NEWSLETTER

AI 事業者ガイドライン（第 1.0 版）の公表と今後の実務対応

ロボット/AI ニューズレター　　2024 年 4 月 26 日号

執筆者:
石川 智也
n.ishikawa@nishimura.com
松下 外
g.matsushita@nishimura.com
角田 龍哉
t.tsunoda@nishimura.com
水井 大
d.mizui@nishimura.com

2024 年 4 月 19 日、総務省及び経済産業省から「AI 事業者ガイドライン（第 1.0 版）」（以下「**AI 事業者 GL**」という）の最終版が公表された。AI 事業者 GL は（法的拘束力は有さないとしても）今後の日本における AI ガバナンスや AI 関連規制の方向性を示すものとして実務上幅広く参照されることが想定される。そこで、本ニューズレターでは、AI 事業者 GL の全体像を概観しつつ、事業者ごとに検討・実践すべき要点を紹介する。

1. AI 事業者 GL 策定に至る経緯と国内外の動向

日本では従前より、AI、ビッグデータ、IoT、5G などのサイバー空間とフィジカル空間とを融合した社会である「Society5.0」に対応するため、AI の利活用に関連し、次の各ガイドラインが整備されていたが、従前より複数のガイドラインへのアクセスの悪さや棲み分けの必要性が説かれていた。

- 総務省「国際的な議論のための AI 開発ガイドライン案」（2017 年）
- 総務省「AI 利活用ガイドライン～AI 利活用のためのプラクティカルリファレンス～」（2019 年）
- 経済産業省「AI 原則実践のためのガバナンス・ガイドライン ver.1.1」（2022 年）

そうしたところ、2022 年末頃から、米 Open AI の ChatGPT に代表されるように、生成 AI に関する LLM 技術の非連続的変化によって、画像、文章、音声、動画に至るまで生成 AI の利活用が急速に進んだのは記憶に新しい。そしてその反面、AI 利活用に伴うリスクがより広く認知され、リスクの内容それ自体はもとよりその影響度も予測しづらい面が生じたことも周知のとおりである[1]。

このような生成 AI の急速な発展と普及、そして新たなリスクの出現によって、2023 年 5 月に開催された G7 広島サミットでは、AI のガバナンスと相互運用性に関する国際的な議論を推進するための枠組みである「広島 AI プロセス」が立ち上がり、「全 AI 関係者向けの広島プロセス国際指針」及び「高度な AI システムを開発する組織向けの広島プロセス国際行動規範」を含む「広島 AI プロセス包括的政策枠組み」が策定された。国内でも、2023 年 5 月 26 日に AI 戦略会議において公開された「AI に関する暫定的な論点整理」において

[1] 例えば、2023 年のインターネットバンキングの不正送金被害について、「令和 5 年 2 月以降、再度被害が急増」しており、前年の 15.2 億円比で、令和 5 年度は 80.1 億円と、約 5 倍強に増加した（金融庁ホームページ「フィッシングによるものとみられるインターネットバンキングによる預金の不正送金被害が急増しています。」令和 6 年 1 月 24 日最終更新）。2023 年 2 月にリリースされたテキスト生成 AI の ChatGPT Plus（GPT-3.5）を境に日本語の壁が無くなったとされることと決して無関係ではなく、日本人にも違和感のないフィッシング用のおとりメールやショートメッセージも生成可能となってしまったことが大きな要因ではないかと考えられる。

「生成 AI の普及を踏まえ、既存のガイドラインに関して必要な改訂などを検討する必要がある」ことが確認された。

国外に目を向けると、AI 利活用に関するルール整備が急ピッチで進められている。EU では、2021 年 4 月に欧州委員会により欧州 AI 法案が提案され、2024 年 3 月 13 日には欧州議会で承認されるに至り（原文。なお、以下「**欧州 AI 法案**」という場合、2024 年 3 月 13 日版を指す）、その最終的な成立が近づきつつある。また、米国では、2020 年 10 月に「AI 権利章典の青写真」が公表され、2023 年 7 月及び 10 月には、AI 関連事業者が AI の安全な開発等に関して「自主的なコミットメント」を約束し、同年 10 月には各連邦政府機関に向けた「AI の安全、安心、信頼できる開発と利用に関する大統領令」が発令されるに至った。

以上の背景の下、内閣府による「人間中心の AI 社会原則」（2019 年）を踏襲しつつ、諸外国の動向や新技術の台頭を踏まえ、上記 3 つの既存のガイドラインを統合し見直す形で策定されたのが AI 事業者 GL である。

満を持して、AI 事業者 GL の素案である「AI 事業者ガイドライン案」は 2024 年 1 月 19 日に公開され、同年 2 月 19 日までパブリックコメントに付された。約 4,000 件の意見が寄せられた後、その結果を踏まえ、2024 年 3 月 14 日には「AI 事業者ガイドライン第 1.0 版（案）」が公開され、最終案である AI 事業者 GL が公表されるに至った（以下、特段の断りがない限り、頁数の表記は最終案の AI 事業者 GL 本編によるものである。）。

2. ソフトローとしての AI 事業者 GL

AI 事業者 GL は、①事業者の自主的な取組の支援、②国際的な議論との協調、③読み手にとっての分かりやすさを基本的な考え方として位置づけている。上記のとおり AI システムの高度化、更に AI をめぐる動向が目まぐるしく変化していることを踏まえ、教育・研究機関、一般消費者を含む市民社会、民間企業等で構成される「マルチステークホルダー」の関与の下、その内容の更新が、継続的に検討される「Living Document」として位置づけられている（3 頁）。マルチステークホルダーからどのようにして意見を吸い上げ、反映していくかについては、総務省の AI ネットワーク社会推進会議と経済産業省の AI 事業者ガイドライン検討会にて取りまとめを行うものとされた（3 頁注 1）。

こうした基本的な考え方は、Society5.0 の国家ビジョンを背景とし、日本が、従前より、拘束的なハードローではなく非拘束的なソフトローによって柔軟に各事業者による AI の社会実装を導く方針を採用してきた事実を再確認するものである。一方、リスクの性質や程度を踏まえ、長期的に実現されるべき価値観や法執行の確実性の担保が必要な場面ではハードローを活用する（新たな法令の策定のみならず既存の法令の明確化や見直しも含む）等、ハードローとソフトローをその性質に応じて適切に使い分けることがより効果的な場面も想定される。

この観点からは、2024 年 2 月 16 日に自民党 AI の進化と実装に関するプロジェクトチーム WG 有志が公表した「責任ある AI 推進基本法（仮）」は、社会的影響力の大きな基盤モデルについて、ソフトローのみならず、制裁を伴うハードローの活用も踏まえた AI 規制の在り方を模索するものである。また、この AI 推進基本法では、技術の進歩に遅れないようにするために、体制整備に求められる具体的な基準については民間の

関与を求める共同規制のアプローチが示唆されていることも興味深い。

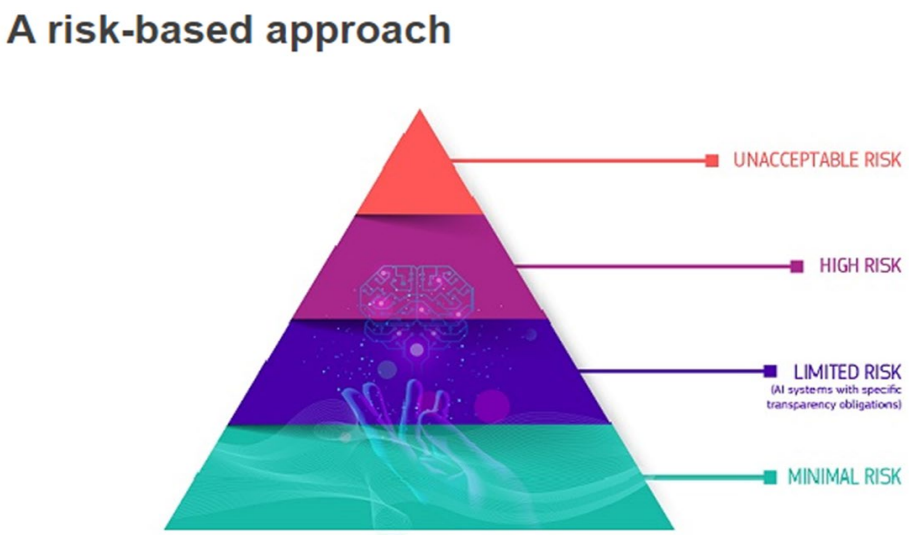
3. リスクベースアプローチと実践的対応

（1） AI 利活用のためのフレームワークとしての「リスクベースアプローチ」

AI 事業者 GL では、「本ガイドラインでは、この『リスクベースアプローチ』にもとづく企業における対策の方向を記載している」とされ、AI 活用のフレームワークとして、「リスクベースアプローチ」が採用されることが明記された（3 頁）。AI 事業者 GL が「リスクベースアプローチにもとづく AI ガバナンスの実践」の視点から整理されている点は重要な視点になる。

この点、2024 年 3 月 13 日に欧州議会で承認された欧州 AI 法案では、AI 利用に伴うリスクを 4 段階に定形的にカテゴライズしており、AI のリスクの程度に応じて「許容できないリスク（Unacceptable Risk）」「ハイリスク（High Risk）」「限定リスク（Limited Risk）」「最小リスク（Minimal Risk）」に区分し、それぞれ異なる規制を課す「リスクベースアプローチ」が採用されている（【図 1】）[2]。

【図 1】



（出所）
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

一方の AI 事業者 GL における「リスクベースアプローチ」は、「予め事前に当該利用分野における利用形態に伴って生じうるリスクの大きさ（危害の大きさ及びその蓋然性）を把握したうえで、その対策の程度をリスクの大きさに対応させる」ものであり（3 頁）、欧州 AI 法案が採用するようなリスク（そして行うべき対応）のカテゴライズはされていない。そのため、各事業者が柔軟かつ主体的に、個別の事案における調整・判断を（少なくともその構造上は）可能としたうえで、AI のリスクを特定・評価し、そのリスクを効果的に

[2] 生成 AI との関連では追加的に「汎用目的 AI モデル」及び「システミックリスクを伴う汎用目的 AI モデル」との区分も導入されている。

低減するための対応を講じるフレームワークとなっている。AI の利活用にあたって、非定型的なリスクを前提とするリスクベースアプローチのフレームワークを採用するからこそ、外部環境を踏まえたリスク評価とその見直しを実践するための体制ないしその運用（AI ガバナンス）がより重要になる。

（2）AI ガバナンス

AI ガバナンスは、AI 事業者 GL では以下のとおり定義されている（9 頁）。

> AI の利活用によって生じるリスク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（便益）を最大化することを目的とする、ステークホルダーによる技術的、組織的、及び社会的システムの設計並びに運用。

すなわち、AI ガバナンスは「ステークホルダーによる技術的、組織的、及び社会的システムの設計並びに運用」等の AI の利活用に向けた具体的な実践と位置づけられており、「リスクベースアプローチ」はその具体的な内容を決定するための基本的な考え方（フレームワーク）として位置づけられうる。

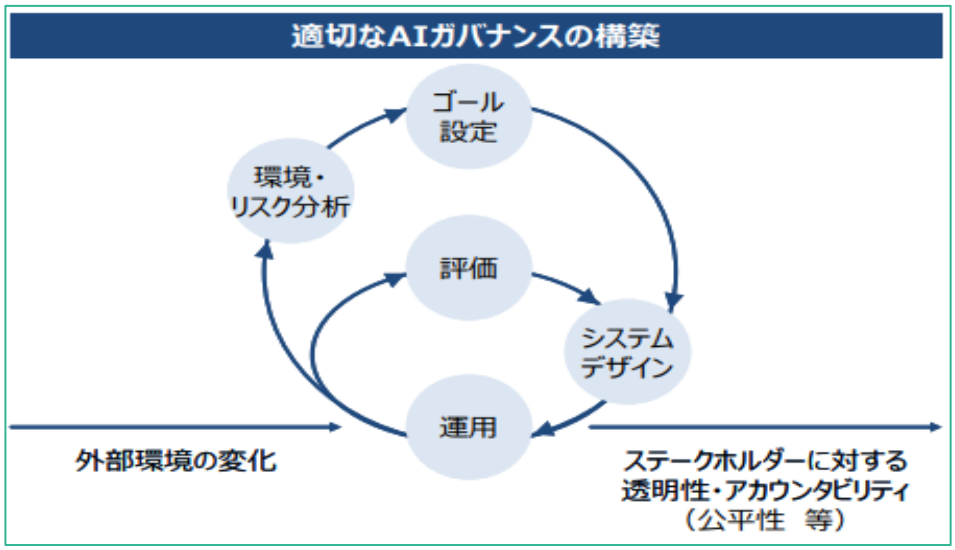
AI 事業者 GL は、「各主体間で連携しバリューチェーン全体で『共通の指針』を実践し AI を安全安心に活用していくためには、AI に関するリスクをステークホルダーにとって受容可能な水準で管理しつつ、そこからもたらされる便益を最大化するための、AI ガバナンスの構築が重要」と述べている（24 頁）。すなわち、リスクベースアプローチを中核的な概念としつつも、AI がもたらすリスクの面のみに着目するのではなく、その便益とのバランスも踏まえたうえで、AI ガバナンスを設計することが期待されている。

AI 事業者 GL は、AI ガバナンスを実践する具体的なアプローチとして、「アジャイル・ガバナンス」の方法を提案する（24 頁、【図 2】）。すなわち、事前にルール又は手続が固定されたいわゆる「伝統的なガバナンスモデル」ではなく、従前より、Society5.0 を実現していくための新たなガバナンスモデルとして説かれる「アジャイル・ガバナンス」を、AI のリスク管理の場面でも実践することを求めているものと思われる[3]。

> 「Society 5.0」を実現するためには、サイバー空間とフィジカル空間を高度に融合させたシステム（CPS）の社会実装を進めつつ、その適切な AI ガバナンスを構築することが不可欠である。CPS を基盤とする社会は、複雑で変化が速く、リスクの統制が困難であり、こうした社会の変化に応じて、AI ガバナンスが目指すゴールも常に変化していく。そのため、事前にルール又は手続が固定された AI ガバナンスではなく、企業・法規制・インフラ・市場・社会規範といった様々なガバナンスシステムにおいて、「環境・リスク分析」「ゴール設定」「システムデザイン」「運用」「評価」といったサイクルを、マルチステークホルダーで継続的かつ高速に回転させていく、「アジャイル・ガバナンス」の実践が重要となる

[3] 経済産業省「アジャイル・ガバナンスの概要と現状」報告書（2022 年 8 月）17 頁等

【図 2】



（出所）AI 事業者 GL25 頁

（3）　考えられる実践的対応

以上のとおり、AI 事業者 GL では、リスクベースアプローチのフレームワークを用いて、AI ガバナンスを実践するものとされた。では、事業者としては具体的に何を実践することが考えられるだろうか。

AI 事業者 GL は、AI ガバナンスのゴールのひとつとして「AI ポリシー」及び「データ活用ポリシー」を設定すること等がありうるとし（24 頁注 27）、またそれらを開示することも考えられる。この点、経済産業省から「『AI 事業者ガイドライン（第 1.0 版）』チェックリスト及び具体的なアプローチ検討のためのワークシート（別添 7A,7B,7C）」（以下「**AI 事業者 GL チェックリスト**」という）が公表されているため、「『AI 事業者ガイドライン（第 1.0 版）』主体横断的な仮想事例（別添 8）」を参照しながら、チェックリスト及びワークシートを作成し、必要に応じて全部又は一部を公表することが有用であろう。

本ガイドラインに記載の「共通の指針」への対応事項からなる自社の取組方針（「AI ポリシー」等、呼称は各主体により相違）及び「共通の指針」への対応事項を包含しつつそれ以外の要素を含む取組方針（データ活用ポリシー等）を設定すること等が考えられる。AI を活用することによって包摂性を向上させる等の便益を高めるための指針を提示してもよい。また、呼称も各主体に委ねられている。

なお、従前より、例えば AML/CFT（マネー・ローンダリング・テロ資金供与）リスクに関しては、規制に基づくリスクベースアプローチが採用されている。具体的には「特定事業者」（犯罪による収益の移転防止に関する法律 2 条各号）に該当する事業者であれば、特定事業者作成書面等（いわゆるリスク評価書）を作成したうえで、必要に応じて見直し、変更を加える義務（同法 11 条 4 号、同法施行規則 32 条 1 項 1 号）等が課されている。

AI における（ソフトローにおける）リスクベースアプローチを堅牢に実践するのであれば、こうした規制にもとづくリスクベースアプローチの取組みも必要に応じて参照しながら、「AI ポリシー」及び「データ活用

ポリシー」、更にはチェックリスト等の作成等の対応を行うことも有用であると思われる[4]。

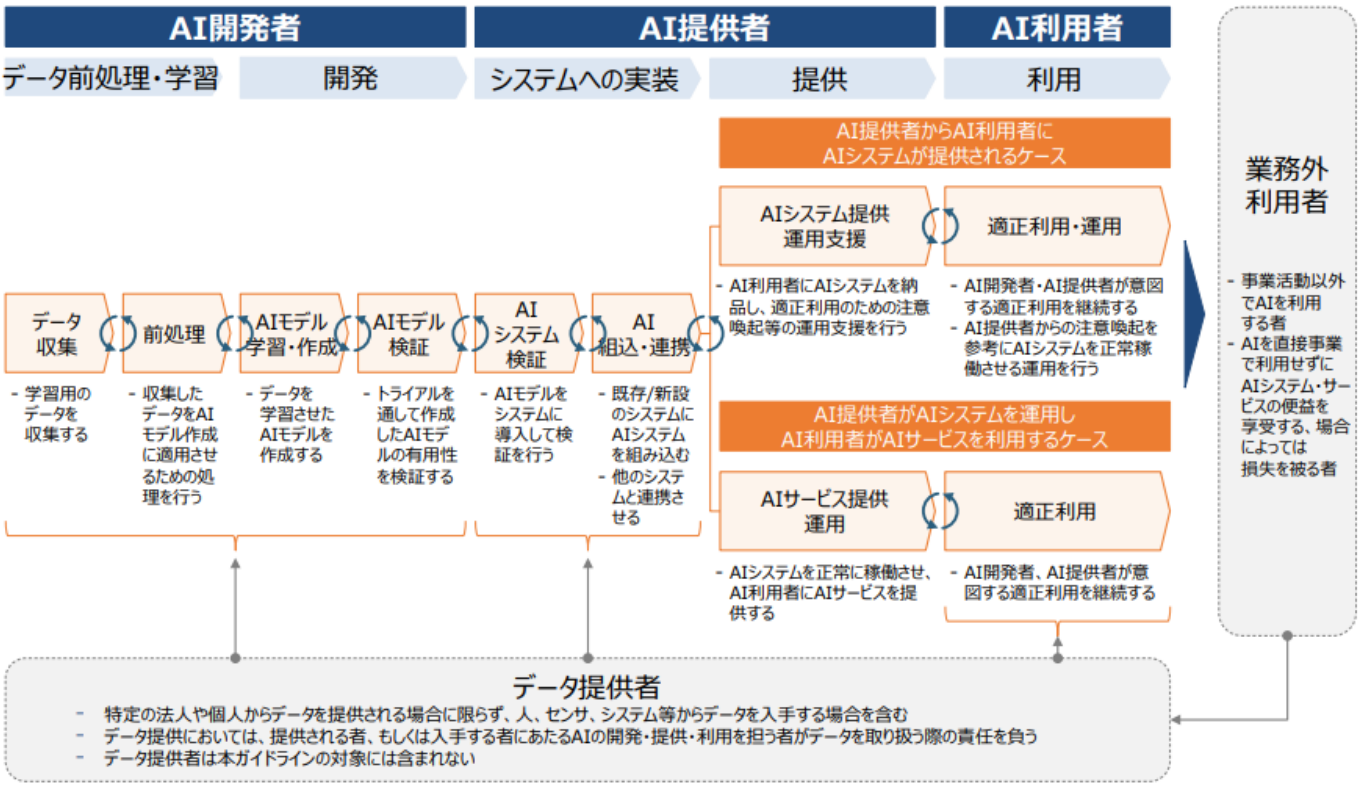
4. バリューチェーンの検討

AI 事業者は、上記 3（3）の対応を実践する場合に、どこまでの主体のリスクを考慮すればよいのか。

AI 事業者 GL は、AI の社会実装の過程として、①AI モデル→②AI システム→③AI サービスの流れを想定し、そのうえで実利用の場面として、②AI システムが利用されるケースと、③AI サービスが利用されるケースを前提に、AI 事業者 GL の対象事業者を整理している。大まかには、AI モデルの学習・生成に携わる事業者が「AI 開発者」であり、AI モデルを AI システムに組み込み AI システム又は AI サービスとして提供する事業者が「AI 提供者」、そしてこれらシステム又はサービスを利用する事業者が「AI 利用者」である（【図 3】）。AI の活用方法によっては同一の事業者が AI 開発者、AI 提供者、AI 利用者の複数を兼ねる場合もあるとされる（4 頁）。一方で、学習等に用いる「データ提供者」や、消費者等の事業者でない AI 利用者である「業務外利用者」は AI 事業者 GL の直接の規律対象ではない。ただし、データ収集に関しては、AI の開発・提供・利用を担う者がデータを取り扱う際の責任を負う形で記載されている。

[4] なお、金融分野においては、AML/CFT リスクにとどまらず金融庁「モデル・リスク管理に関する原則」（2021 年 11 月 12 日）に基づき、AI や関連する技術をめぐるモデル・リスクの管理にあたって、リスクベースアプローチのフレームワークを採用している（その意味で、AI 事業者 GL のうちリスクベースアプローチの考え方に関していえば、全く真新しいものではない）が、例えばこうした業界標準に沿った取組みが、AI 事業者 GL で推奨される取組みとの関係で、どのように評価されうるのかは課題である。また、AI 事業者 GL で推奨される取組みと既存の業法を含む規制との関係性も必ずしも判然としないが、個別の事案ごとに規制の適合性も検討するほかないと思われる。

【図 3】



（出所）AI 事業者 GL5 頁

このような整理は、AI 利活用ガイドラインを踏襲するものである。もっとも、AI 事業者 GL は、更に一歩進んで、AI 利活用に関する「バリューチェーン」「リスクチェーン」をより明確に想定し、リスクベースアプローチで考慮すべき主体の範囲として、「バリューチェーン/リスクチェーン全体での便益の確保、リスクの削減」も指向されている。

特に欧州 AI 法案では、AI システムの提供者（provider）や実装者（deployer）に加えて、提供者の正当な代理人（authorized representative)、製品製造者（product manufacturer）、輸入者（importer）、販売事業者（distributor）等のバリューチェーン上の利害関係者を事業者（operator）として総称し、その規律の対象としている。こうした国際的な規制動向に照らして、特定の企業だけでなく、バリューチェーンを意識しながらリスクベースアプローチを実践する必要もある。

更に関連して、AI 事業者 GL では、複数主体にまたがる重要な論点として、①主体間、又は主体内の責任分配、②AI システム・サービス全体の品質向上、③各 AI システム・サービスが相互に繋がることによる新たな価値の創出の可能性（System of Systems）及び④AI 利用者又は業務外利用者のリテラシー向上等が指摘されている。その中でも特に①主体間、又は主体内の責任分配に関しては、個別の事業者の利害関係が直接的に対立しうる側面があり、実務上必ずしもその解決が容易ではない。この点については、契約の文脈において、AI 事業者 GL で次のような重要な指摘がなされている（別添 156 頁）。

リスクを直接コントロールできない当事者によるリスクの引き受けという問題もある。･･･仮に、その提供や利用の方法にかかわらず、AI に起因する損害に対する責任の一切を AI 開発者が負うとすれば、自身では直接コントロールできないリスクも AI 開発者が引き受けることになる。このような事態は、交渉力に格差のある当事者間で起こりやすい一方で、影響力の大きい AI であれば、その AI 開発者がコントロールすべきリスクの範囲を広く捉えるべき場合もある。

更に近時では、AI の社会実装が国内のみで完結しない場面は少なくない。すなわち、バリューチェーン上の責任分配は、バリューチェーンが連なる限り、国内のみならず国際的な広がりをもち、各国の専門家を活用して適切な規制調査と契約上の対応を行うことが必要となる。AI の利活用にはデータが不可欠であることに照らせば、データの自由な流通（DFFT）の確保のための適切な制度設計や契約の枠組みもまた、益々重要度を増すだろう。

5. 説明責任（透明性・アカウンタビリティ）への対応

（1） 説明責任の内容

AI 事業者は、上記 3（3）の対応を実践する場合に、どの範囲のステークホルダーに対し情報提供を行い、そして具体的に何を開示すべきか。

この点、AI 事業者 GL では、各主体が連携してバリューチェーン全体で取り組むべき共通の指針として、幾つかの要素が挙げられており、その中でも実務上特に重要な概念は透明性及びアカウンタビリティである。これらの概念は相互互換的に用いられている。AI 事業者 GL では、透明性を「情報開示に関する事項」（16 頁注 20）、また「アカウンタビリティ」を「AI に関する事実上・法律上の責任を負うこと及びその責任を負うための前提条件の整備に関する概念」（18 頁注 21）と整理しており、これらを総称して「説明責任」と理解されている（35 頁）。

AI 事業者 GL では、透明性及びアカウンタビリティに関して、各事業者に共通して次の各対応が期待されている（AI 事業者 GL チェックリスト参照。なお、AI 開発者、AI 提供者及び AI 利用者のそれぞれについて、更に追加的な対応指針も示されている。）。大まかには、①ログの記録・保存、②ステークホルダーへの情報提供、③責任者の明示、④責任の明確化、⑤文書化等がその内容である。

【透明性】

① **検証可能性の確保**	AI の判断にかかわる検証可能性を確保するため、**データ量又はデータ内容に照らし合理的な範囲で、AI システム・サービスの開発過程、利用時の入出力等、AI の学習プロセス、推論過程、判断根拠等のログを記録・保存**しているか？
	ログの記録・保存にあたっては、**利用する技術の特性及び用途に照らして、事故の原因究明、再発防止策の検討、損害賠償責任要件の立証上の重要性等を踏まえて、記録方法、頻度、保存期間等について検討**しているか？

②	**関連するステークホルダーへの情報提供**	AI との関係の仕方、AI の性質、目的等に照らして、それぞれが有する知識及び能力に応じ、例えば、**下記について取りまとめた情報の提供及び説明**を行っているか？ ✓ AI を利用しているという事実及び活用している範囲 ✓ データ収集及びアノテーションの手法 ✓ 学習及び評価の手法 ✓ 基盤としている AI モデルに関する情報 ✓ AI システム・サービスの能力、限界及び提供先における適切/不適切な利用方法 ✓ AI システム・サービスの提供先、AI 利用者が所在する国・地域等において適用される関連法令等
		多様なステークホルダーとの対話を通じて積極的な関与を促し、社会的な影響及び安全性に関する様々な意見を収集しているか？
		実態に即して、AI システム・サービスを提供・利用することの優位性、それに伴うリスク等を関連するステークホルダーに示しているか？
③	**合理的かつ誠実な対応**	「関連するステークホルダーへの情報提供」は、**アルゴリズム又はソースコードの開示を想定するものではなく、プライバシー及び営業秘密を尊重して、採用する技術の特性及び用途に照らし、社会的合理性が認められる範囲で実施**しているか？
		公開されている技術を用いる際には、それぞれ定められている規程に準拠しているか？
		開発した **AI システムのオープンソース化にあたっても、社会的な影響を検討**しているか？
④	**関連するステークホルダーへの説明可能性・解釈可能性の向上**	関連するステークホルダーの納得感及び安心感の獲得、また、そのための AI の動作に対する証拠の提示等を目的として、説明する主体がどのような説明が求められるかを分析・把握できるよう、説明を受ける主体がどのような説明が必要かを共有し、必要な対応を講じているか？ ✓ AI 提供者：AI 開発者に、どのような説明が必要となるかを共有しているか？ ✓ AI 利用者：AI 開発者・AI 提供者に、どのような説明が必要となるかを共有しているか？

【アカウンタビリティ】

①	**トレーサビリティの向上**	データの出所、AI システム・サービスの開発・提供・利用中に行われた意思決定等について、**技術的に可能かつ合理的な範囲で追跡・遡求が可能な状態を確保**しているか？

②	**「共通の指針」の対応状況の説明**	「共通の指針」の対応状況について、ステークホルダー（サプライヤーを含む）に対してそれぞれが有する知識及び能力に応じ、**例えば以下の事項を取りまとめた情報の提供及び説明を定期的に行っている**か？ ✓ **全般：** 「共通の指針」の実践を妨げるリスクの有無及び程度に関する評価、「共通の指針」の実践の進捗状況 ✓ **「人間中心」関連：** 偽情報等への留意、多様性・包摂性、利用者支援、持続可能性の確保の対応状況 ✓ **「安全性」関連：** AIシステム・サービスに関する既知のリスク及び対応策、並びに安全性確保の仕組み ✓ **「公平性」関連：** AI モデルを構成する各技術要素（学習データ、AI モデルの学習過程、AI 利用者又は業務外利用者が入力すると想定するプロンプト、AI モデルの推論時に参照する情報、連携する外部サービス等）によってバイアスが含まれうること ✓ **「プライバシー保護」関連：** AI システム・サービスにより自己又はステークホルダーのプライバシーが侵害されるリスク及び対応策、並びにプライバシー侵害が発生した場合に講ずることが期待される措置 ✓ **「セキュリティ確保」関連：** AIシステム・サービスの相互間連携又は他システムとの連携が発生する場合、その促進のために必要な標準準拠等。AIシステム・サービスがインターネットを通じて他の AI システム・サービス等と連携する場合に発生しうるリスク及びその対応策
③	**責任者の明示**	各主体において**アカウンタビリティを果たす責任者を設定**しているか？
④	**関係者間の責任の分配**	関係者間の責任について、**業務外利用者も含めた各主体間の契約、社会的な約束（ボランタリーコミットメント）等により、責任の所在を明確化**しているか？
⑤	**ステークホルダーへの具体的な対応**	必要に応じ、AIシステム・サービスの利用に伴うリスク管理、安全性確保のための**各主体の AI ガバナンスに関するポリシー、プライバシーポリシー等の方針を策定し、公表**しているか？（社会及び一般市民に対するビジョンの共有、情報発信・提供を行うといった社会的責任を含む）
		必要に応じ、**AI の出力の誤り等について、ステークホルダーからの指摘を受け付ける機会を設けるとともに、客観的なモニタリングを実施**しているか？
		ステークホルダーの利益を損なう事態が生じた場合、どのように対応するか方針を策定してこれを着実に実施し、進捗状況については必要に応じて**定期的にステークホルダーに報告**しているか？
⑥	**文書化**	**上記に関する情報を文書化して一定期間保管し、必要なときに、必要なところで、入手可能かつ利用に適した形で参照可能な状態**としているか？

（2）　考えられる実践的対応

以上を踏まえても、どの範囲のステークホルダーに対し情報提供を行い、そして具体的に何を開示すべきかという説明責任は、個社ごとにリスクベースアプローチに従って決定されるべきものであるとの方針は理解できるが、あるべき/望ましい具体的な例示等の実例が示されなかったこともあり、実際のところ悩ましい。

まず、AI 事業者 GL で情報開示の対象とされる「ステークホルダー」は「AI 開発者、AI 提供者、AI 利用者、業務外利用者以外の第三者を含む AI の活用によって直接・間接の影響を受ける可能性がある全ての主体」を指す（11 頁注 9）。

次いで、具体的に開示が必要な情報に関しては、AI システム・サービスに関する事項と、共通の指針の履践状況が挙げられており、各主体の AI ガバナンスに関するポリシー、プライバシーポリシー等の方針の策定・公表や客観的なモニタリング、定期的な報告等が求められている。もっとも、透明性・アカウンタビリティの履践に際しては、特に開発側から懸念が挙げられることが少なくなかったアルゴリズム又はソースコードの開示が求められるものではないことや、情報提供の範囲も社会的に合理的な範囲に留まる点を明記する等の対応はなされているものの、具体的なタスクが見えにくいかもしれない。

結局のところ、一体、何を目的として透明性やアカウンタビリティが要求されているのか十分に検討しなければ、その適切な実施は困難である。これらの履践が要求される一因には、AI がデータにもとづく帰納的な工程により生成されることに起因しており、AI に入力されたデータから如何なる過程を経て出力（AI 生成物）が生成されたのかを説明することの技術的な困難性（AI のブラックボックス性と呼ばれることもある）が指摘できる。このような困難性に対し何らかの手当がなされなければ、①ステークホルダーが AI の利活用の適否を判断する際に十分な判断根拠を欠くことになり、また、②何らかの事故が生じた際の責任の所在の確定等が困難となる等、その利活用の前後にリスク要因が残り、結果として AI の利活用が忌避される状況が生じうる。

上記の観点から整理を試みると、少なくとも①については、このような AI の不確実性を低減するに足る情報が提供される必要があるだろう。如何なる情報提供・開示が必要であるかは具体的な事案次第であるが、法的・社会的な受容可能性があることや、ステークホルダーがこれを利用するか否かを判断するに足る程度の判断の不確実性等の技術的な制約に関する情報は、情報提供・開示の中核的な内容を構成すると思われる。

他方、②の観点については、当事者間の契約の文脈であるが、AI 事業者 GL の以下の指摘は示唆に富む（別添 156、157 頁）。AI システムやサービスの出力の理由を説明することが困難であるとしても、そのような困難性を前提に自らが如何に合理的な対応を取っていたかを立証できることが重要になってくると考えられる。

> AI の普及や応用が進むにつれ、AI の開発・提供・利用に伴うリスクが増えるとともに、そうしたリスクが顕在化するケースも今後増えていくことが予想される。
> そうしたリスクの中に、AI を組み込んだソフトウェアや、これを利用したサービスに関連して何らかの事故が起こり、それにより AI の開発・提供・利用の当事者や第三者に損害を生じさせるリスクがある。このケー

スで問題となるのは、事故を回避するために尽くすべき注意を尽くしていたかや、事故がそもそも AI に起因していたかの証明に関する問題がある。AI の開発・提供・利用の当事者には、それぞれのプロセスにどのような関与を行ったかについて、合理的な説明を行うことが求められる可能性がある。
こうした説明に対する責任は、AI の開発・提供・利用のすべての当事者の間でどのような契約が締結されていたとしても、事故について一次的な責任を負う当事者に発生する可能性があるものである。契約で定めることができるのは、契約の当事者限りでの責任の分担に限定される。契約の当事者以外の者により責任追及をされた場合に、AI のバリューチェーンに連なる者はすべて、一定の説明を求められる立場に立つ可能性がある。説明の合理性との関係で問題となるのは、説明の内容に加えてその客観的な根拠であり、AI の開発・提供・利用に関する契約を締結の前後で、そうした根拠を整理しておくことが期待される。

AI 事業者 GL は、事業者による自主的な情報提供を前提にしている。もっとも、欧州 AI 法案では、高リスク AI システムの提供者については実装者がその機能を十分理解できるような一定の情報の開示等が義務づけられているし、汎用目的 AI についても学習に用いられたコンテンツ等に関する要約の公開のほか、AI オフィス等の規制当局や実装者それぞれに対するきめ細かな情報開示が求められている。

特にグローバルに展開する企業は今後欧州 AI 法案に沿った情報開示を検討・実施する可能性が高く、結果として欧州 AI 法案が定立する実務上の取扱いが、わが国でも一種のデファクトスタンダードとして流入していく事態も十分にありうる。上記 AI 事業者 GL に照らして優先度が高いと思われる情報の項目のほか、欧州 AI 法案上の情報開示の項目にも目配せをして、必要十分な透明性の確保と、これを通じた AI サービスに対する市場、市民社会等からの信頼を得ていくことが有用だろう。とりわけ透明性・アカウンタビリティについては、国内の実務に留まらず、国際的なスタンダードを視野に入れてその対応を決することが重要になり、また、AI 活用を推進する企業経営層にも幅広い視点が求められる。

以上

西村あさひ　広報課 newsletter@nishimura.com