者のデジタル化への対応に関する検討会 AI ワーキンググループ」において、AI に関する消費者アンケートを実施している。たとえば、同ワーキンググループにおける第２回消費者アンケート結果は以下のウェブサイトから入手可能である。
https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/assets/consumer_policy_cms101_200616_1.pdf.

[10] たとえば、BEUC など。https://www.beuc.eu/publications/survey-consumers-see-potential-artificial-intelligence-raise-serious-concerns/html.

費者に負のインパクトを与える可能性が低い用途など、**負のインパクトの大きさにしたがって区分するなどして、リスクベースで社会的受容を整理している**。

【実践例２】

当社では、実践例１に加えて、大学や産業団体が開催する **AI 倫理や品質に関するセミナーやカンファレンスに担当者を積極的に派遣している**。最近では、これらのセミナー等がウェビナー形式で開催されることも多く、以前よりも効率的に情報が得られるようになってきた。海外のウェビナーにアクセスすれば、AI 倫理や品質の国際的な動向を把握することも可能である。

【実践例３】

これまで当社では実践例２のような実務を採用してきたが、AI システムを本格的かつ広範に開発・運用していることから、当社が適切に AI を利用することに対するステークホルダーからの期待が比較的高いと理解している。そのため、経営層のリーダーシップの下、ステークホルダーの意見を間接的・受動的に把握するのではなく、直接的・積極的に把握するという方針に切り替えた。

この新しい方針の下、当社では、**AI の社会的受容の事情に詳しい有識者を招聘し有識者会議を定期的に開催**している。当社の AI マネジメントシステムや運用に対する評価を得るだけではなく、**AI に対する一般的な社会的受容など、当社が置かれた環境への理解を深めるためにも、この有識者会議を活用**している。また、実践例１、２で得られる一般的な情報と比較して、**有識者会議で得られる情報は当社向けに深掘りされたもの**であり、かつ、広く知られていない情報であることが多いという特徴があると認識している。そして、この有識者会議で得られた情報と実践例１、２で得られた一般的な情報と組み合わせて、社会的な受容についてリスクベースで精緻に分析している。分析結果は運営層で整理され、運営層から経営層（業務執行担当）に報告されている。

【実践例４】

当社は実践例３と概ね同じ取り組みをしているが、**有識者会議で得られた意見を取締役会に直接報告している**点で異なる。**当社向けに深掘りされた意見を直接聞けることから、AI 倫理や品質に対する経営層の感度が高まった**と認識している。たとえば、AI 倫理や品質が全従

業員の必須の研修科目になるという目に見える効果があった。AI 倫理や品質を経営層の課題として強く感じてもらうためには、このような仕掛けも有効である。

コラム：AI 習熟度の高め方

行動目標 1 － 3 に先立ち、AI 習熟度に関する一般的な情報を提供したい。Google Brain の共同設立者であり、スタンフォード大学教授でもある、Andrew Ng 氏がまとめた Landing AI のプレイブックには、まずは小さくても意味のある成功を目指し、その成功を梃子に AI システムを社内外に広げいく AI 習熟度向上モデルが示されている。RIETI の BBL セミナー『ディープラーニングの最前線と活用への課題』において、スピーカーの井崎武士氏から、日本の AI 活用の成功事例は Andrew Ng 氏のプレイブックに沿っており、シリコンバレーのようなソフトウェア企業だけではなく製造業においても Landing AI のプレイブックは有効であるとの指摘があった。

Landing AI のプレイブックには技術的な側面の記述が多く、AI 習熟度に合わせたリスク管理の発展の在り方への言及はほとんどない。AI 技術の社会実装をさらに促進するためには、この AI ガバナンス・ガイドラインの行動目標を用いて、AI 原則の実践に関するベストプラクティスを整理し、共有していくことも重要になってくるだろう。

(3) 自社の AI 習熟度を理解する

> 行動目標１－３：AI システムを開発・運用する企業は、経営層のリーダーシップの下、行動目標１－１、１－２の実施を踏まえ、自社の事業領域や規模等に照らして負のインパクトが軽微であると判断した場合を除き、自社の AI システムの開発・運用の経験の程度、AI システムの開発・運用に関与するエンジニアを含む従業員の人数や経験の程度、当該従業員の AI 技術及び倫理に関するリテラシーの程度等に基づいて、自社の AI 習熟度を評価し、適時に再評価すべきである。負のインパクトが軽微であると判断し、AI 習熟度の評価をしない場合には、その理由等をステークホルダーに説明できるようにしておくべきである。

【実践例１】

AI システムのビジネスへの導入、すなわち AI システムを利用して生産過程やサービス提供のオペレーションを効率化することに成功した場合は、人材不足の解消、生産性の向上、高付加価値事業の開発などのビジネスにとって正のインパクトをもたらしうる。一方で、野放図な AI システムのビジネス提供は、意図せずして、公平性を損なってしまったり、安全性の問題が生じたりするなど、AI 特有のリスクも伴うことから、AI 事業者には、これらの AI 導入の負の側面とも言うべきリスクをよく把握した上で、導入にとりかかることが求められる。そこで、**AI の負のインパクトへの対応力を見える化する AI 習熟度（AI システムの開発・運用時に求められる準備がどれだけできているのか）という指標が重要**になる。

当社では、AI システムの開発・運用の際に正のインパクトだけに気をとられて、負のインパクトやリスクへの配慮が不足して、結果として AI システムの導入により他の事業者が大きなダメージを被ったりすることのないように、**経営層のリーダーシップの下、自社の AI 習熟度を評価し、適時に再評価している**。

AI 習熟度の評価には、日本経済団体連合会の『AI 活用戦略　～AI-Ready な社会の実現に向けて～』（2019 年 2 月 19 日）の「AI-Ready 化ガイドライン」を用いている[11]。その理由は、**自社の AI システムが社会に与えるインパクトの大きさ及び関連するステークホルダーの広がりが、自社の AI 習熟度に相応しているか否かについて評価するためである**。そして当社

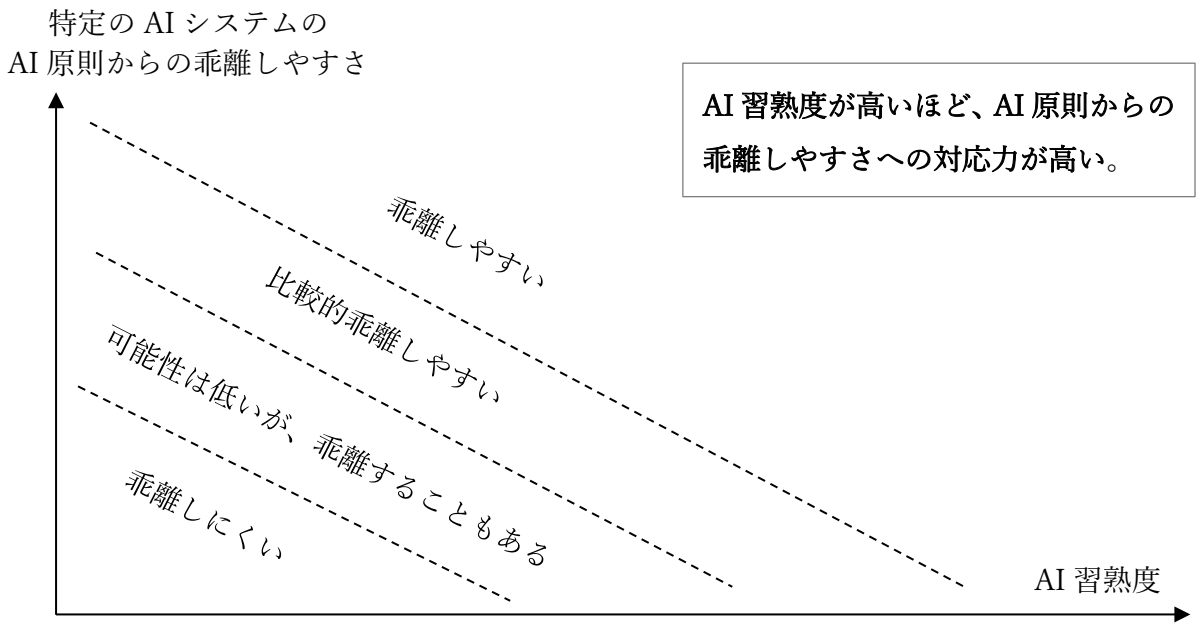
[11] 日本経済団体連合会『AI 活用戦略～AI-Ready な社会の実現に向けて～』（2019 年 2 月 19 日）、https://www.keidanren.or.jp/policy/2019/013_honbun.pdf. 一覧性のある表は https://www.keidanren.or.jp/policy/2019/013_sanko.pdf から入手可能である。

は、**AI 習熟度を AI ガバナンス・ゴールの検討を含む、AI ガバナンス全体の検討に役立てている**。

【実践例２】

当社は、AI システムを自前で開発せず、AI システムの開発を外部に依頼し、納品された AI システムを運用し、AI システム利用者に提供している。このように当社は、AI システム運用者にすぎないことから、当社の利用者向けに AI システムを提供しはじめた当初は自社の AI 習熟度には関心を持っていなかった。しかし、AI システム利用者からの苦情が増え、当社が「期待」したとおりに AI システムが動作していないことがわかり、その後、AI システム開発者との適切な意思疎通だけではなく、意思疎通を支えるリテラシーにも問題があることがわかってきた。

日本経済団体連合会の「AI-Ready 化ガイドライン」は、AI システムを開発する企業向けであると思っていたが、**外部に開発を依頼する AI システム運用者にも関連する AI 習熟度の指標**も含まれており、今では、それらを自社向けに再構成し、経営層のリーダーシップの下、自社の AI 習熟度を評価し、適時に再評価している。たとえば、**自社の AI システムが社会に与えるインパクトの大きさ及び関連するステークホルダーの広がりが、自社の AI 習熟度に相応しているか否かを評価する際に AI Ready 度を用いている**。そして当社は、**AI 習熟度を AI ガバナンス・ゴールの検討を含む、AI ガバナンス全体の検討に役立てている**。



AI 原則からの乖離しやすさへの対応力に関する概念図

２．ゴール設定

(1) AI ガバナンス・ゴールの設定を検討する

行動目標２－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、「人間中心の AI 社会原則」を踏まえ、AI システムがもたらしうる正負のインパクト、AI システムの開発や運用に関する社会的受容、自社の AI 習熟度を考慮しつつ、設定に至るプロセスの重要性にも留意しながら、自社の AI ガバナンス・ゴール（たとえば AI ポリシー）*を設定するか否かについて検討すべきであり、潜在的な負のインパクトが軽微であることを理由に AI ガバナンス・ゴールを設定しない場合には、その理由等をステークホルダーに説明できるようにしておくべきである。「人間中心の AI 社会原則」が十分に機能すると判断した場合は、自社の AI ガバナンス・ゴールに代えて「人間中心の AI 社会原則」をゴールとしてもよい。なお、ゴールを設定しない場合であっても、「人間中心の AI 社会原則」の重要性を理解し、行動目標３から５に係る取り組みを適宜実施することが望ましい。

＊AI ガバナンス・ゴールには、AI 社会原則への対応事項のみからなる AI ポリシーだけではなく、AI 社会原則への対応事項を包含しつつそれ以外の要素を含むデータ活用ポリシー等も含まれる。AI ポリシーと呼称するか否かは各社に委ねられていることは当然である。AI ガバナンス・ゴールの代表例は様々な文献で紹介されている[12]。

【実践例１】

当社は、AI システムの開発・運用を開始して間もなく AI 習熟度はそれほど高くないため、**当面は社会に対する潜在的な負のインパクトが軽微な用途の AI システムのみを扱う予定**である。そのため、当社は**AI ガバナンス・ゴールを設定していない**が、潜在的な負のインパクトが軽微とは言えない用途まで事業範囲を拡大する際には、AI ガバナンス・ゴールの設定について検討するつもりである。もちろん、検討内容を記録し、AI ガバナンス・ゴールを設定しない理由等をステークホルダーに説明できるようにしている。

【実践例２】

当社は、AI 技術の開発・運用を開始して間もないが、一部の AI システムの用途は潜在的

[12] 一例として、舟山聡『AI の責任と倫理（第２回）AI 倫理に対する企業の取組み(1)』NBL No. 1170 (2020.5.15) 第 75 頁に、日米独等の例がまとめられている。

な負のインパクトが軽微であるとは言えないため、AI ガバナンス・ゴールの設定を検討したところ、**プライバシーの保護など、当社が重視すべき事項が「人間中心の AI 社会原則」に適切にまとめられているため、当面は「人間中心の AI 社会原則」をゴールにする**こととし、「人間中心の AI 社会原則」の 7 つの原則を尊重することを徹底している。たとえば、運用の現場にも 7 つの原則を尊重してもらうべく、個々の職場での e-ラーニングを含む研修を通して意識の共有を図っている。将来的に AI システムの開発・運用の範囲が拡大した段階においては、自社独自の AI ガバナンス・ゴールを掲げることが必要であると考えており、**ゴール設定に向けて他社の事例等に関する勉強会を社内で開催**している。

【実践例 3】

当社は事業ポートフォリオが多様な企業であり、事業部ごとに AI 技術への関わり方が異なる。また、それぞれが独立しているカンパニー制を採用していることから、単一の AI ガバナンス・ゴールに合意することは容易ではない。そのため、現時点では「人間中心の AI 社会原則」を尊重することとし、それと並行して AI に関する全社的な研修の一部に AI 倫理や品質を追加することで AI 倫理や品質に対する理解を高めていくことを狙っている。さらに AI 相談窓口を社内に設置して、事業部からの事例集めを行っている。対外的には動きが遅く見えるかもしれないが、**AI ガバナンス・ゴールの合意に向けたプロセスに価値がある**と考えている。なお、企業全体の AI ガバナンス・ゴールを設定する手前の段階で、AI システムを開発・運用する**事業部ごとの AI ガバナンス・ゴールの必要性や内容を検討することもありうる**と考えている。

【実践例 4】

当社は、AI システムの開発・運用にも AI システムを運用する企業の支援にも豊富な経験を有し、潜在的な負のインパクトが軽微ではないと見られている用途向けの AI システムも開発・運用している。これまで自社で運用した AI システムや他社に提供した AI システムから重大なインシデントが発生したことはないが、当社が提供する AI システムの用途の中には**社会的な受容が定まっていないものも多いと理解**している。そこで当社は、**消費者を含むステークホルダーとのコミュニケーションの強化を図るために AI ガバナンス・ゴールを設定し、公表している**。ステークホルダーが当社のポリシーを理解しているため、**AI システムを開発する担当者と顧客を含むステークホルダーとが AI 技術に対する基本姿勢を共有でき、コミュニケーションが円滑になった**と評価されている。

3．システムデザイン（AI マネジメントシステムの構築）

(1) AI ガバナンス・ゴールからの乖離の評価と乖離への対応を必須プロセスとする

行動目標３－１：AI システムを開発･運用する企業は、経営層のリーダーシップの下、自社が開発･運用している AI システムの AI ガバナンス･ゴールからの乖離を特定し、乖離により生じる影響を評価した上、負のインパクトが認められる場合、その大きさ、範囲、発生頻度等を考慮して、その受容の合理性の有無を判定し、受容に合理性が認められない場合に AI の開発・運用の在り方について再考を促すプロセスを、AI システムの設計段階、開発段階、利用開始前、利用開始後などの適切な段階に組み込むべきである。運営層はこのプロセスの具体化を行うべきである。そして、AI ガバナンス・ゴールとの乖離評価には AI システムの開発や運用に直接関わっていない者が加わるようにすべきである。なお、乖離があることのみを理由として AI の開発・提供を不可とする対応は適当ではない。そのため、乖離評価は負のインパクトを評価するためのステップであって、改善のためのきっかけにすぎない。

＊この行動目標の実施にあたっては、必要に応じて、以下の実践例だけではなく、別添２（AI ガバナンス・ゴールとの乖離を評価するための実務的な対応例）を参照されたい。

【実践例１】

当社は小規模企業であり、技術担当役員と開発担当者の距離が近く、プロジェクト数がそれほど多くないこともあり、**技術担当役員は全てのプロジェクトを十分に把握**できている。**技術担当役員は、「人間中心の AI 社会原則」からの乖離を評価するための観点を設定し、開発担当者に対し、全ての AI システム開発プロジェクトについて、実務上可能な限り早い段階に、観点ごとに乖離を特定し、乖離により生じる影響を評価し、技術担当役員に報告するように指示**している。そして、**技術担当役員は、開発担当者以外の者も加えた開発担当者との会議において、開発担当者の報告内容に基づき、乖離により生じる影響を改めて評価し、負のインパクトがある場合には、その受容の合理性の有無を判定し、受容に合理性が認められない場合に AI の提供の在り方について再考する**こととしている。

このプロセスの運用にあたっては、行動目標３－１－１にしたがって、**当社が属する業界における標準的な乖離評価や本ガイドラインの別添 2 を参考にしながら、社内運用の標準化に努めている**。

【実践例２】

多数の事業部を有する当社は、**AI ガバナンス担当役員を決め、この役員の下に AI 倫理審査委員会を設置**している。この委員会は、**特定の AI システムの開発・運用プロジェクトを担当する者以外から構成**されており、「人間中心の AI 社会原則」を踏まえて当社が策定した **AI ポリシーからの乖離評価をプロジェクトごとに実施**することを任務としている。具体的には、**AI ポリシーに基づいた評価リストを作成**し、AI システムの開発・運用について**当該評価リストを用いて乖離を特定し、乖離により生じる影響を評価し、負のインパクトがある場合、その受容の合理性の有無を判定し、受容に合理性が認められない場合に AI の開発・提供の在り方を再考するよう、プロジェクト担当者に促すこととしている**。**乖離評価のためのリストについては、行動目標３－１－１にしたがって、当社が属する業界における標準的な乖離評価や本ガイドラインの別添 2 を参考にしながら作成している**が、**実際のプロジェクトを選定し、AI マネジメント担当者がプロジェクト担当者に伴走することで、リストの精緻化や運用の定着化を図る工夫**もしている。なお、AI 倫理審査委員会では、プロジェクト担当者に対し、その再考の結果を報告するよう求めることとし、その報告内容の合理性に懸念がある場合には、AI ガバナンス担当役員からプロジェクトを所管する役員にその旨を通知し、適宜調整を図ることとしている。

なお、AI システムに伴う負のインパクトは、用途、範囲、使用態様によって大きく異なり、プロジェクトを推進している担当者がその性質や程度を最もよく知っているとも考えられることから、**潜在的な負のインパクトが軽微であることが明らかな場合に AI マネジメント担当者がプロジェクト会議に同席して簡素な乖離評価とするなど、厳格な乖離評価を一律に求めない運用も考えられる**。しかし、現時点では乖離やリスクを評価するためのノウハウがまだ当社内に十分蓄積されていないこともあり、AI 倫理審査委員会による一律の乖離評価を全てのプロジェクトが通過すべき必須のゲートとし、今後の経過を見ることとしている。

【実践例３】

乖離評価のプロセスは、複数社によって担われるべき場合がある。例えば、サービスを他者に提供する AI システム運用企業が、AI システムの開発を自ら行うのではなく、AI システム開発者にその開発を委託する場合、AI システム開発者と AI システム運用者の両者が乖離評価プロセスを分担することが合理的である場合がある。そしてこの場合、**AI システムの開発から運用に至るまでに想定される流れはもちろん、乖離評価の方法や基準を開発者と運用者の間で共有することが重要**である。AI システム運用者が AI システムを用いたサービ

スの提供に伴うリスクを軽視する場合には、AI システム開発者は難しい立場に置かれることになるため、このような対応は大切である。

このような委託を受けて AI システムの開発を行うことがある当社では、当社の責に帰すべき事情がある場合を除いて、AI システムの運用上の事故はサービスを提供する運用者が負うこととなる契約を結んでいるが、それでも、この種の事故が発生したときに当社も紛争に巻き込まれるリスクはある。そのため、納入した AI システムの運用方法にも無関心ではいられない。実際、プロジェクトの終盤で運用上のリスクに気がつき、当該プロジェクトの再設計を運用者に助言した上で、その再設計のコストの一部を負担せざるを得なかった経験がある。そのため、**当社が属する業界における標準的な乖離評価や本ガイドラインの別添 2 を参考にしながら、個々の評価項目の意味を十分に理解した上で、乖離評価プロセスを確立し、自社で開発せず他者にサービスを提供するのみである AI システム運用者にも共有**するようにした。懸念項目を網羅している乖離評価プロセスを活用し、しかも早めに乖離評価を行うことで、顧客との交渉はスムーズになってきている。

＊次の実践例のように、通常の乖離評価プロセスに加えて、広く議論を行うことが必要な場合がある。

【実践例 4】

当社は AI システムの開発を主たる事業とする小規模企業である。技術担当役員が全てのプロジェクトについて進捗報告を受けることになっており、その報告の中には、公平性などの AI 倫理に関する事項も含まれている。**AI 倫理の問題の中には、妥当な出力結果が得られるように十分なデータセットを用意するなど技術的な配慮で対応できる事項もあるが、社会的にセンシティブな領域ではそれだけでは不十分な場合がある**。

そこで当社では、そのような**センシティブな領域における AI システムのプロジェクトの場合には、法務担当役員などを含めて話し合いをする**ことにしている。センシティブな領域の特定には、すでに広範囲に AI システムの開発・運用しているリーディング企業の考え方などを参考にしている。このような情報収集には実務的な雑誌が有効である[13]。そのような雑誌には概要記事が掲載されることが多く、その概要記事を手がかりにインターネット等で深い情報に当たることが効率的かつ効果的である。

個々のプロジェクトに関して外部の有識者や専門家を招いて意見交換している企業もあることは知っている。当社も事業の拡大に合わせて、そのような意見交換の場も設置していき

[13] センシティブな領域への対応の参考例に、舟山聡『AI 倫理に対する企業の取組み(1)』NBL No. 1170（2020 年 5 月 15 日）などがある。

たいと考えている。

【実践例５】

当社は AI システムを開発している部門と運用している部門が混在する大規模企業である。すでに AI ポリシーを定め、当該ポリシーからの乖離評価を全てのプロジェクトに対して実施している。過去に対応したことのある分野におけるプロジェクトであれば、プロジェクトの早い段階で AI マネジメント担当者が対応すれば十分であるが、**これまでに対応したことがないセンシティブな分野において AI システムを開発したり利用したりする場合には、通常のプロセスではなく個別に相談**してもらうようにしている。そして、**そのような相談を受けたときには、開発部門、運用部門、法務部門等の責任者からなる横断的な会議を開催**し、議論することとしている。AI マネジメント担当者が通常の乖離評価プロセスにおいてそのようなプロジェクトを発見した場合も同様である。

当社では、**定期的に外部の有識者や専門家を招いて、最近の AI インシデントやセンシティブ分野に関する情報を早い段階でキャッチできるようにしている**。そのため、今のところは、有識者や専門家から入手した情報や一般的な助言を踏まえて、横断的な会議で議論すれば十分に対応可能である。他方で、**当社の AI システムの用途先が広がってきていることから、今後は個別のプロジェクトに関しても外部の有識者等に意見を求める必要が出てくるのではないかと考えている**。

① 業界の標準的な乖離評価プロセスとの整合性を確保する

> 行動目標３－１－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、業界における標準的な乖離評価プロセスの有無を確認し、そのようなプロセスが存在する場合には、それを自社のプロセスに取り込むべきである。

【実践例１】

AI 原則の実践では多様な視点が欠かせない上に、**他社との認識共有も必要**であるから、**自社だけで考えるのでなく他社や団体等の取り組みを参考にすべき**ある。このように考えている当社では、AI マネジメント担当者に対して、乖離評価プロセスを構築するにあたって社外の取り組みを調査するように指示した。

当社は産業用途の AI システムの開発を主たる事業としていることから、産業用途を中心

に調査を行った。調査を進めていくと、たとえば、経済産業省、厚生労働省、消防庁が、**「プラント保安分野 AI 信頼性評価ガイドライン」**、それを実施するための**「実施内容記録フォーマット」**、記載例をまとめた**「信頼性評価実用例」**を公表していることがわかった。また、AI プロダクト品質保証コンソーシアムが公表している**「AI プロダクト品質保証ガイドライン」**には、Voice User Interface、産業用プロセス、自動運転、OCR の事例が掲載されていることがわかった。さらに、国立研究開発法人産業技術総合研究所が**「機械学習品質マネジメントガイドライン」**を公表しており、産業用途別の実アプリケーションを対象とする具体的適用事例としての**リファレンスガイドの作成を予定**していることもわかった。**当社の現在の乖離評価プロセスには、これらの具体的な取り組みの一部が反映されている**。

【実践例２】

当社では、AI システム利用者から得られたデータに基づいた AI システムを開発・運用している。AI 原則の実践、特に**プライバシー確保の原則の実践にあたっては、**AI モデルの構築とアウトプットへの配慮だけではなく、**AI モデルに対するインプットデータの扱いへの配慮が必要である**と認識している。当社には個人情報の扱いに関する豊富な経験があるが、そうであっても社外の取り組みに積極的に目を向けるべきであると考えている。そこで、AI マネジメント担当者に対して、乖離評価プロセスを構築するにあたって社外の取り組みを調査するように指示した。

AI モデルの構築とアウトプットへの配慮については、たとえば、パーソナルデータ＋α研究会が提示している**プロファイリングにおける「自主的取組みに関するチェックリスト」**があることがわかった。また、AI モデルに対するインプットデータの扱いについては、改めて**個人情報保護委員会が公表しているガイドライン**を確認した。**「DX 時代における企業のプライバシーガバナンスガイドブック ver1.0」**については、インプットとアウトプットの両方の観点から AI に関する記述も見られるため参考になった。**当社の現在の乖離評価プロセスには、これらの具体的な取り組みの一部が反映されている**。

② 利用者に対して乖離の可能性や対応策に関する十分な情報を提供する

行動目標３－１－２：AI システム利用者にサービスを提供している AI システム運用企業は、経営層のリーダーシップの下、提供している AI システムに一定の乖離が発生しうる場合には、AI システム利用者に対して、その事実や当該乖離への対応策に関する十分な情報を提供するとともに、問い合わせ先も明確にすべきである。

【実践例１】

当社は AI システムを運用し、不特定多数の消費者を中心とする AI システム利用者に対してサービスを提供している。**サービス提供相手の AI に関するリテラシーに大きな幅があ**ることが予想されることから、当社では、**AI システムの運用にあたり、**適切なリスク管理を行い、負のインパクトを最小限にするための措置をとっていることや、情報の厳格な安全管理を行っていることなど、**リスクに関連する情報を、不慣れな消費者でも理解できるようにわかりやすく整理して提供するとともに、問い合わせ先を明確にしている**。これらの情報に加え、上述のとおり、サービス提供相手の AI に関するリテラシーに大きな幅があることが予想されることから、当社では、**提供される情報等に AI システムの出力が用いられていることが利用者に明らかな場合を除き、AI を使っていることをわかりやすく表示するとともに AI を利用したときのメリットとデメリットを明示**している。そして、**AI 機能を好まない AI システム利用者には代替サービスがあることも表示**している。個人情報を扱う場合もあることから、個人情報保護委員会のガイドラインに準拠することはもちろんのこと、「DX 時代における企業のプライバシーガバナンスガイドブック ver1.0」を参考にしつつ、**消費者との継続的なコミュニケーションを確立**している。

【実践例２】

当社は、実践例１と同様、AI システムを運用し、外部にサービスを提供しているが、**ビジネスで利用する企業に提供**している点で実践例１と異なる。**当社のサービスの提供先は AI リテラシーが比較的高い**ため、提供している AI システムには一定の乖離が発生しうる可能性や当該乖離への対応策について、**専門的な用語も交えながらメリハリをつけた説明**をするとともに、問い合わせ先を明確にしている。

今後、一般消費者向けに AI システムを用いたサービスを提供する可能性があるが、**サービス提供先の AI へのリテラシーに応じて十分な情報を提供**していきたいと考えている。

【実践例３】

当社は、実践例１と同様の対応をしているが、**AI システム利用者からの問い合わせに対応するために必要な情報を AI システム開発者から提供してもらえるように、その旨を契約で明確にしている**。AI システム利用者からの「フィードバック」は AI システム開発者にとっても貴重な情報であることもあり、迅速に対応してもらっている。

【実践例４】

当社は、実践例１と同様の対応をしているが、**AI システム利用者が自らの判断で AI システムを用いたサービスを選択できるようにすること自体に付加価値があると信じており、他社との差別化のために情報提供の在り方を工夫**している。また、**AI システムだけではなく情報提供の在り方についてもフィードバックをもらう**ように工夫している。

③ データ事業者は乖離評価に十分な情報を AI システム開発者に提供する

行動目標３－１－３：データを提供する企業は、AI システムを開発する企業が適切に乖離評価をできるようにするために、データの収集元、収集方針、収集基準、アノテーション付与基準、利用制約等のデータセットに関する情報を提供すべきであり、AI システム開発者は十分な情報を提供するデータ事業者からデータセットを取得すべきである。

【実践例１】

当社は、AI システムを開発する企業にデータを提供しているデータ事業者であり、**AI システムを開発する企業が適切に乖離評価をできるようにするために、**データの収集元、収集方針、収集基準、アノテーション付与基準、利用制約等の**データセットに関する情報を提供している**。また、**十分に整理されていないデータセットを提供する場合であっても、乖離評価に必要なデータの収集元等の基本的な情報を十分に提供**している。

コラム：公平性（Fairness）確保のための取り組み

公平性（fairness）の基準を決めたり、公平性の原則を実践したりすることは難しい。そこで OECD では原則から実践を支援する活動をしている。The OECD

Network of Experts Working Group on Implementing Trustworthy AI は、公平性などの原則の実践するための取り組みを収集し、それらを技術的、手続的、教育的の３つにわけている。このうち公平性を確保するための技術的なツールとして AT&T、Microsoft、LinkedIn、Google、IBM の取り組みが紹介されているが、日本企業の取り組みは紹介されていない。OECD Framework of Tools for Trustworthy AI で紹介されているサンプルツールは LinkedIn の Fairness Toolkit (LiFT)である。

TYPES OF TOOLS THAT EMERGED FROM THE SURVEY

Approach	Type of tool
Technical	Toolkits / toolboxes / software tools
	Technical documentation
	Technical certification
	Technical standards
	Product development / lifecycle tools
	Technical validation tools
Procedural	Guidelines
	Governance frameworks
	Product development / lifecycle tools
	Risk management tools
	Sector-specific codes of conduct
	Collective agreements
	Certification
	Process-related documentation
	Process standards
Educational	Change management processes
	Capacity / awareness building
	Inclusive design guidance
	Educational materials / training programmes

ツールの分類図、OECD.AI の AI Wonk から引用 [14]

このような国際的な議論を通じて、公平性の基準や実践の相場感がデファクトで設定されていく可能性がある。文化やビジネス慣習などの違いによって公平性の基準が異なる可能性があることから、日本の多様性も反映されるように、日本企業の取り組みも発信していく必要がある。実際、日本、米国、英国の 3 地域に対応したローン審査 AI について、地域ごとの公平性により判断が異なることが明らかになっている [15]。

[14] Carolyn Nguyen, Adam Murray, and Barry O'Brien, "What are the tools for implementing trustworthy AI? A comparative framework and database," The AI Wonk, OECD.AI (May 25, 2021), https://oecd.ai/wonk/tools-for-trustworthy-ai.

[15] 富士通『文化やビジネス慣習によって異なる公平性を設計段階から考慮する AI 開発手法 Fairness by Design を開発』(2021 年 3 月 31 日)、https://pr.fujitsu.com/jp/news/2021/03/31-1.html.

(2) AI マネジメントシステムを担う人材のリテラシーを向上させる

行動目標３－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、AI マネジメントシステムを適切に運営するために、外部の教材の活用を検討し、AI リテラシーを戦略的に向上させるべきである。たとえば、AI システムの開発・運用における法的・倫理的側面に責任を負う役員、マネジメントチーム、担当者には AI 倫理に関する一般的なリテラシー向上のための教育を、AI システムの開発・運用プロジェクトの担当者には AI 倫理だけではなく AI 技術に関する研修を提供することが考えられる。データを提供する企業は、AI システム開発者・運用者の実践例を参考に、データ提供に関わる担当者の AI 倫理に関する一般的なリテラシーを向上させるべきである。

【実践例１】

当社は小規模企業であり、**研修対象者が少ないことから、AI リテラシーの向上の研修プログラムを自前で用意せず、外部の教材を用いることとした**。米国の教育技術の営利団体である Coursera や日本ディープラーニング協会（JDLA）などが提供しているオンライン講座やテキスト、経済産業省が紹介している巣ごもり DX ステップ講座など、国内外を含めて様々な教育プログラムが利用可能である。たとえば、理系のバックグランドのない人向けのオンライン教材に、スタンフォード大学 Andrew Ng 教授による AI for Everyone（Coursera、約 4 時間、トランスクリプト付き）がある。「Coursera（コーセラ）をはじめ、インターネット上の学習サイトで 40 時間ほど勉強すれば、専門家に近いレベルに到達できる」[16]との指摘があり、**外部の教材でも十分**であると判断した。

当社は、研修対象者の到達度を計るための JDLA の検定試験シラバスに基づいたプログラムを活用している。JDLA の G 検定は、AI 技術の基礎から AI 倫理まで幅広く含む内容である。また、JDLA 主催の「G 検定合格体験談オンラインセミナー」（2020 年 5 月 30 日開催）では、AI ベンチャーで営業を 2 年経験した人が、合格までに 25 から 30 時間の学習が必要だったと振り返っており、研修対象者に過度な負担にならないことも確認している。

これまで実施してきて、当社が期待している効果が出ていると思っている。たとえば、AI システムのインシデントについてニュースで断片的に聞いたことがあった程度の人が、AI 技術の初歩から倫理的な側面まで習得したことで、**AI の負のインパクトについても当事者意識を持って考えてくれるようになった**。

[16] 松尾豊『研究の第一人者が語る「AI との向き合い方」』週刊東洋経済（2020 年 5 月 16 日）、第 56 頁。

【実践例２】

当社は、AI システムの開発・運用を事業の柱の１つとする大企業である。AI 技術や倫理に関する教材が外部にあることは知っているが、**AI システムの提供数が多く、社会へのインパクトが大きいことから、**汎用的な外部教材ではなく、**自社の AI システムの用途を想定した事例を充実させた自社教材を使っている**。

AI に関する研修プログラムを作成した当初は、AI 技術に関する講義の最後に AI 倫理のパートを設けていたが、外部有識者を招聘した委員会からの指摘をきっかけに、**AI 倫理に対する経営層の関心が一層高まり、AI 倫理だけを独立させた e-learning を作成し、全社員に受講してもらっている**。この e-learning には講義と確認テストが含まれており、AI 倫理に詳しくない人でも１時間程度で終えることができるように工夫されている。**自社の AI システムの用途に関連づけることで短時間でも高い学習効果が得られている**と考えている。

コラム：JDLA の検定試験の活用事例

JDLA の G 検定は多くの企業に利用されている[17]。全社的な DX 推進プロジェクトの人材育成カリキュラムに G 検定を組み込み、数百名規模で団体受験している企業もある。たとえば ENEOS は、全社員のデジタルリテラシーを AI Analytics、Business Intelligence、Cyber Security、Design Thinking の４つの項目に分け、G 検定を AI Analytics に組み込んでいる[18]。別の企業では、IT 部門の自主活動からグループ/部署横断で 100 名規模が参加する G 検定対策勉強会へと発展しているという。公的機関が提供する研修プログラムでも G 検定が使用されている[19]。

G 検定の利用者からは「これからのプロジェクトマネージャーやソリューションアーキテクトが知っておかなければいけない知識」が得られると評価されており、受験希望者が若手エンジニアから、役員、管理職、リーダー層にも広がったという声もある。G 検定は AI 技術の基礎や AI 倫理を学ぶきっかけとなっているようだ。

[17] JDLA、団体受験企業様の声。https://www.jdla.org/certificate/general/#general_No04.
[18] 以下のスライド番号 22 では AI Analytics の下に E 検定、G 検定が位置づけられている。https://www.hd.eneos.co.jp/csr/meeting/pdf/esg_ex_20201202.pdf.
[19] 埼玉県産業振興公社が中小企業向けの『AI・IoT 人材育成研修（技術者養成コース）』で G 検定を取り入れている。https://www.saitama-j.or.jp/iot/jinzai/.

(3) 適切な情報共有等の事業者間・部門間の協力により AI マネジメントを強化する

行動目標３－３：AI システムを開発・運用する企業、及び、データを提供する企業は、学習等用のデータセットの準備から AI システムの開発・運用までの全てを自部門で行う場合を除き、経営層のリーダーシップの下、営業秘密等に留意しつつ、自社や自部門のみでは十分に実施できない AI システムの運用上の課題と当該課題の解決に必要な情報を明確にし、積極的に共有すべきである。その際に、必要な情報交換が円滑に行われるよう、AI システム開発者、AI システム運用者、データ事業者の間で予め情報の開示範囲について合意し、秘密保持契約の締結等を検討することが望ましい。

【実践例１】

当社は、開発した AI システムを顧客に納入し、当該顧客が AI システムの運用にあたっている。この AI システムは運用環境の変化によって精度が低下し、場合によっては設備の破損等の損害につながるおそれがある。そのため、顧客に対しては、**AI システムの出力のモニタリングを依頼し、品質劣化の判断の仕方も伝えている**。

AI に詳しくない顧客に対して、モニタリング等を単に依頼するだけでは機能しない。**AI システムのメンテナンスが必要な理由とその原因（学習データと運用時の入力データの分布が変化する等）、当該原因による出力の変化の傾向などについて、時間をかけて説明して納得してもらう必要がある**。標準的な情報を提供すれば十分な場合もあるが、AI システム開発側がそのように考えた場合でも、納入先に積極的に質問を促し、可能な限り認識を一致させるべきである。必要に応じて、保守サービス契約等を締結し、納入後であっても積極的に質問を受け付ける体制を整えることも重要である。また、**AI システムの再学習を行った場合には、再学習によって出力がどのように変化したかを丁寧に説明すべき**である。

当社は、このような情報共有が円滑に行われるように、AI システム開発者と AI システム運用者の間で予め情報の開示範囲について合意しておき、秘密保持契約の締結も締結している。

【実践例２】

当社が開発している AI システムは、特定のデータセットによって学習させたものであり、データセットに含まれていない対象に適用すると好ましくない出力結果となる可能性がある。そのため、**当該 AI システムを AI システム利用者に提供しようとする AI システム運用者**

に対して、学習等に利用したデータ、利用したモデルの概要や精度などの性能を説明するだけでなく、AI システムを利用すべきではない状況や対象についても伝えている。情報提供を徹底するために、紙書面や電子書面で伝えるだけでなく、別途時間を確保して口頭でも説明し、そのような説明を行ったことにサインしてもらうようにしている。

① 複数事業者間の情報共有の現状を理解する

> 行動目標３－３－１：AI システムを開発・運用する企業、及び、データを提供する企業は、経営層のリーダーシップの下、学習等用のデータセットの準備から AI システムの開発・運用までの全てを自社で行う場合を除き、営業秘密に留意しつつ、複数事業者間の情報共有の現状を理解し、適時に理解を更新すべきである。

【実践例１】

AI システムの開発は、それがどのような場面で利用されるものであるかを踏まえて行う必要があり、また、AI システムの運用は、それがどのような制約の下で開発されたものであるのかを正しく理解した上で行われる必要がある。そのため、データの提供、データへのアノテーションの付加、AI システムの開発、AI システムの運用が複数事業者によって担われる場合、複数事業者間での情報共有が重要になる。**AI 技術の社会実装を促進するためには、共有情報の標準化が望ましい**。このような問題意識から、当社では、**自社の情報提供の在り方を決めるにあたって、経営層のリーダーシップの下、営業秘密に留意しつつ、複数事業者間の情報共有の現状を理解し、定期的に理解を更新する**こととした。

情報収集を進めていくと、**複数事業者間の情報共有の標準化に向けた様々な取り組みがなされている**こともわかった。たとえば、国立研究開発法人産業技術総合研究所は、機械学習利用システムの品質に関する社会合意としての基準とすることを目的の１つに掲げ、「機械学習品質マネジメントガイドライン」を公表しており、経済産業省、厚生労働省、消防庁が、このガイドラインを基礎として、プラント保安分野の信頼性評価実施記録フォーマットを作成していることもわかった。また、食品の成分表示等が人々の責任ある意思決定に貢献しているように、AI モデルの性能も表示していくべきであるとの認識の下、モデルカードの提案がなされていることも把握した[20]。

現時点で学習済みの機械学習モデル等の性能や品質を複数事業者間で共有するための標

[20] Google, “The value of a shared understanding of AI models, https://modelcards.withgoogle.com/about.

準的な文書化手続きはないが、**社内の体制を整備するにあたっては、自社の独自基準を一から考えるのではなく、様々な取り組みを参考にする**つもりである。

【実践例２】

当社は、**AI 倫理や品質に関する団体に所属し、AI システムの性能等に関する情報提供のベストプラクティスについて他の所属企業と積極的に意見交換している**。AI システム利用者には AI システムに関する十分な情報を提供すべきであるが、利用者が消費者であってもそれ以外の利用者であっても、全ての利用者が AI の性質や限界などに詳しいわけではないことから、専門家以外には理解が難しいような情報や、膨大かつ詳細な情報を一方的に提供しておけばよいと考えることは適当ではない。情報提供の適切な在り方を考えるためには、自社の直接的な経験だけではなく、他社との意見交換を通じて間接的に多くの利用者と触れていくことも大切である。

AI システム開発者から AI システム運用者に伝えるべきと思われる情報には、たとえば、AI システムの開発に用いたデータに関する情報がある。たとえば、データの取得源（オープンデータということもある）、データの量や分布、これに含まれるカテゴリー毎の概要などを挙げることができる。また、開発の際に選択した（選択しなかった）アルゴリズムや、生成されたモデルの概要、特に、どのような条件下でテストを行い、その結果、どの程度の精度が得られたかなどを説明することも重要である。

これらの観点は AI システムの開発や運用の経験が豊富な企業にとっては目新しいものではないが、当社は**「伝え方」が重要**であると考えている。**どのような内容をどの程度の深さで説明するか**である。**複数事業者間の情報共有の現状を理解することは、AI ガバナンスの全体設計を考える上で重要**であり、そこに AI 倫理や品質に関する団体に参加する意義がある。

② 環境・リスク分析のために日常的な情報収集や意見交換を奨励する

> 行動目標３－３－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、日常的に、AI システムの開発や運用に関するルール整備、ベストプラクティス、インシデントなどの情報を収集するとともに、社内外の意見交換を奨励すべきである。

【実践例１】

アジャイル・ガバナンスを支えるのは日常的な情報収集であり、そのための意見交換である。AI システムの適切な開発・運用のためのガバナンスには、様々なステークホルダーの

関与が必要であると言われているように、日常的な情報収集や意見交換でも様々なステークホルダーの声を聞くことが求められる。たとえば、**AI マネジメントチームを社内に設置している場合であっても、社内の他部門との勉強会を開催したり、他社も参加する団体活動に関与したりすることが必要**である。

これまで当社は、部門外や社外と情報収集や意見交換を行う場合には、その担当者に対して重要な情報を取得できる蓋然性を高いことを説明させてきた。しかし**AI 倫理に関しては、**原則こそ定まりつつあるが、**原則の尊重の在り方については正解のない中で模索していくしかないこと、さらには他社も同様に活動していることから、AI マネジメント担当者に対して AI の適切な開発・運用に関する情報収集や意見交換を奨励し、部門を越えた社内の勉強会で共有するように指示している**。

このような活動を継続することで、決定版のような解決策はないものの大きなトレンドをつかめるようになってきた。このような活動の成果を、適時に実施される環境・リスクの分析に反映している。

【実践例２】

当社は AI システムを開発する小規模企業である。社内には AI 倫理の尊重よりも成長を重視すべきという意見があるため、**法務部門と技術部門で AI 倫理に関する社内勉強会から始めることとした**。言葉の定義や使い方が部門ごとに異なる可能性があるためファシリテーター役を設置したところ、円滑に議論が進み、成長を重視すべきと話していたエンジニアもすでに公平性などを扱う論文に接していて、AI 倫理に対する認識に大きな違いがないことがわかってきた。**エンジニアは AI 倫理の尊重を技術によって実現することに関心を示し始めてから、開発プロセスが AI 倫理に整合的なものに変化**しつつある。**今後は社外との意見交換も進めたい**。

コラム：様々なステークホルダーによる共創環境整備

行動目標３－３はAIシステムの負のインパクトへの対応としてのAIマネジメントの強化という視点でまとめられているが、別の視点からみれば、複数事業者間の情報共有などを共創環境基盤の整備と位置づけることもできる。これまでのソフトウェアやシステムにおいても、安全性等の確保には複数事業者の協力が欠かせなかったが、AIシステムではより一層の協力が求められる。

このような状況に対応するように、共創領域を厚くするための様々な取り組みがなされている。契約や法的責任の在り方について、JDLAは「契約締結におけるAI品質保証の在り方」研究会を開催している。有志の集まりであるAI法研究会では、データ、プライバシー、知的財産などの部会にわかれて法的・倫理的問題について意見交換している。複数事業者間の共有情報の標準化については、Partnership on AIがABOUT MLというプロジェクトを実施しており、Google、Microsoft、IBMのサンプルが提供されている。

Socieity5.0で求められるアジャイル・ガバナンスでは、様々なステークホルダーが積極的にガバナンス設計に携わることが期待されている。そして、このような活動が有機的に結合していくことで、AIシステムの開発・運用に必要な共創環境が整備されることが期待される。本ガイドラインには有機的な結合への貢献が期待されている。

(4) インシデントの予防と早期対応により利用者のインシデント関連の負担を軽減する

行動目標３－４：AI システムを開発・運用する企業、及び、データを提供する企業は、経営層のリーダーシップの下、インシデントの予防と早期対応を通じて利用者のインシデント関連の負担を軽減すべきである。

【実践例１】

当社では、AI システムへの信頼性を高めるためには、問題が生じないように工夫し、問題が発生したときに迅速に対応することで利用者の負担をできる限り軽減することが重要であると考えている。**AI システムの開発・運用には、**データ事業者、AI システム開発者、AI システム運用者、ビジネスで AI システムを利用する者、消費者など**様々な立場の企業や個人が関与することが多く**、さらに **AI にはいわゆるブラックボックス性があることから、責任の所在が曖昧になりやすい**。インシデントを予防するためには、**負のインパクトを軽減できる者に責任を分配することが重要**である。また、**インシデントの発生に備えて事前に準備することでインシデントへの早期対応力を高めることも大切**である。

【実践例２】

当社は、実践例１の実施を基本としつつ、一部の用途では保険の利用を検討している。**社会全体への恩恵が大きいにもかかわらず、AI システムの動作に関して一定の不確実性が避けられず、まれに一定の経済的な損失が発生してしまう用途では、保険を活用することでインシデントから生じうる経済的な損失に早期に対応することで利用者の負担を軽減する**ことが重要であると考えている。もちろん、利用者の信頼を継続的に高めていくためには AI システムの不確実性を低減していくことが重要であると認識しており、そのための研究開発を継続している。

① 複数事業者間の不確実性への対応負担を適切に分配する

行動目標３－４－１：AI システムを開発・運用する企業、及び、データを提供する企業は、経営層のリーダーシップの下、学習等用のデータセットの準備からシステムの開発・運用までの全てを自社で行う場合を除き、負のインパクトを全体で最小化できるように AI システムの不確実性への対応負担を複数事業者間で分配すべきである。

【実践例１】

機械学習の手法により学習用データを基礎として帰納的に構築される AI システムの利用には、AI システムが導出した推論の結果が常に正しいものであるとは限らないという不確実性の問題が伴う。このような不確実性への対応策としては、適切なデータセットの準備、適切なモデルの選択、AI システム利用開始前の検証や試験の実施などの AI システムの開発時の対応によって不確実性の低減を目指すアプローチがあるが、AI システムの運用のモニタリングなどの AI システム運用時の対応によって不確実性の制御を目指すアプローチも重要である。**AI システムの開発者と運用者が異なる場合には、その提供に伴う不確実性に対し、各事業者が現実的に採りうる選択肢を踏まえ、こうした不確実性への対応負担が事業者間で契約等によって適切に分配されることを原則とすべきである**。『AI・データの利用に関する契約ガイドライン』でもこの原則を確認している[21]。

AI システムを開発する当社は、このような原則を踏まえ、**AI システム運用者に適切に使ってもらうことが AI 技術への社会的な信頼向上に資する**と考えている。情報を集めていくと、AI システム運用者の中には、AI システムは従来型のソフトウェアの延長上にあると考え、AI システム開発者が AI システムの品質に関する全ての責任を負うべきと考えている企業もあることがわかった。他方で、**AI システム運用者が AI システムへの期待を理解し、運用者自身が腹落ちするまで丁寧に説明することで、運用者自身が再学習のタイミングを判断できる場合がある**ことをわかった。そして、「AI システム品質保証ガイドライン」に記載されているように、「品質保証の技術者やチーム、組織は、開発や営業と共に、AI システムに関する顧客の理解を深めるような活動を行う」ことが大切であるという考え方が少しずつ広がっている現状を理解した。ただ、**依然として AI システム開発者が品質を保証すべきという考え方が根強い**ことから、「AI システム品質保証ガイドライン」のような活動が与える好影響が広がることに期待しつつ、不確実性への対応負担に関する調査を今後も定期的に行いたいと考えている。

【実践例２】

当社は、他社が開発した AI システムを用いてサービスを提供している AI システム運用者である。AI システム開発者とは、「AI・データの利用に関する契約ガイドライン」のモデ

[21] 経済産業省『AI・データの利用に関する契約ガイドライン（AI 編）』（2018 年 6 月）、第 106 頁（「ユーザの課題解決は、ユーザの事業や社内の既存ルール・制約や組織と深く関連し、ユーザの意思決定の下に行われることや、ベンダのコントロール下にない未知の入力（データ）に対する学習済みモデル等の挙動について、ベンダが性能保証をすることが困難である」）。

ル契約書を参考にした契約書を用いて契約を締結している。これによれば、AI システム（学習済みモデル）の開発者は、仕事の完成や成果の性能・品質等の保証は行わない一方、一定以上の注意水準をもって業務を行わなければならないことになっている。当社はあくまで他社が開発した AI システムを運用しているにすぎないという意識があり、AI システムの運用に関連して不適切な事例が発生したり、それ以外の場面で最終利用者から説明を求められたりした場合に、運用者である当社がどのような説明責任を果たすべきであるかを真剣に考えていなかった。

しかし、**最終的な法的責任の所在はともかく、AI システム利用者に対して直接サービスを提供しているのは当社である以上、当社が運用している AI システムについて利用者から説明を求められた場合に、少なくとも一次的にこうした要求に対応する責任の一切を免れることはできないことと、十分な説明ができない場合に当社にレピュテーションリスクが生じる**ことに気がついてからは、開発者の協力を得ながら、**リスク低減のために AI システム運用者ができることを行い、必要に応じてそのことを説明していくという方針**に変えた。

【実践例３】

当社では、当社が保有するデータを利用した AI システムの開発を他社に委託することを計画しているが、当社にはデータの扱いに関するノウハウが乏しく、クレンジングなどのデータの前処理だけでなく、データの品質の確保を含めた一切合切を他社に任せたいと考えていた。当社は、現在保有するデータを集めて AI システムの開発者に提供してしまえば、データを扱うプロである AI システム開発者がデータに必要な処理を行ってくれ、当社が希望する AI システムを開発してくれるものだと誤って認識していた。

しかし、開発委託前に情報を収集していくと、一般の事業者間におけるデータの提供においても参考となる内容としてまとめられている「AI・データサイエンス人材育成に向けたデータ提供に関する実務ガイドブック」[22]があることがわかり、そこには「**提供前の委託データの品質をコントロールできるのは委託者のみ**」であるという考え方や、一定の前提の下では「成果の利用による利益も委託者のみに帰属することから・・・**危険責任と報償責任の考え方に基づいて**・・・**創出された成果の利用・実施等に伴う損害の責任は、原則として委託者が負う**」べきであると考えられる場合があることなど、データ事業者の留意事項がまとめられていた。

現在は、AI システムの開発に必要なデータの内容は、当社が開発を予定している AI シス

[22] 経済産業省『AI・データサイエンス人材育成に向けたデータ提供に関する実務ガイドブック』（2021 年 3 月 1 日）

テムの内容によって定まってくること、AI システム開発者の側で対応できることには限界があることなどを理解している。当社はここで一度立ち止まり、**データ提供段階であってもAI システムの開発・運用のライフサイクルの重要な一部**であることを踏まえ、複数事業者間の不確実性への対応負担について再検討することが考えている。

② インシデント/紛争発生時の対応をあらかじめ検討しておく

> 行動目標３－４－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、AI インシデント又は紛争発生時に、AI システム利用者への説明、影響範囲や損害の特定、法的関係の整理、被害救済措置、被害拡大防止措置、再発防止策の検討等を速やかに実施するため、対応方針の決定や計画の策定を検討するとともに、当該対応方針又は計画に関して適宜実践的な予行演習の実施を検討すべきである。

【実践例１】

当社は AI システムを開発・運用する中小企業である。AI インシデントの発生可能性をできるだけ低くすることはもちろん重要であるが、インシデント発生の可能性をゼロにすることは困難であるため、**発生時の損害を最小限に抑えるための計画の策定・発動が重要であると認識**している。

具体的には、**インシデントが発生した場合に備えて、連絡受付窓口の設置、対応を担当する役員のアサイン、社内における連絡体制はもちろん、社外の関係者・専門家への連絡体制を整備**している。あらゆるインシデントに万全に対応することは困難だが、自社の AI システムの内容に鑑みて想定される主なインシデントについて、ある程度類型的に整理した上で、大まかな対応方針を策定している。また、策定した対応方針の実施可否の確認のため、定期的に予行演習も実施している。

【実践例２】

当社は AI システムを運用する大企業である。**インシデントの発生時の対応を速やかに行えるように、連絡受付窓口の設置、担当役員のアサイン、リスク管理部門、法務部門、広報部門、危機管理部門との連絡・連携体制はもちろん、社外の関係者、専門家への連絡体制も整備**している。

また、想定しうるインシデントを複数パターン想定し、どういった法的責任が発生しうる

かについてあらかじめ専門家に相談して整理し、その上でリスク評価を実施している。人身・物損事故、プライバシー侵害、財産的損害など、様々な類型の被害が生じうるので、類型ごとに開発者、運用者、利用者等の法的責任関係を予め整理しておくことは有用である。また、AI システムに固有の考慮要素として、異常な結果を出力する原因が多様であること（アルゴリズムの異常、学習データの真正性、学習データの偏り等）や、想定外の影響が生じやすいことも念頭に置いておく必要がある。想定外の事態が発生した場合であっても、システムリスクを低減させるような技術上・運用上の仕組みを定期的にアップデートするよう努力している。

当社は、全社的に BCP（事業継続計画）を策定しているが、当社が運用している AI システムが停止した場合に事業継続に支障を生じるおそれがあるため、BCP の発動トリガーの１つに AI インシデントを盛り込むこととし、AI システムの全部又は一部を停止することとなった場合に備えた初動対応及び事業継続のための計画を策定している。また、計画を策定するだけでは意味がなく、有事に計画を実行できないことが大きなリスクになることを認識し、毎年少なくとも一度は計画を実践するための演習を行っている。

４．運用

(1) AI マネジメントシステムの運用状況について説明可能な状態を確保する

行動目標４－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、たとえば、行動目標３－１の乖離評価プロセスの実施状況について記録するなど、AI マネジメントシステムの運用状況について対外的に説明可能な状態を確保すべきである。

【実践例１】

環境・リスク分析、ゴール設定、システムデザイン、運用、評価、環境・リスクの再分析等からなるダイナミックなプロセスのうち、「受け身」になりがちと指摘される「運用」プロセスの実践は意外と難しい。「運用」とは、簡潔に言えば、記録を残し、行動目標４－３に示すように、必要に応じて実施状況を公表することである。**この「運用」においてデータや情報を得ておくことが改善に向けた意思決定につながる**ため、当社では、環境・リスクの再分析、評価などを通じた改善の肝は「運用」にあると考えている。

当社では、**AI ガバナンスの実践に限らず、さらなる改善のために記録を残すことを重視**しており、行動目標３の記録を残すことは当然であると考えている。たとえば、個々の AI システム開発プロジェクトにおける乖離評価を記録し、AI に関する研修を実施した場合には実施概要を作成し、AI システムの開発や運用に関する社内の会議や他の事業者との会議の議事録を残し、担当者以外の関係者もそれらを閲覧できるようにしている。

当社は比較的規模の大きな企業であるため、一般的なコーポレートガバナンスに関連する行動目標には困難を感じていない。しかし、**企業内の組織的分化が進んでいることから、比較的新しい技術である AI については、部門間の専門性や理解度にギャップが生まれ、組織間連携に影響が出ないか心配**している。たとえば、行動目標３－１－２にしたがって設置した問い合わせ窓口に関し、**問い合わせ担当者が、技術的な内容を理解できずに、重大インシデントへの気づきが遅れる**ことも恐れている。行動目標３－２にしたがって、従業員のリテラシーの向上のための取り組みを実施しているが、当面は、**外部からの問い合わせについては、概要だけではなく詳細も AI マネジメント担当者に積極的に報告する**こととしている。

【実践例２】

当社はAIシステムを開発する小規模企業である。技術担当役員は全てのプロジェクトを把握しており、自らプログラミングしたり論文を読解したりするなど AI に大変詳しく、AI倫理の問題についても強い関心を持っている。そのため当社では、部門間の専門性のギャップが問題になることはないと考えている。他方で、**プロジェクトに関わる人たちの専門性が高いために、いちいち確認しなくても、行動目標ができているであろうと思い込みがち**である。そのため、**プロジェクトの進捗報告のレポートに乖離評価チェックリストを添付し、技術担当役員が必要に応じて聞き取りできるように工夫**している。

また、当社は専門性が高い集団であることから、世間の認識とのずれが生じやすい傾向があると分析している。そのため、運用状況を確認しつつ、行動目標３－３－２にしたがって**日常的な情報収集や意見交換から得られた状況を定期的に共有する**ことで、社会的受容に意識を向けるようにしている。

(2) 個々の AI システムの運用状況について説明可能な状態を確保する

行動目標４－２：AI システムを運用する企業は、経営層のリーダーシップの下、個々の AI システムの仮運用及び本格運用における乖離評価を継続的に実施するために、仮運用及び本格運用の状況をモニタリングし、結果を記録すべきである。AI システムを開発する企業は、AI システムを運用する企業による当該モニタリングを支援すべきである。

【実践例１】

当社は AI システムを運用し、当該システムを AI システム利用者に提供している企業である。AI システム開発者に依頼して AI システムを開発したが、精度だけではなく公平性にも対応できるように、データセットの内容から AI モデルの振る舞いの確認に至るまで、開発担当者から説明を受けてきた。この開発担当者からは、**開発時に想定した利用者像と実際の利用者像に違いが生じてきた場合には、精度や公平性の確保のために AI システムのメンテナンスが必要**であると言われた。

AI システムのコードを解釈できるほどの知識を持つ従業員は当社にいないことから、開発者に依頼し、**性能に大きく影響する入力や出力のログを自動的に取れるようにしてもらうとともに、モニタリングの仕方を教えてもらった**。その後、行動目標３－１の一環として、**別添２の乖離評価リストを参考にしながら性能維持のための管理方法を定めた**。現在は、この管理方法により継続的にモニタリングを行い、記録を残している。

【実践例２】

当社は、他社が運用する AI システムを開発する企業である。当社は AI システムを法的に所有しているわけではないが、保守契約を通じて、他者の運用に一定の責任を負っていて、運用者としての側面も有している。このような状況では、AI システムの性能維持のためのモニタリングにおいて、日常的に AI システムを運用している企業（実運用者）の協力が欠かせない。実際、**この実運用者は、AI システムからの出力を記録し、品質の著しい劣化を出力から判断し、実際の状況も確認した上で、再学習の必要性について当社に報告することになっており、その後の再学習の必要性に関する会議にも実運用者が加わることとしている**。

実運用者が再学習のタイミングを判断できる理由は、実運用者自身が AI システムに対して具体的に何を求めていて、具体的に何ができるかを良く理解しているからである。**AI シ**

ステム開発者は、AI システム運用者の AI システムへの期待を理解し、運用者自身が腹落ちするまで何ができるかを丁寧に説明することが重要である。「AI システム品質保証ガイドライン」に記載されているように、「品質保証の技術者やチーム、組織は、開発や営業と共に、AI システムに関する顧客の理解を深めるような活動を行う」ことが大切である[23]。

コラム：ソフトウェアによる自動モニタリング

Governance Innovation Ver.2 －アジャイル・ガバナンスのデザインと実装に向けて－では、Society5.0 におけるガバナンスの在り方が議論されている。ガバナンスシステムの運用に関して、Governance Innovation Ver.2 は、従来は断片的にしか取得できなかったデータをリアルタイムで得られるようになってきていることから、こうしたリアルタイムデータを活用することで、より効率的かつ精緻なモニタリングを行うことが可能になるだけではなく、リスク状況やゴールの達成状況を随時に判断することで、ゴールを達成するための手段を柔軟に選択できるようになり、コンプライアンスを確保しつつ持続的なイノベーションを実現していくことが可能になると指摘する。

AI システムの本格運用後のモニタリングに関しては、2020 年の CEATEC に出展された株式会社グリッドの取り組みが興味深い。学習時とは異なる特徴の時系列データや画像データが入力されていることを自動的に検知する AI 監視サービスを提供しているという。このような AI 監視サービスの利用は、AI モデルの精度の維持を通じた収益への貢献だけではなく、AI モデルの性能維持に関してモニタリングをしていることをステークホルダーに説明する責務であるアカウンタビリティとも関係する。

同様の取り組みは他にも見られる。富士通株式会社は、AI システムの本格運用後に入力データの変化の傾向を追跡して AI モデルの精度を自動推定する技術や既存モデルを再学習させることなく精度低下を抑制できる技術を開発した。この技術も AI システム運用者のアカウンタビリティや利用者の信頼度を高めることに貢献することが期待される。

AI システム運用者が、AI システム開発者との対話を通じて、AI システムの可能性や限界を理解することは重要であり、自動モニタリングツール等への過度な依存は避けるべきであるが、Society5.0 時代には自動モニタリングツールを賢く併用することも求められている。

[23] AI システム品質保証コンソーシアム『AI システム品質保証ガイドライン　2020.08 版』(2020 年 8 月)、p. 2-6.

(3) AI ガバナンスの実践状況を非財務情報に位置づけて積極的な開示を検討する

> 行動目標４－３：AI システムを開発・運用する企業は、AI ガバナンス・ゴールの設定、AI マネジメントシステムの整備や運用等に関する情報を、コーポレートガバナンス・コードの非財務情報に位置づけ、積極的に開示することを検討すべきである。上場会社以外であっても、AI ガバナンスに関する活動の情報を積極的に開示することを検討すべきである。そして、検討の結果、開示しないと判断した場合には、その理由を対外的に説明できるようにしておくべきである。

【実践例１】

当社は AI システムを開発する小規模企業である。AI システムの開発は単なる技術的な営みではなく、社会に対する深い理解に支えられていなければならないと考えており、AI ガバナンス・ゴールを明示的に設定することよりも、この考え方を社内に浸透させることを優先している。顧客や株主はこの姿勢を支持してくれている。**もちろん「人間中心の AI 社会原則」を尊重すべきであると考えているが、その背後にある思想の理解こそが重要**である。

当社は非上場会社であるため、コーポレートガバナンス・コードの対象ではないが、**ホームページ等で AI に対する上述の考え方を積極的に発信**している。当社の潜在的な顧客や当社の AI システムの利用者は、AI システムを技術的なツールではなく、社会技術的なツールであると受け止めてくれており、他社との差別化にもつながっている。

【実践例２】

当社は AI システムを開発する上場企業である。AI の適切な開発は当社の重要なテーマであるところ、すでに自社の AI ポリシーを設定し、その達成に向けた体制の整備を終えている。そして、これらの**活動内容を自社のホームページで公表し、プレス発表もした**。他方で、これらの活動について経営層から強いメッセージを発することを検討したが、当社の AI 関連事業は、現時点では、中長期的な収益に直接影響を与えないことから、そのようなメッセージを発するまでには至っていない。

このような中、先日、**ある機関投資家から企業ガバナンスに関するアンケートが届き、そこには AI 倫理への対応ぶりを聞く設問があった**。このようなアンケートに企業の中長期的な発展に対する投資家の意向が反映されているとすれば、AI 倫理も企業の健全な発展を判断するために必要な情報であることがうかがえる。今後は、改めて、**AI 倫理の取り組みを統**

合報告書に掲載することを含め、経営層からの積極的な情報発信を検討していく予定である。

コラム：コーポレートガバナンスにおける情報開示と AI 倫理の取り組み

AI ガバナンスをコーポレートガバナンスと切り離すことはできない。ここでは、AI ガバナンスの文脈で、コーポレートガバナンス・コードの基本原則の 1 つである「適切な情報開示と透明性の確保」を取り上げたい。この基本原則では、「上場会社は、会社の財政状態・経営成績等の財務情報や、経営戦略・経営課題、リスクやガバナンスに係る情報等の非財務情報について、法令に基づく開示を適切に行うとともに、法令に基づく開示以外の情報提供にも主体的に取り組むべきである。その際、取締役会は、開示・提供される情報が株主との間で建設的な対話を行う上での基盤となることも踏まえ、そうした情報（とりわけ非財務情報）が、正確で利用者にとって分かりやすく、情報として有用性の高いものとなるようにすべきである」と述べられている。AI システムの開発・運用に伴うリスクの評価や対応に関する情報についても、非財務情報の一部として開示されることが期待される場合があると考えられる。

この基本原則に関連し、本ガイドラインの作成の際に行ったヒアリングにおいて、欧州の機関投資家から AI ガバナンスについて問い合わせがあったという情報をヒアリング先から得た。AI ガバナンスへの関心が投資家の間で少しずつ高まっていることが、このような問い合わせの背景にあると考えられる。たとえば、Hermes EOS (Equity Ownership Services)は、グーグルの親会社であるアルファベットの取締役会に対して、責任を持った AI 技術の利用に関してリーダーシップを示すことを投資家が期待していると述べている[24]。AI システムの用途によっては、その利用の拡大に伴って社会へ与えるインパクトが大きくなりうるため、投資家から責任を持った AI システムの開発・運用や AI ガバナンスの整備が求められる可能性がある。

[24] Alex Rolandi, “Hermes EOS urges Alphabet to lead responsible AI practice,” funds europe (June 18, 2019), https://www.funds-europe.com/news/hermes-eos-urges-alphabet-to-lead-responsible-ai-practice.

5．評価

(1) AI マネジメントシステムが適切に機能しているかを検証する

> 行動目標５－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、AI マネジメントシステムの設計や運用から独立した者に、AI ガバナンス・ゴールに照らして、乖離評価プロセス等の AI マネジメントシステムが適切に設計され、適切に運用されている否か、つまり行動目標３、４の実践を通じ、AI ガバナンス・ゴールの達成に向けて、AI マネジメントシステムが適切に機能しているか否かを検証させるべきである。

【実践例１】

当社には AI マネジメントシステムの導入前から社内規定の運用等を監査する独立した内部監査部門がある。**AI マネジメントシステムの導入時に内部監査部門の業務範囲を拡張**し、AI マネジメントシステムをその対象に加えた。当社では、内部監査担当者が、各部門の協力を仰ぎながら、組織や規定等が、適切に運用され、有効に機能しているかを調査及び確認し、不適切な運用や機能不全が見られた場合は、当該部門に改善を求めるとともに、他部門のベストプラクティスがあれば、それを共有している。

AI システムに対する社会的受容は変化している。当社では、**社会的受容に歩調を合わせた改善こそが重要**であると考えており、**環境・リスク分析を参考にしながら、社会からの期待が高い分野やインシデントの報告数が多い分野を中心に内部監査を行っている**。改善に向けた各部門の協力が得られるように、全ての分野に対して一律に社内ルール等への厳密な適合性評価を行うのではなく、リスクの高い分野を選定している。**選定理由を伝えると各部門の協力を得られやすい**。

【実践例２】

当社は AI システムを開発する小規模企業である。AI マネジメントシステムを評価する内部監査部門を設けず、**AI マネジメントシステムに直接関与していない開発部門内の者を加えて自己監査**を行ってもらっている。この自己監査という第一の監査ラインはチェック機能不全に陥り、身内に甘くなる傾向があるため、**自己監査結果を AI ガバナンス担当役員直属の監査担当者に報告させ、報告内容を整理して、AI ガバナンス担当役員に報告することにしている**。AI ガバナンス担当役員は AI 技術や倫理に詳しいことから、自己監査中心でありなが

らも十分に機能していると考えている。現在は、第三者的な視点を強化するとともに、内部監査は改善のためにあることを伝えるために、部門横断的なフィードバック会合を開催して監査結果を共有するとともに意見交換を行うことを検討している。

【実践例３】

当社には内部監査部門があるが、AI マネジメントシステムに関しては外部監査を活用してみることにした。**外部監査には高い専門性と他社の監査経験の横展開を期待**している。**AI システムに対する社会的受容は変化していて相場感が形成されていない**。自社なりに十分に対応できていると自負していても死角があるかもしれない。

外部監査サービスはコンサルティングファームなどを中心に提供されている。外部専門家による監査を受けることで、社内外の専門的な情報を活用したアドバイスを受けることができる。また、外部専門家のアドバイスの第三者性と客観性によって、社内へのフィードバックがよりスムーズになる効果も期待している。

このようなメリットがある一方で、**受け身になる可能性**を心配している。**外部専門家はそれぞれの企業に固有の課題等に必ずしも詳しいわけではない**。外部専門家のアドバイスを最大限活用するためには、**外部監査に頼った場合でも AI に対する社会的受容を能動的に理解しようという姿勢が重要**である。

(2) 社外ステークホルダーから意見を求めることを検討する

行動目標５－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、株主だけではなく、ビジネスパートナー、消費者を含む利用者、AI システムの適切な運用をめぐる動向に詳しい有識者などの様々なステークホルダーから、AI マネジメントシステムやその運用に対する意見を求めることを検討すべきである。そして、検討の結果、実施しないと判断した場合には、その理由を対外的に説明できるようにしておくべきである。

【実践例１】

コーポレートガバナンス・コードの「株主以外のステークホルダーとの適切な協働」の章には、従業員、顧客、取引先、債権者、地域社会をはじめとする様々なステークホルダーとの適切な協働に努め、とりわけ取締役会・経営陣は、これらのステークホルダーの権利・立場や健全な事業活動倫理を尊重する企業文化・風土の醸成に向けてリーダーシップを発揮すべきであるとの原則がまとめられている。**AI システムの適切な開発・運用への関心が高まっていることから、上場会社はもちろんのこと非上場会社も、AI ガバナンスやマネジメントシステムの評価や見直しにあたっては様々なステークホルダーとの協働が求められる**。

当社は、AI ポリシーの設定やポリシー達成に向けた体制作りなどの初期設定は企業自身が行うべきものであり、その後の改善も企業自身が主体的に行うべきと考えているが、**「社会からの見え方」を知るために様々なステークホルダーとの協働も重視**している。

当社はすでに AI ポリシーを定めるとともに、AI ポリシーの意味やポリシー達成のための活動を公表している。しかし、「社会からの見え方」を知り、客観的な倫理性を確保する必要があると考え、ステークホルダーと対話を重ねていくことを目的として、**AI やそれ以外の分野の専門家で構成される AI 倫理委員会を設置**することとし、AI 技術の専門家だけでなく、法律、環境問題、消費者問題の専門家も招聘している。一般的な指摘を受けるだけでは不十分であることから、**当社の具体的な課題を提示して深い洞察を得られるように工夫**している。

【実践例２】

実践例１のような外部有識者委員会の設置のような「見える施策」に目が行きがちであるが、そのような場だけが全てではないと考えている。**重要なことは、AI 倫理や品質に関心の**

高い人々とのネットワークに緩やかにつながり、この情報交換網の中に入ることである。当社の AI マネジメント担当者には、AI 倫理や品質に関する意見交換の場で積極的に発言したり、カンファレンスなどのスピーカーを積極的に引き受けたりするように促している。もちろん、そのような活動を業績評価に含めている。

このようなアプローチでは、意見が集まらないという懸念の声を聞く。この懸念の背景には、日本人は意見交換やカンファレンスの場において本音で話さないことがあると考えられる。しかし、自分から意見を発信することで相手の意見を引き出す、いわゆる「アクティブソナー型」の人たちは、意見交換やカンファレンスの後に個人的に意見をくれる人がいることを知っている。このような意見こそが大切である。

この AI マネジメント担当者の人脈をたどり、外部講師を交えた社内研修を開催したことがあった。この研修では、当社の AI ガバナンスの取り組みを AI 関連業務に従事する従業員に対して説明することに加え、当社の取り組みをこの外部講師に評価してもらった。この外部講師は当社の AI マネジメント担当者と日常的に意見交換をしていることから、当社の事情に即したアドバイスを得ることができ、研修受講者から好評価を得たところである。

このような状況であるため、外部有識者委員会の設置についても検討しているが、今のところは必要性を感じていない。

【実践例３】

当社は、AI システムを開発し、他の事業者に納入しているが、これらの他の事業者は当該 AI システムを自身の事業で活用しているだけであり消費者には提供していない。そのため、消費者の安全、財産的価値の重大な毀損、差別などの倫理的な課題とは直接関係していないと考えている。もちろん、予測精度の維持向上の観点から納入先とはきめ細かく意見交換しているが、消費者や AI 倫理に詳しい有識者などの意見を聞く必要性はないと判断している。

６．環境・リスクの再分析

(1) 行動目標１－１から１－３を適時に再実施する

行動目標６－１:AI システムを開発･運用する企業は､経営層のリーダーシップの下、行動目標１－１から１－３について、適時に再評価、理解の更新、新たな視点の獲得などを行うべきである。なお､行動目標５－２を実施する際に､既存の AI マネジメントシステムやその運用だけではなく、環境・リスク分析を含め、AI ガバナンス全体の見直しに向けた意見を得ることも検討すべきである。

【実践例１】

現時点では AI システムに対する社会の受け止め方が定まっていないことから、行動目標１－１から１－３で挙げられている、AI システムがもたらしうる正負のインパクト、AI システムの開発や運用に関する社会的受容、自社の AI 習熟度について、適時に再評価、理解の更新、新たな視点の獲得などを行うべきである。当社では、**重大な「ヒヤリ・ハット」が生じた場合､特定のインシデントへの社会の注目が大きく高まった場合､規制環境が変化した場合等を除き､定期的に環境･リスクの分析を行い､経営層にレポート**することとしている。AI システムの適切な開発・運用をめぐる議論は非常に活発であるが、アジャイルな再分析によるガバナンス疲れを防ぎ､大きなトレンドをアジャイルに把握することを重視している。経営層への報告機会は大きなトレンドに目を向ける良い機会である。

【実践例２】

当社は、実践例１のように定期的に環境・リスクの分析を行っているが、AI ガバナンスと AI マネジメントシステムの検証には重複する要素もあることから、**定期的に開催される外部有識者を招聘した AI 倫理委員会の議題に、AI システムがもたらしうる正負のインパクトと AI システムの開発や運用に関する社会的受容を盛り込み､外部有識者からこれらの論点に関する大きなトレンドを得るようにしている**。

D. AI ガバナンス・ガイドラインに関わった有識者等

1．AI 原則の実践の在り方に関する検討会（AI 社会実装アーキテクチャー検討会）

渡部 俊也（座長）	東京大学　未来ビジョン研究センター　教授
雨宮 俊一	株式会社 NTT データ　技術開発本部　本部長
生貝 直人	一橋大学　大学院法学研究科　准教授
上野山 勝也	株式会社 PKSHA Technology　代表取締役
川上 登福	株式会社経営共創基盤　共同経営者　マネージングディレクター 一般社団法人日本ディープラーニング協会　理事
齊藤 友紀	法律事務所 LAB-01　弁護士
杉村 領一	国立研究開発法人産業技術総合研究所　情報・人間工学領域 人工知能研究戦略部　上席イノベーションコーディネータ
角田 美穂子	一橋大学　大学院法学研究科　教授
妹尾 義樹	国立研究開発法人産業技術総合研究所　イノベーション推進本部 標準化推進センター　審議役
田丸 健三郎	日本マイクロソフト株式会社　業務執行役員 ナショナル テクノロジー オフィサー
土屋 嘉寛	東京海上日動火災保険株式会社　企業商品業務部　部長
中条 薫	株式会社 SoW Insight　代表取締役社長
原 聡	大阪大学　産業科学研究所　准教授
福岡 真之介	西村あさひ法律事務所　弁護士
古谷 由紀子	サステイナビリティ消費者会議　代表
増田 悦子	公益社団法人全国消費生活相談員協会　理事長
丸山 友朗	パナソニック株式会社 イノベーション推進部門 テクノロジー本部 デジタル・AI 技術センター AI ソリューション部 主任技師
宮村 和谷	PwC あらた有限責任監査法人　パートナー
山本 龍彦	慶應義塾大学　大学院法務研究科　教授

以下、昨年度の AI 社会実装アーキテクチャー検討会のみ参加

青島 武伸	パナソニック株式会社 イノベーション推進部門 テクノロジー本部 デジタル・AI 技術センター　データアナリシス部 部長（当時）

2．AI ガバナンス・ガイドライン ワーキンググループ

生貝 直人（主査）	一橋大学　大学院法学研究科 准教授
岡田 淳	森･濱田松本法律事務所　弁護士
齊藤 友紀	法律事務所 LAB-01　弁護士
中崎 尚	アンダーソン・毛利・友常法律事務所外国法共同事業　弁護士
本橋 洋介	日本電気株式会社　AI・アナリティクス事業部 シニアマネジャー
宮村 和谷	PwC あらた有限責任監査法人　パートナー

3．協力者

本ガイドラインの作成にあたっては、上記の委員に加え、多くの専門家にご指導いただいている。この場を借りて御礼を申し上げたい。

(1) 上記検討会における講演（講演順）

荻野 武	キユーピー株式会社 生産本部 未来技術推進担当部長（当時）
大岩 寛	国立研究開発法人産業技術総合研究所　情報・人間工学領域 デジタルアーキテクチャ研究センター　副研究センター長
杉村 領一	国立研究開発法人産業技術総合研究所　情報・人間工学領域 人工知能研究戦略部 上席イノベーションコーディネータ
長宗 豊和	Japan Business Council in Europe 事務局長
齋藤 千紘	外務省 在ストラスブール総領事館 領事

(2) 上記ワーキンググループを拡大した事前コンサルテーションへの参加

今田 俊一	ソニーグループ株式会社　AI コラボレーション・オフィス AI 倫理室　統括課長
大岩 寛	国立研究開発法人産業技術総合研究所　情報・人間工学領域 デジタルアーキテクチャ研究センター　副研究センター長
菊池 慎司	富士通株式会社　富士通研究所　人工知能研究所　AI 品質 PJ 主管研究員
杉村 領一	国立研究開発法人産業技術総合研究所　情報・人間工学領域 人工知能研究戦略部 上席イノベーションコーディネータ
妹尾 義樹	国立研究開発法人産業技術総合研究所　イノベーション推進本部 標準化推進センター　審議役

曽我部　完	株式会社グリッド 代表取締役
藤田 雅博	ソニーグループ株式会社 AI コラボレーション・オフィス 担当 VP
松本 敬史	有限責任監査法人トーマツ マネジャー 東京大学 未来ビジョン研究センター 客員研究員
美馬 正司	株式会社日立コンサルティング スマート社会基盤コンサルティング第 2 本部　ディレクター 慶應義塾大学　政策・メディア研究科　特任教授
山本 力弥	ソフトバンクロボティクス株式会社　事業開発本部 Humanoid Division Director

(3) 事務局による個別ヒアリング

上述の検討会等に加え、企業等の取り組みを個別にお聞きした。Musashi AI 株式会社、一般社団法人日本ディープラーニング協会、日本電気株式会社、国立研究開発法人産業技術総合研究所、エヌビディア合同会社、富士通株式会社、株式会社日立製作所、株式会社メルカリ、横河電機株式会社、損害保険ジャパン株式会社、一般社団法人 AI ビジネス推進コンソーシアム、株式会社 J.Score、有限責任監査法人トーマツ、東京海上日動火災保険株式会社、パナソニック株式会社、株式会社グリッド、千代田化工建設株式会社、ソニーグループ株式会社、日立造船株式会社、ソフトバンク株式会社、ソフトバンクロボティクス株式会社に所属の多くの専門家・有識者にお世話になった。漏れがあるとすれば、全て事務局の責任である。

なお、協力者は本ガイドラインの決定には加わっていない。

4．事務局

松田　洋平	経済産業省 商務情報政策局 情報経済課 課長（～2021 年 6 月 30 日）
須賀　千鶴	経済産業省 商務情報政策局 情報経済課 課長（2021 年 7 月 1 日～）
泉 卓也	経済産業省 商務情報政策局 情報経済課 情報政策企画調整官 （執筆主担当）
羽深 宏樹	経済産業省 商務情報政策局 情報経済課 ガバナンス戦略国際調整官 （別添 3 主担当）
野村 至	経済産業省 商務情報政策局 情報経済課 課長補佐

E. 参考文献

１．政府・公的機関の文献

- 経済産業省『「GOVERNANCE INNOVATION Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」報告書（案)』(2021 年 2 月 19 日)（パブリックコメントを反映した「GOVERNANCE INNVOATION Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」は、2021 年 7 月中に公表される予定）
- 経済産業省『「GOVERNANCE INNOVATION： Society5.0 の実現に向けた法とアーキテクチャのリ・デザイン」報告書』(2020 年 7 月 13 日）
- 経済産業省『AI・データの利用に関する契約ガイドライン 1.1 版』(2019 年 12 月）
- 経済産業省、厚生労働省、消防庁『プラント保安分野 AI 信頼性評価ガイドライン 第 2 版』、『信頼性評価実用例実施記録』(2021 年 3 月 30 日）
- 経済産業省『AI・データサイエンス人材育成に向けたデータ提供に関する実務ガイドブック』(2021 年 3 月 1 日）
- 経済産業省『デジタルガバナンス・コード』(2020 年 11 月 9 日）
- 総務省、経済産業省『DX 時代における企業のプライバシーガバナンスガイドブック ver1.0』(2020 年 8 月）
- 総務省､AI ネットワーク社会推進会議『国際的な議論のための AI 開発ガイドライン案』（2017 年 7 月）
- 総務省、AI ネットワーク社会推進会議『AI 利活用ガイドライン』(2019 年 8 月）
- 消費者のデジタル化への対応に関する検討会 AI ワーキンググループ『消費者のデジタル化への対応に関する検討会 AI ワーキンググループ報告書』(2020 年 7 月）
- 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)』(2020 年 7 月）（デジタルアーキテクチャ研究センター・サイバーフィジカルセキュリティ研究センター・人工知能研究センター テクニカルレポート DigiARC-TR-2021-01 / CPSEC-TR-2021001）
- 独立行政法人情報処理推進機構『AI 白書 2020 ～広がる AI 化格差（ギャップ）と 5 年先を見据えた企業戦略～』(2020 年 3 月 2 日）
- 東京証券取引所『コーポレートガバナンス・コード　～会社の持続的な成長と中長期的な企業価値の向上のために～』(2021 年 6 月 11 日）
- Executive Office of the President, Office of Management and Budget, “Memorandum to

the Heads of Executive Departments and Agencies, Guidance for Regulation of Artificial Intelligence Applications" (November 17, 2020)

- European Commission, "White Paper on Artificial Intelligence – A European approach to excellence and trust" (February 19, 2020)
- The High-Level Expert Group on Artificial Intelligence (AI HLEG), "Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment" (July 17, 2020)
- European Commission, "Proposal for a Regulation laying down harmonised rules on artificial intelligence" (April 21, 2021)
- Info-communications Media Development Authority and Personal Data Protection Commission, "Model Artificial Intelligence Governance Framework Second Edition" (January 21, 2020)
- World Economic Forum, "Companion to the Model AI Governance Framework – Implementation and Self-Assessment Guide for Organizations" (January 2020)
- P J. Phillips, Amanda C. Hahn, Peter C. Fontana, David A. Broniatowski, Mark A. Przybocki, "Four Principles of Explainable Artificial Intelligence (Draft), NIST Interagency/Internal Report (NISTIR) - 8312-draft" (August 19, 2020)
- OECD, "OECD FRAMEWORK FOR THE CLASSIFICATION OF AI SYSTEMS – PUBLIC CONSULTATION ON PRELIMINARY FINDINGS" (May 2020).
- Carolyn Nguyen, Adam Murray, and Barry O'Brien, "What are the tools for implementing trustworthy AI? A comparative framework and database," The AI Wonk, OECD.AI (May 25, 2021)

２．企業・産業界の文献

- AI システム品質保証コンソーシアム『AI システム品質保証ガイドライン（2020 年 08 版)』(2020 年 8 月)
- 日本経済団体連合会『AI 活用戦略　～AI-Ready な社会の実現に向けて～』(2019 年 2 月 19 日)
- 日本電気株式会社『NEC グループ　AI と人権に関するポリシー』(2019 年 4 月)
- 富士通株式会社『富士通グループ AI コミットメント』(2019 年 3 月 13 日)
- ソニーグループ株式会社『ソニーグループ AI 倫理ガイドライン』(2019 年 3 月改定)

- 株式会社日立製作所『AI 倫理原則』（2021 年 2 月 22 日）
- Andrew Ng, “AI Transformation Playbook”, available at https://landing.ai.

３．個人・学術系の文献

- パーソナルデータ＋α研究会『プロファイリングに関する提言案』NBL 1137 号（2019 年 1 月 1 日）
- パーソナルデータ＋α研究会『プロファイリングに関する提言案付属 中間報告書』NBL 1137 号（2019 年 1 月 1 日）
- 山本龍彦『AI と憲法』日本経済新聞出版社（2018 年 8 月 24 日）
- 浅川伸一、江間有紗、工藤郁子、巣籠悠輔、瀬谷啓介、松井孝之、松尾豊『ディープラーニング G 検定（ジェネラリスト）公式テキスト』翔泳社（2018 年 10 月 22 日）
- 舟山聡『AI の責任と倫理(第 2 回)AI 倫理に対する企業の取組み(1)』NBL 1170 号（2020 年 5 月 15 日）
- 荒堀淳一『AI の責任と倫理（第 3 回）AI 倫理に対する企業の取組み(2)』NBL 1172 号（2020 年 6 月 15 日）
- 齊藤友紀『AI の責任と倫理（第 4 回）AI 倫理とアカウンタビリティ、法的責任』NBL 1174（2020 年 7 月 15 日）
- 松田千恵子『これならわかる コーポレートガバナンスの教科書』日経 BP 社（2015 年 8 月 11 日）
- 松田千恵子『ESG 経営を強くする コーポレートガバナンスの実践』日経 BP 社（2018 年 12 月 24 日）
- 國廣正『企業不祥事を防ぐ』日本経済新聞出版社（2019 年 10 月 17 日）
- 平林良人、奥野麻衣子『ISO 共通テキスト《附属書 SL》解説と活用　ISO マネジメントシステム　構築組織のパフォーマンス向上』日本規格協会（2015 年 10 月 14 日）
- 古谷由紀子『現代の消費者主権』芙蓉書房出版（2017 年 5 月 19 日）
- Cathy O’Neil, “Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy” New York: Crown Publishers (2016)

F. 別添１（行動目標一覧）

行動目標１－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、AI システムから得られる正のインパクトだけではなく意図せざるリスク等の負のインパクトがあることも理解し、これらを経営層に報告し、経営層で共有し、適時に理解を更新すべきである。	
行動目標１－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、本格的な AI の提供に先立ち、直接的なステークホルダーだけではなく潜在的なステークホルダーの意見に基づいて、社会的な受容の現状を理解すべきである。また、本格的な AI システムの運用後も、適時にステークホルダーの意見を再確認するとともに、新しい視点を更新すべきである。	
行動目標１－３：AI システムを開発・運用する企業は、経営層のリーダーシップの下、行動目標１－１、１－２の実施を踏まえ、自社の事業領域や規模等に照らして負のインパクトが軽微であると判断した場合を除き、自社の AI システムの開発・運用の経験の程度、AI システムの開発・運用に関与するエンジニアを含む従業員の人数や経験の程度、当該従業員の AI 技術及び倫理に関するリテラシーの程度等に基づいて、自社の AI 習熟度を評価し、適時に再評価すべきである。負のインパクトが軽微であると判断し、AI 習熟度の評価をしない場合には、その理由等をステークホルダーに説明できるようにしておくべきである。	
行動目標２－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、「人間中心の AI 社会原則」を踏まえ、AI システムがもたらしうる正負のインパクト、AI システムの開発や運用に関する社会的受容、自社の AI 習熟度を考慮しつつ、設定に至るプロセスの重要性にも留意しながら、自社の AI ガバナンス・ゴール（たとえば AI ポリシー）＊を設定するか否かについて検討すべきであり、潜在的な負のインパクトが軽微であることを理由に AI ガバナンス・ゴールを設定しない場合には、その理由等をステークホルダーに説明できるようにしておくべきである。「人間中心の AI 社会原則」が十分に機能すると判断した場合は、自社の AI ガバナンス・ゴールに代えて「人間中心の AI 社会原則」をゴールとしてもよい。なお、ゴールを設定しない場合であっても、「人間中心の AI 社会原則」の重要性を理解し、行動目標３から５に係る取り組みを適宜実施することが望ましい。	
行動目標３－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、自社が開発・運用している AI システムの AI ガバナンス・ゴールからの乖離を特定し、乖離により生じる影響を評価した上、負のインパクトが認められる場合、その大きさ、範囲、発生頻度等を考慮して、その受容の合理性の有無を判定し、受容に合理性が認められない場合に AI の開発・運用の在り方について再考を促すプロセスを、AI システムの設計段階、開発段階、利用開始前、利用開始後などの適切な段階に組み込むべきである。運営層はこのプロセスの具体化を行うべきである。そして、AI ガバナンス・ゴールとの乖離評価には AI システムの開発や運用に直接関わっていない者が加わるようにすべきである。なお、乖離があることのみを理由として AI の開発・提	

供を不可とする対応は適当ではない。そのため、乖離評価は負のインパクトを評価するためのステップであって、改善のためのきっかけにすぎない。	
行動目標３－１－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、業界における標準的な乖離評価プロセスの有無を確認し、そのようなプロセスが存在する場合には、それを自社のプロセスに取り込むべきである。	
行動目標３－１－２：AI システム利用者にサービスを提供している AI システム運用企業は、経営層のリーダーシップの下、提供している AI システムに一定の乖離が発生しうる場合には、AI システム利用者に対して、その事実や当該乖離への対応策に関する十分な情報を提供するとともに、問い合わせ先も明確にすべきである。	
行動目標３－１－３：データを提供する企業は、AI システムを開発する企業が適切に乖離評価をできるようにするために、データの収集元、収集方針、収集基準、アノテーション付与基準、利用制約等のデータセットに関する情報を提供すべきであり、AI システム開発者は十分な情報を提供するデータ事業者からデータセットを取得すべきである。	
行動目標３－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、AI マネジメントシステムを適切に運営するために、外部の教材の活用を検討し、AI リテラシーを戦略的に向上させるべきである。たとえば、AI システムの開発・運用における法的・倫理的側面に責任を負う役員、マネジメントチーム、担当者には AI 倫理に関する一般的なリテラシー向上のための教育を、AI システムの開発・運用プロジェクトの担当者には AI 倫理だけではなく AI 技術に関する研修を提供することが考えられる。データを提供する企業は、AI システム開発者・運用者の実践例を参考に、データ提供に関わる担当者の AI 倫理に関する一般的なリテラシーを向上させるべきである。	
行動目標３－３：AI システムを開発・運用する企業、及び、データを提供する企業は、学習等用のデータセットの準備から AI システムの開発・運用までの全てを自部門で行う場合を除き、経営層のリーダーシップの下、営業秘密等に留意しつつ、自社や自部門のみでは十分に実施できない AI システムの運用上の課題と当該課題の解決に必要な情報を明確にし、積極的に共有すべきである。その際に、必要な情報交換が円滑に行われるよう、AI システム開発者、AI システム運用者、データ事業者の間で予め情報の開示範囲について合意し、秘密保持契約の締結等を検討することが望ましい。	
行動目標３－３－１：AI システムを開発・運用する企業、及び、データを提供する企業は、経営層のリーダーシップの下、学習等用のデータセットの準備から AI システムの開発・運用までの全てを自社で行う場合を除き、営業秘密に留意しつつ、複数事業者間の情報共有の現状を理解し、適時に理解を更新すべきである。	

行動目標３－３－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、日常的に、AI システムの開発や運用に関するルール整備、ベストプラクティス、インシデントなどの情報を収集するとともに、社内外の意見交換を奨励すべきである。	
行動目標３－４：AI システムを開発・運用する企業、及び、データを提供する企業は、経営層のリーダーシップの下、インシデントの予防と早期対応を通じて利用者のインシデント関連の負担を軽減すべきである。	
行動目標３－４－１：AI システムを開発・運用する企業、及び、データを提供する企業は、経営層のリーダーシップの下、学習等用のデータセットの準備からシステムの開発・運用までの全てを自社で行う場合を除き、負のインパクトを全体で最小化できるように AI システムの不確実性への対応負担を複数事業者間で分配すべきである。	
行動目標３－４－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、AI インシデント又は紛争発生時に、AI システム利用者への説明、影響範囲や損害の特定、法的関係の整理、被害救済措置、被害拡大防止措置、再発防止策の検討等を速やかに実施するため、対応方針の決定や計画の策定を検討するとともに、当該対応方針又は計画に関して適宜実践的な予行演習の実施を検討すべきである。	
行動目標４－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、たとえば、行動目標３－１の乖離評価プロセスの実施状況について記録するなど、AI マネジメントシステムの運用状況について対外的に説明可能な状態を確保すべきである。	
行動目標４－２：AI システムを運用する企業は、経営層のリーダーシップの下、個々の AI システムの仮運用及び本格運用における乖離評価を継続的に実施するために、仮運用及び本格運用の状況をモニタリングし、結果を記録すべきである。AI システムを開発する企業は、AI システムを運用する企業による当該モニタリングを支援すべきである。	
行動目標４－３：AI システムを開発・運用する企業は、AI ガバナンス・ゴールの設定、AI マネジメントシステムの整備や運用等に関する情報を、コーポレートガバナンス・コードの非財務情報に位置づけ、積極的に開示することを検討すべきである。上場会社以外であっても、AI ガバナンスに関する活動の情報を積極的に開示することを検討すべきである。そして、検討の結果、開示しないと判断した場合には、その理由を対外的に説明できるようにしておくべきである。	
行動目標５－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、AI マネジメントシステムの設計や運用から独立した者に、AI ガバナンス・ゴールに照らして、乖離評価プロセス等の AI マネジメントシステムが適切に設計され、適切に運用されている否か、つまり行動目標３、４の実践を通じ、AI ガバナンス・ゴールの達成に向けて、AI マネジメントシステムが適切に機能しているか否かを検証させるべきである。	

行動目標５－２：AI システムを開発・運用する企業は、経営層のリーダーシップの下、株主だけではなく、ビジネスパートナー、消費者を含む利用者、AI システムの適切な運用をめぐる動向に詳しい有識者などの様々なステークホルダーから、AI マネジメントシステムやその運用に対する意見を求めることを検討すべきである。そして、検討の結果、実施しないと判断した場合には、その理由を対外的に説明できるようにしておくべきである。	
行動目標６－１：AI システムを開発・運用する企業は、経営層のリーダーシップの下、行動目標１－１から１－３について、適時に再評価、理解の更新、新たな視点の獲得などを行うべきである。なお、行動目標５－２を実施する際に、既存の AI マネジメントシステムやその運用だけではなく、環境・リスク分析を含め、AI ガバナンス全体の見直しに向けた意見を得ることも検討すべきである。	

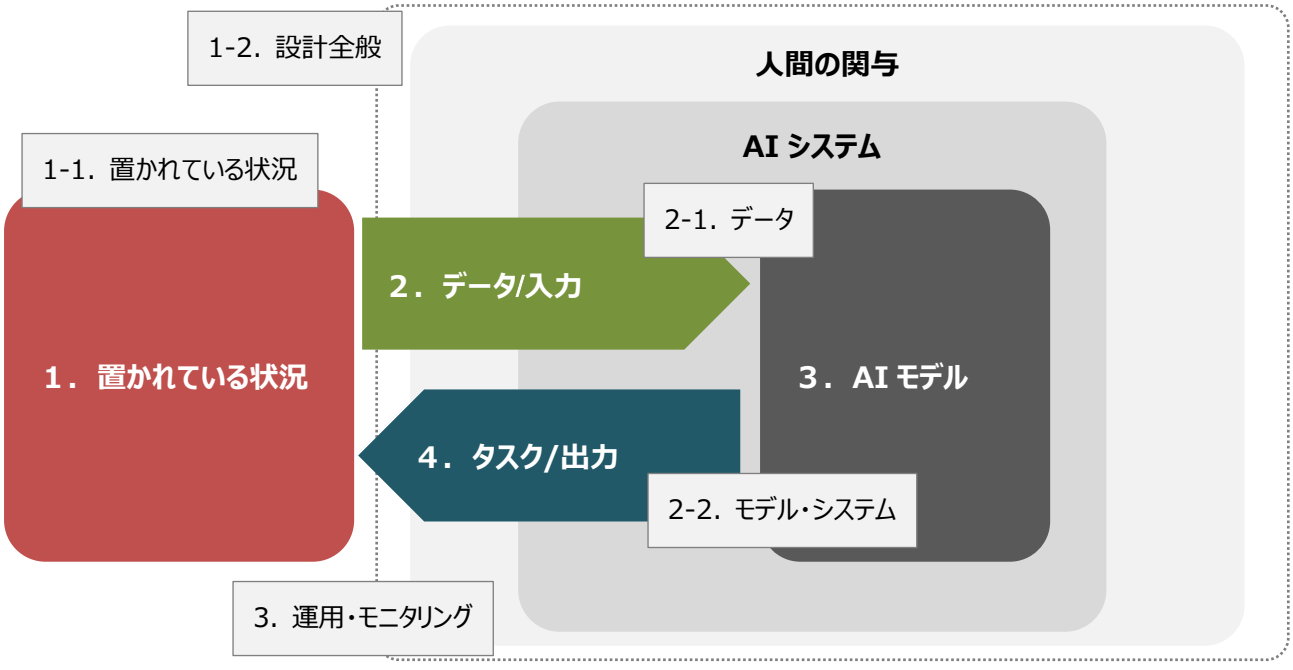
G. 別添２（AI ガバナンス・ゴールとの乖離を評価するための実務的な対応例）

この乖離評価例は、個々の AI システムの開発及び運用において、AI ガバナンス・ゴールへの適合性の評価を支援するツールである。ここで例示している評価項目例は、これを用いた評価プロセスとその他の AI ガバナンス要素とともに、必要に応じて、各 AI システム開発・運用者のガバナンスシステムに組み込まれ、「人間中心の AI 社会原則」という社会的なゴールの達成に向けた支援ツールとして機能することが期待される。

AI システムの開発・運用の内容、目的、方法は多様であり、実証事業と本格開発との間でも事情は異なることから、乖離評価をあらゆる AI システムの開発・運用を対象として一律に行うことは必ずしも想定されていない。評価の対象の選定及びその基準の策定は AI システムの開発者・運用者の合理的な裁量に委ねられる。自らの事業等に照らし、AI ガバナンス・ゴールからの乖離による負のインパクトが大きくなりうる AI システムの開発・運用の類型をリスクベースで抽出し、乖離評価を行うべきである。

「A．はじめに」でも述べたように、**乖離評価例は、各 AI 利用者が置かれた個別具体的な状況までは考慮されておらず、評価の対象によっては、乖離評価例として提供する評価項目例は不十分であることも過剰であることもありうる**。そのため、**その採否は AI システムの開発・運用者の任意**に委ねられることはもちろん、**採用する場合であっても各自の事情に応じた修正や取捨選択を検討する必要がある**。

この乖離評価例の観点は、個々のプロジェクトが置かれている状況、個々のプロジェクトの設計全般、データ、モデル・システム、運用・モニタリングからなる。これらの観点を OECD のフレームワークを参考にした概念図にマッピングすると以下のとおりとなる。



1. 企画・設計段階		
評価項目例	具体的な項目例	
1-1. 置かれている状況		
A. AI システム開発者及び運用者は潜在的な利用者を想定できているか → 1-2. 設計全般 A → 3. 運用・モニタリング A	例：AI システム開発者及び運用者は、開発・運用しようとしている AI システムの潜在的な利用者の AI リテラシー及び AI システム利用経験を想定できているか 例：AI システム開発者及び運用者は、開発・運用しようとしている AI システムの潜在的な利用者に子供、高齢者、社会的弱者などが含まれうること/含まれえないことを把握できているか	
B. AI システム開発者及び運用者は潜在的な用途を想定できているか → 1-2. 設計全般 B → 3. 運用・モニタリング B	例：AI システム開発者は、開発しようとしている AI システムの提供目的を AI システム運用者から聴取したか 例：AI システム開発者と運用者が同じ用途を想定しているか 例：AI システム開発者及び運用者は、開発・運用しようとしている AI システムの典型的な用途だけではなく、悪用されうる可能性があることも把握できているか	
C. AI システム開発者は AI システム運用者の AI リテラシー及び経験を把握しているか → 1-2. 設計全般 C → 3. 運用・モニタリング C	例：AI システム開発者は、AI システム運用者の AI 人材の数、AI に関する研修機会の数、AI システムの運用実績を把握しているか 例：AI システム開発者は、AI リテラシー及び経験が不足している場合、AI システム運用者に AI リテラシーを高める意欲があるか把握しているか	
D. AI システム開発者及び運用者は、AI システムに求められる身体、精神、財産等への悪影響を把握しているか	例：AI システム開発者及び運用者は、開発・運用しようとしている AI システムや類似の AI システムの身体、精神、財産等への悪影響に関するインシデント事例を調査したか	

→ 1-2. 設計全般 D → 3. 運用・モニタリング D	例：AI システム開発者及び運用者は、開発・運用しようとしているAIシステムが身体、精神、財産等に与えうる損害の大きさ及び発生頻度を把握しているか 例：AIシステム開発者は、開発しようとしているAIシステムが身体、精神、財産等に影響を与えうる場合、業界等の標準的な実務などに照らして許容されるリスクの評価を行ったか	
E. AIシステム開発者及び運用者は、AIシステムに求められる公平性を把握しているか → 1-2. 設計全般 E → 3. 運用・モニタリング E	例：AI システム開発者及び運用者は、開発・運用しようとしている AI システムや類似の AI システムの公平性に関するインシデント事例を調査したか 例：AIシステム開発者及び運用者は、AIシステムが利用される国・地域において、想定される利用者全体の一部に関連して、偏見や差別的な扱いやそれらが残存しているとの指摘があるか否かを確認したか 例：AIシステム開発者及び運用者は、AIシステムが利用される国・地域において、想定される利用者全体の一部に関連して、同一又は類似のAIシステムが偏見や差別的な扱いを生み出したり、再現したり、拡大したりしたとの指摘があるか否かを確認したか 例：AI システム開発者は、可能な場合には、公平性に関して提案されている指標から適切なものを選択し、許容される範囲に関する評価を行ったか	
F. AIシステム開発者及び運用者は、AIシステムに期待されている個人への配慮事項を理解しているか → 1-2. 設計全般 F → 3. 運用・モニタリング F	例：AI システム開発者及び運用者は、開発・運用しようとしている AI システムや類似の AI システムについて、個人への配慮を欠いたインシデント事例を調査したか 例：AIシステム開発者及び運用者は、AIシステムが多くの断片的な情報を用いて特定の個人の細かい特徴を明らかにする能力があることだけではなく、その能力に対する懸念が	

	示されていることを理解しているか	
G. AI システム開発者及び運用者は、AI システムに求められるサイバーセキュリティ上の課題を把握しているか →　1-2. 設計全般 G →　3. 運用・モニタリング G	例：AI システム開発者及び運用者は、開発・運用しようとしている AI システムがインターネットと接続して利用される場合には、その AI システムや類似の AI システムのサイバーセキュリティに関するインシデント事例を調査したか 例：AI システム開発者及び運用者は、開発・運用しようとしている AI システムがインターネットと接続して利用される場合には、業界等の標準的な実務などに照らしてサイバーセキュリティ上の課題を把握したか	
1-2. 設計全般		
A. AI システム開発者は、潜在的な利用者のリテラシーや経験不足の課題に対処したか	例：AI システム開発者は、利用者の理解を高めるために、開発しようとしている AI システムのインターフェースの工夫をしたり、AI システム運用者及び利用者への注意事項をまとめたりしたか 例：AI システム開発者は、利用者の理解を高めるために、AI システム運用者及び利用者向けに、開発しようとしている AI システムの利点だけではなく、限界をわかりやすくまとめたか	
B. AI システム開発者は、予見可能な悪用に関する課題に対処したか	例：AI システム開発者は、開発しようとしている AI システムに関する予見可能な悪用を設計によって排除可能か否か検討し、排除可能な場合には、そのような設計を施したか 例：AI システム開発者は、開発しようとしている AI システムに関する予見可能な悪用を設計によって排除できない場合、AI システム運用者への注意事項をまとめたか 例：AI システム開発者は、AI システムを受け渡した時に特定されていた以外の目的に利用されないように、AI システム運用者に目的外利用の禁止について契約で明確にしたか	

C. AI システム開発者は、AI システム運用者のリテラシーや経験不足の課題に対処したか	例：AI システム開発者は、AI システム運用者の理解を高めるために、開発しようとしている AI システムのインターフェースの工夫をしたり、AI システム運用者への注意事項をまとめたりしたか 例：AI システム開発者は、AI システム運用者の理解を高めるために、AI システム運用者に、開発しようとしている AI システムの利点だけではなく、限界をわかりやすくまとめたか	
D. AI システム開発者は、AI システムの身体、精神、財産等への悪影響に関する課題に対処したか	例：AI システム開発者は、身体、精神、財産等への悪影響に関し、AI システム全体で除去できる予見可能なリスクと対応できない残存リスクに分け、残存リスクについては可能な限り緩和するとともに、AI システム運用者及び利用者への注意喚起や利用方法の工夫によって残存リスクを管理できることを確認したか 例：AI システム開発者は、入力データに異常値が含まれた場合に生じうる身体、精神、財産等への悪影響を想定し、対応策を講じたか 例：AI システム開発者は、入力データに悪意のあるデータや敵対的データが含まれる可能性を検討し、それらが含まれた場合に生じうる身体、精神、財産等への悪影響を想定し、対応策を講じたか 例：AI システム開発者は、出力データに異常値が含まれた場合に生じうる身体、精神、財産等への悪影響を想定し、対応策を講じたか 例：AI システム開発者は、AI モデルの説明可能性に伴う課題に対処するために、AI モデル自体への配慮だけではなく、AI モデルを含むシステム全体の冗長性や安全機能の追加についても検討したか	

E. AI システム開発者は、AI システムの公平性に関する課題に対処したか	例：AI システム開発者は、可能な限り、AI システム開発チームの多様性を高めたか 例：AI システム開発者は、他国・地域に AI システムを提供することを想定している場合に、可能な限り、その国・地域の規制、慣習、商慣行等を理解する人を AI システム開発チームや AI システムレビュー役に加えたか 例：AI システム開発者は、公平性に関して提案されている指標から適切なものを選択した場合、許容される範囲を超えた出力データに対して警告を発する措置を講じたか 例：AI システム開発者は、公平性に関して提案されている指標から適切なものを選択した場合、許容される範囲を超えたデータが出力される可能性がある場合に、その旨を AI システム運用者及び利用者への注意事項としてまとめたか	
F. AI システム開発者は、AI システムに対して期待されている個人への配慮事項に対処したか	例：AI システム開発者は、個人を分析する AI システムの開発にあたって、個人の機会や意思決定を不当に害しないことを確保するとともに、そのために実施した対応策を記録し、AI システム運用者及び利用者に説明できるようにしたか 例：AI システム開発者は、匿名化されたデータ等から個人を特定するような分析を行っているか否かを確認したか	
G. AI システム開発者は、AI システムのサイバーセキュリティに関する課題に対処したか	例：AI システム開発者は、開発・運用しようとしている AI システムがインターネットと接続して利用される場合には、業界等の標準的な実務などに照らして、不正なアクセスやデータの入力に対処する措置を講じたか 例：AI システム開発者は、不正なアクセスやデータの入力に対処する措置について AI システム運用者への注意事項としてまとめたか	
H. AI システム開発者は、必要な場合	例：AI システム開発者は、身体、精神、財産等への悪影響の	

に、システムの設計上、人間の主体的な関与の機会を確保したか	低減や公平性向上の観点から、人間による制御可能性を含め、人間の主体的な関与の機会の必要性を検討したか 例：AI システム開発者は、必要な場合には、AI システムの採否や利用の中断・停止を決定する自由や機会を AI システム利用者に与える設計としたか 例：AI システム開発者は、AI システムの振る舞いに問題があった場合に、AI を使わないプロセスに変更するなど、リスク回避の仕組みを整えたか 例：AI システム開発者は、必要な場合には、AI システム利用者が意思決定に際して AI システムに過剰に依存しないような設計を採用したか	
I. AI システム開発者は、AI システムの機能、効果について、AI 以外のシステムと比較しながら、AI システム運用者とすりあわせをしたか	例：AI システム開発者は、AI システム運用者が AI システムに期待する機能、効果を理解したか 例：AI システム開発者は、AI システム運用者が AI システムに期待する機能、効果について、AI 以外のシステムで実施可能であり、かつ、精度等に大きな変化が生じない場合には、AI 以外のシステムの代替案を AI システム運用者に伝え再考を促し、依然として AI システムの開発を選好する場合には、自社及び社会の利益が明確か確認したか 例：AI システム開発者は、精度と説明可能性等の AI システムのトレードオフを考慮して、AI システム運用者が AI システムに期待する機能、効果を提供できない場合には、その旨を説明し、必要に応じて代替策を提案したか	
J. AI システム開発者は、AI システム運用時のモニタリングを容易にする設計をしたか	例：AI システム開発者は、AI システム運用者が AI システムの動作状況をモニタリングできるように、入力・出力データのログの取得を可能としたか 例：AI システム開発者は、AI システム運用者のモニタリングを支援するために、AI システムの動作状況をまとめた一	

	覧表示（AI システムの運用担当者、運用時間、入力・出力ログなど）を可能としたか 例：AI システム開発者は、AI システム運用者が、経営層や社外から要請を受けた場合に、適当な時間で運用状況を整理できるような設計としたか 例：AI システム開発者は、AI システムの性能変化や運用環境におけるデータ分布の変化など、AI システム運用者が再学習の必要性に気がつく工夫を施したか	

2. 開発段階		
評価項目例	具体的な項目例	
2-1. データ		
A. データ事業者及び AI システム開発者は、適法、公正、一般的に妥当な方法でデータを取得・収集しているか	例：データ事業者及び AI システム開発者は、AI システムの開発にあたって、学習・検証・試験に必要なデータの範囲を明確にし、必要な範囲でデータを取得・収集したか 例：データ事業者及び AI システム開発者は、他者からデータを取得・収集する場合には、関連する法令、ガイドライン、業界における標準的な手法の有無及び内容を確認し、存在する場合には、関連する法令に従い、ガイドラインや標準的な手法を尊重したか 例：データ事業者及び AI システム開発者は、データの提供の可否についてデータの主体の選択の余地が事実上ない状態を作り出さないように配慮したか	
B. データ事業者及び AI システム開発者は、適法、公正、一般的に妥当な方法でデータを管理・利用しているか	例：データ事業者及び AI システム開発者は、他者から取得したデータを利用する場合には、関連する法令、ガイドライン、業界における標準的な手法の有無及び内容を確認し、存在する場合には、関連する法令に従い、ガイドラインや標準	

	的な手法を尊重したか 例：データ事業者及び AI システム開発者は、データの来歴、データの加工方法等のデータの管理に必要な情報を記録しているか 例：データ事業者及び AI システム開発者は、一部のデータの利用を停止したり、削除したりできる形式でデータを管理しているか 例：データ事業者及び AI システム開発者は、学習・検証・試験用に管理しているデータの漏洩や改ざんを防止するための措置を講じたか 例：データ事業者及び AI システム開発者は、学習・検証・試験用に管理しているデータへの権限のない者によるアクセスを監視するために、アクセスログを取得しているか 例：データ事業者及び AI システム開発者は、データの収集・加工・利用、データへのアクセスの監視などのデータの管理に関する状況を説明できるようにしているか 例：データ事業者及び AI システム開発者は、不特定の者からデータを広く収集している場合には、疑問やトラブルが発生した際に対応する窓口等を設置しているか	
C. データ事業者及び AI システム開発者は、データの品質を確保しているか	例：データ事業者及び AI システム開発者は、学習・検証・試験用のデータセットに、理由が示されていない、異常値、外れ値、エラー値、欠損値が多く含まれないようにしているか 例：データ事業者及び AI システム開発者は、学習・検証・試験用のデータセットに外れ値や欠損値を補う加工データが存在するか、存在する場合にはその加工方法について確認	

	したか 例：データ事業者及びAIシステム開発者は、学習・検証・試験用のデータセットに、悪意のあるデータや敵対的データが含まれないようにしているか 例：データの加工に関して、再現性があるプロセスが規定されているか	
D. AIシステム開発者は、開発しようとするAIシステムに要求される機能の獲得に必要なデータの属性を考慮したか	例：AIシステム開発者は、AIシステム運用時の典型的なケースだけではなく、出現頻度が低いケースを想定しながら、開発しようとするAIシステムに要求される機能の獲得に必要なデータの属性を特定したか 例：AIシステム開発者は、想定可能なシナリオを考慮して、開発しようとするAIシステムに要求される機能の獲得に必要なデータの属性を特定したか 例：AIシステム開発者は、AIシステムが運用・利用される分野の専門家の支援を得て、特定したデータの属性の妥当性や網羅性の検証を行ったか 例：AIシステム開発者は、開発しようとするAIシステムに要求される機能の獲得に必要なデータの属性の特定にあたり、他国・地域にAIシステムを提供することを想定している場合には、その国・地域の規制、慣習、商慣行、文化等を考慮したか 例：AI システム開発者は、人間と機械の識別特性に違いがあることを考慮して、開発しようとするAIシステムに要求される機能の獲得に必要なデータの属性を特定したか	
E. AIシステム開発者は、開発しようと	例：AIシステム開発者は、開発しようとするAIシステムに	

するAIシステムに要求される機能の獲得に必要とされる全ての属性のデータを網羅的かつ十分に用意したか	要求される機能の獲得に必要とされる全ての属性に対応する学習・検証・試験用のデータを漏れなく用意したか 例：AIシステム開発者は、開発しようとするAIシステムに要求される機能の獲得に必要とされる全ての属性ごとに十分な量の学習・検証・試験用のデータを用意したか	
F. AIシステム開発者は、データ属性ごとのデータ量に大きなばらつきがないことを確認したか	例：AIシステム開発者は、開発しようとするAIシステムに要求される機能の獲得に必要とされる全ての属性の間で、用意された学習・検証・試験用のデータの量に大きなばらつきがないことを確認したか 例：AI システム開発者は、上記確認の結果、大きなばらつきが確認された場合に、出現頻度の低いケースの重点的な学習について検討したか	
G. AIシステム開発者は、データセットの設計にあたり、特定の社会属性に基づく不当な差別を維持・助長しないよう配慮したか	例：AIシステム開発者は、現実の再現性を優先することで、結果的に、特定の社会属性に基づく不当な差別までも再現してしまう可能性を検討したか 例：AIシステム開発者は、現実の再現性を犠牲にしてでも、特定の社会属性に基づく不当な差別を維持・助長するようなデータセットを用いないようにしたか 例：AI システム開発者は、特定の社会属性に基づく不当な差別を維持・助長しないようするために、可能な限り、データセットの設計に関わるチームの多様性を高めたか 例：AI システム開発者は、特定の社会属性に基づく不当な差別を維持・助長しないようするために、他国・地域に AIシステムを提供することを想定している場合には、その国・地域の規制、慣習、商慣行、文化等を考慮したか	

2-2. モデル・システム		
A. AI システム開発者は、開発しようとしているAIシステムに求められる十分な精度を確保したか	例：AI システム開発者は、AI システムの精度を評価するための指標を定義し、AI システムの運用時を想定して AI システムの精度を評価したか 例：AI システム開発者は、AI システムの運用時に想定される典型的な入力だけではなく、出現頻度の低い入力に対しても AI システムの挙動が安定的であることを評価したか 例：AI システム開発者は、要請に応じて AI システム運用者及び利用者に提供できるようにするために評価結果を記録したか	
B. AI システム開発者は、開発しようとしているAIシステムに求められる十分な堅牢性を確保したか	例：AI システム開発者は、外れ値や欠損を含む入力、不連続な入力などの異常入力に対しても AI システムの挙動が安定的であることを確認したか 例：AI システム開発者は、ノイズデータを調製し、ノイズを含むデータが入力された場合の AI システムの挙動が許容範囲であることを確認したか	
C. AI システム開発者は、開発しようとしているAIシステムの公平性を確保したか	例：AI システム開発者は、可能な限り多様性を高めたチームで、特定の社会属性に基づく不当な差別を維持・助長するような出力になっていないか否かを評価したか 例：AI システム開発者は、特定の社会属性に基づく不当な差別を維持・助長するような出力になっていないか否かについて、AI システム運用者とも議論し、評価したか 例：AI システム開発者は、複数の属性を独立して変化させ、出力への影響度や感度などを評価したか 例：AI システム開発者は、公平性の定義や指標の提案を調	

	査し、適当な場合には、公平性の定義や指標を定めて、公平性を客観的に評価したか	
D. AI システム開発者は、開発しようとしているAIシステムの妥当性を確保したか	例：AI システム開発者は、必要に応じて AI システムが運用・利用される分野の専門家の支援を得て、業界常識に照らして明らかに妥当でない出力が見られないことを確認したか 例：AI システム開発者は、AI システムの出力に対する寄与度の高いパラメータを割り出し、必要に応じてAI システムが運用・利用される分野の専門家の支援を得て、当該パラメータが業界常識に照らし合わせて妥当であることを確認したか 例：AI システム開発者は、複数の機械学習アルゴリズムを試行したり、1 種類であってもすでに妥当性が検証されている単純な汎用モデル等と比較したりしたか 例：AI システム開発者は、同じ機械学習アルゴリズムを使って同じデータで複数回学習を行った結果、それらの推定結果に説明できないような差異がないことを確認するなど、学習の再現性を確認したか 例：AI システム開発者は、開発中の AI システムと過去に妥当と判断された AI システムとを比較できるようにするために、AI システム開発に関する情報を管理しているか 例：AI システム開発者は、開発中の AI システムの開発過程の妥当性を検証できるように、開発プロセスのトレーサビリティを確保しているか	
E. AI システム開発者は、開発しようとしているAIシステムの説明可能性に配慮したか	例：AI システム開発者は、全ての出力について、人間が理解できるような一定の説明を加えることができるか否か確認したか	

	例：AI システム開発者は、AI システムについて一定の説明を加えることができる範囲を特定し、AI システム運用者にその範囲を伝えたか 例：AI システム開発者は、精度を犠牲にして説明可能性を高める選択肢を、AI システムの開発を依頼した AI システム運用者に提示したか	
F. AI システム開発者は、AI モデルおよび AI システムの管理方法を定めたか	例：AI システム開発者は、AI モデルが更新される頻度・方法について明確にしたか 例：AI システム開発者は、モデル更新時に更新前のモデルと比較して大幅に悪化していないことを確認したか 例：AI システム開発者は、過去の AI モデルも含めて保存し、問題が起きたときに前の AI モデルに戻す仕組みを整えたか 例：AI システム開発者は、AI モデル作成時のデータを保存し、後で再作成ができるようになっているか	

3. 運用・モニタリング		
評価項目例	具体的な項目例	
A. AI システム運用者は、潜在的な利用者のリテラシーや経験不足に対処したか	例：AI システム運用者は、AI システム開発者が利用者の理解を高めるためにまとめた AI システム運用者への注意事項に関する理解を徹底したか 例：AI システム運用者は、上記注意事項を理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか 例：AI システム運用者は、AI システム開発者が利用者の理解を高めるためにまとめた AI システムの限界に関する理解	

	を徹底するとともに、AI システム利用者にわかりやすく伝達したか	
B. AI システム運用者は、予見可能な悪用に関する課題に対処したか	例：AI システム運用者は、開発しようとしている AI システムに関する予見可能な悪用を設計によって排除できない場合に対処すべく、AI システム開発者が AI システム運用者のためにまとめた注意事項の理解を徹底したか 例：AI システム運用者は、AI システムが受け渡された時に特定されていた以外の目的への利用が禁止されているか否かを確認したか。	
C. AI システム運用者は、自らのリテラシーや経験不足の課題に対処したか	例：AI システム運用者は、AI システム開発者が AI システム運用者の理解を高めるためにまとめた AI システム運用者への注意事項の理解を徹底したか 例：AI システム運用者は、上記注意事項を理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか	
D. AI システム運用者は、AI システム開発者による AI システムの身体、精神、財産等への悪影響に関する課題への対処の内容を理解したか	例：AI システム運用者は、身体、精神、財産等への悪影響に関し、AI システムの残存リスクと残存リスクの管理方法を AI システム開発者から聴取、理解し、理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか 例：AI システム運用者は、入力データに異常値が含まれた場合に生じうる身体、精神、財産等への悪影響への対応策を AI システム開発者から聴取し、理解し、理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか 例：AI システム運用者は、入力データに悪意のあるデータや敵対的データが含まれた場合に生じうる身体、精神、財産等への悪影響への対応策を AI システム開発者から聴取し、理解し、理解できない場合に、AI システム開発者に問い合	

	わせ、疑問を解消したか 例：AI システム運用者は、出力データに異常値が含まれた場合に生じうる身体、精神、財産等への悪影響への対応策を AI システム開発者から聴取し、理解し、理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか	
E. AI システム運用者は、AI システム開発者による AI システムの公平性に関する課題への対処の内容を理解したか	例：AI システム運用者は、AI システム開発者が公平性に関して提案されている指標から適切なものを選択した場合であって、許容される範囲を超えた出力データに対する警告を発する措置を講じた場合に、その警告の意味を理解し、理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか 例：AI システム運用者は、公平性に関して AI システム開発者がまとめた注意事項を理解し、理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか	
F. AI システム運用者は、AI システム開発者による個人への配慮事項への対処の内容を理解したか	例：AI システム運用者は、個人を分析する AI システムの運用にあたって、AI システム開発者が実施した、個人の機会や意思決定を不当に害しないこと確保するための対応策を AI システム開発者から聴取し、理解し、理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか 例：AI システム運用者は、匿名化されたデータ等から個人を特定するような分析が行われているか否かを AI システム開発者に確認したか	
G. AI システム運用者は、AI システム開発者による AI システムのサイバーセキュリティに関する課題への対処の内容を理解したか	例：AI システム運用者は、不正なアクセスやデータの入力に対処する措置について AI システム開発者がまとめた AI システム運用者への注意事項を理解し、理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか	

H. AI システム運用者は、AI システムにおける人間の主体的な関与の機会を理解し、そのような機会を AI システム利用者に適切に提供しているか	例：AI システム運用者は、身体、精神、財産等への悪影響や公平性などの観点から、AI システム開発者が採用した人間の主体的な関与の機会（たとえば、AI システムの採否や利用の中断・停止を決定する自由や機会）や AI システム利用者が意思決定に際して AI システムに過剰に依存しないような設計について、その採用理由も含めて内容を理解し、理解できない場合に、AI システム開発者に問い合わせ、疑問を解消したか 例：AI システム運用者は、AI システム利用者に対して人間の主体的な関与の機会を適切に提供しているか	
I. AI システム運用者は、AI システムの機能、効果について理解しているか	例：AI システム運用者は、AI システム開発者との対話を通じて、運用しようとしている AI システムの機能、効果に関する理解を深めたか 例：AI システム運用者は、説明可能性等を高めるために精度を犠牲にしている場合があること（精度と説明可能性等の AI システムのトレードオフ）を理解しているか	
J. AI システム運用者は、AI システム運用時のモニタリング支援機能や AI モデルや AI システムの管理方法を理解して、適切に運用しているか	例：AI システム運用者は、入力・出力データのログ取得機能、AI システムの動作状況をまとめた一覧表示（AI システムの運用担当者、運用時間、入力・出力ログなど）などのモニタリング支援機能を理解しているか 例：AI システム運用者は、経営層や社外の要請を受け、適当な時間で運用状況を整理できる程度まで、モニタリング支援機能を使えているか 例：AI システム運用者は、入力・出力データのログ、AI システムの動作状況をまとめた一覧表示（AI システムの運用担当者、運用時間、入力・出力ログなど）を定期的に確認しているか	

	例：AI システム運用者は、モデルの性能変化や運用環境におけるデータ分布の変化など、再学習の必要性を定期的に確認しているか 例：AI システム運用者は、AI モデルが更新される頻度・方法など、AI モデルや AI システムの管理方法について理解しているか	
K. AI システム運用者は、適法、公正、一般的に妥当な方法でデータを取得・管理しているか	例：AI システム運用者は、AI システムの利用に必要なデータのみを取得・管理しているか 例：AI システム運用者は、AI システムの利用に必要なデータを取得・管理するにあたり、関連する法令、ガイドライン、業界における標準的な手法の有無及び内容を確認し、存在する場合には、関連する法令に従い、ガイドラインや標準的な手法を尊重したか 例：AI システム運用者は、AI システムの利用に必要なデータの入力の可否についてデータの主体の選択の余地が事実上ない状態を作り出さないように配慮したか 例：AI システム運用者は、データの管理に必要な情報を記録しているか 例：AI システム運用者は、一部のデータの利用を停止したり、削除したりできる形式でデータを管理しているか 例：AI システム運用者は、管理しているデータの漏洩や改ざんを防止するための措置を講じたか 例：AI システム運用者は、管理しているデータへの権限のない者によるアクセスを監視するために、アクセスログを取得しているか	

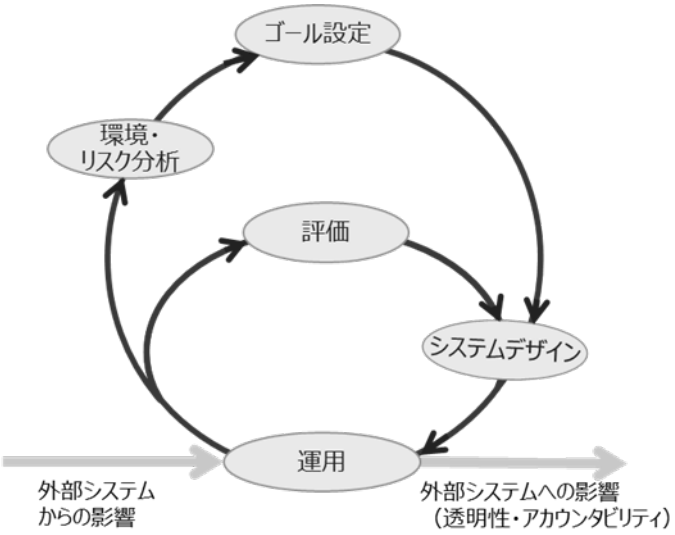
	例：AI システム運用者は、データの取得、データへのアクセスの監視などのデータの管理に関する状況を説明できるようにしているか 例：AI システム運用者は、AI システムへの入力データに関する疑問やトラブルが発生した際に対応する窓口等を設置しているか	
L. AI システム運用者は、AI システム利用者に対する説明責任を果たしているか	例：AI システム運用者は、AI システム開発者が特定した一定の説明を加えることができる範囲内で運用しているか 例：AI システム運用者は、AI システム開発者の支援を仰ぎながら、全ての出力について、要請があった場合に、人間が理解できるような一定の説明を加えることができることを確認したか	
M. AI システム運用者は、人的資源や運用体制を含め、AI システムの運用方法を明確に定めたか	例：AI システム運用者は、AI システムの振る舞いに問題を検知した場合に、リスクに応じた早さで適切な対応ができるような体制を確保しているか 例：AI システム運用者は、AI システムの振る舞いに問題があった場合に、AI を使わないプロセスに変更するなど、リスク回避の仕組みを適切に活用できているか	

H. 別添３（補論：アジャイル・ガバナンスの実践）

冒頭に紹介したとおり本ガイドラインの C.1.-6.の構成は、経済産業省の「Governance Innovation Ver.2」報告書[25]の中で整理した「アジャイル・ガバナンス」のフレームワークに準拠している。本補論では、その「アジャイル・ガバナンス」のコンセプトの背景について解説する。

AI システムに代表される、サイバー空間とフィジカル空間を高度に融合させるシステム（CPS：サイバー・フィジカルシステム）を基盤とする社会は、複雑で変化が速く、予見可能性に欠け、リスクの統制が困難である場合が多い。また、そうした社会の変化に応じて、ガバナンスが目指すゴールも常に変化していく。したがって、CPS を基盤とする社会のガバナンスモデルは、常に変化する環境とゴールを踏まえ、最適な解決策を見直し続けるものである必要がある。そのためには、ゴールや手段が予め設定されている固定的なガバナンスモデルを適用することは、妥当ではないと考えられる。最適な解決策を見直し続けるガバナンスのモデルとなる枠組が、「アジャイル・ガバナンス」である。

【アジャイル・ガバナンスの基本的なモデル】



このガバナンスモデルは、以下のような特徴を有する。

[25] パブリックコメントを反映した「GOVERNANCE INNVOATION Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」は、2021 年 7 月中に公表される予定。

各ガバナンスの主体（企業、政府、NGO 等、ガバナンスを担う様々な主体が含まれる。）は、まず、以下のようなプロセスを実施することが求められる。

① 環境・リスク分析

ガバナンスの主体は、常に外部環境及びその変化と、これに基づくリスク状況を分析し続ける必要がある。

② ゴール設定

ガバナンスの主体は、外部環境の変化や技術の与える影響の変化に伴い、ガバナンスの「ゴール」を設定し、常時見直すべきである。

③ ガバナンスシステムのデザイン

ガバナンスの主体は、設定されたゴールに基づいて、ガバナンスシステムのデザインを行う。ここでの「システム」とは、技術的なシステムだけではなく、組織のシステムやこれに適用されるルールを含む。そのデザインあたっては、(i) 透明性とアカウンタビリティ、(ii) 適切な質と量の選択肢の確保、(iii) ステークホルダーの参加、(iv) インクルーシブネス、(v) 適切な責任分配、(vi) 救済手段の確保、といった要素が、ガバナンスシステムをデザインする上で尊重されるべき基本原則となる。

④ ガバナンスシステムの運用

デザインされたガバナンスシステムを運用するプロセスである。ガバナンスの主体は、システム運用の状況について、リアルタイムデータ等を使って継続的にモニタリングしていくことが求められる。また、影響を受けるステークホルダーに対して、自らのシステムのゴール、それを達成するためのシステムのデザイン、そこから生じるリスク、運用体制、運用結果、救済措置等について、適切な開示を行うことが不可欠である。

こうした運用の過程・結果を踏まえて、ガバナンスの主体は、以下の 2 つの評価・分析をいずれも実施する必要がある。

⑤ ガバナンスシステムの評価

ガバナンスの主体は、当初設定されたゴールが達成されているかを評価する。設定したゴールが達成されていなければ、再度システムデザインを行う（下側の楕円型のサイクル）。

⑥ 環境・リスクの再分析

第 2 に、外部システムからの影響によって、ガバナンスのゴール自体を見直さなければならない可能性がある（外側の円形のサイクル）。そのため、ガバナンスシステムの置かれた環境やリスク状況に変化があるか、これによってゴールを変更する必要があるか、という点を継続的に分析する必要がある。

「Governance Innovation Ver.2」報告書では、**アジャイル・ガバナンスの実践に向けた企業の取組を後押しするために、標準やガイドラインといったソフトローによって、官民共同で政策ツールを策定していくことが重要**であると述べている（4.3.3）。本ガイドラインは、AI のガバナンスという局面で、まさにそのような企業の取組みを後押しするツールとしての役割を果たすものであるといえる。さらに、**本ガイドライン自体もアジャイル・ガバナンスのプロセスに則って継続的に評価・見直し・アップデートされるべき**であり、**AI 技術の発展や AI に対する社会的な受容の変化などを適時に反映した官民共同の Living Document として、継続的にメンテナンスされ、参照され続けていくことが望ましい**。