信託セミナー

生成 AI と金融機関の AI ガバナンス

KPMG コンサルティング株式会社プリンシパル　津　田　圭　司

— 目　次 —

はじめに

私からは AI ガバナンスについて説明します。生成 AI の登場により、AI のリスクの性質が少し変わってきており、AI ガバナンスの必要性が叫ばれているところです。この動向についてお話をさせていただきます。

まず、生成 AI 登場によって AI リスクがどのように変わっているかを説明し、各国の AI 規制の動向についてもお話しします。その上で、金融機関に求められる AI ガバナンスもしくはリスクマネジメントについて説明させていただきます。

1．AI リスクの変遷

⑴　AI の利活用拡大と生成 AI の特徴

生成 AI の登場によって AI の世界はずいぶん広がりましたが、まずは従来型の AI との違いを説明します。この点を認識いただくことにより、最近なぜ AI リスクが特に話題になっているのかということについて、理解が深まると考えています。

違いとして最も大きいのは、利用の目的です。資料５頁に記載のとおり、従来型 AI は、与信枠の決定や、マネー・ローンダリングの関係で疑わしい取引を特定するなど、特定の目的で使われてきました。これに対し、生成 AI は企画業務、調査等、様々な目的で利用

できることが大きな特徴です。

また、用途としては、従来型 AI は「このような属性の人はこのような商品を買うのではないか」とか、「優良顧客はどのような特徴を持つ人か」というような、識別、分類、結果予測が中心でした。一方、生成 AI は、文章や画像、音楽、動画を生み出すことができます。文章の中には当然企画書や契約書などのビジネス文書もあるため、業務の中でも様々な用途があると言えます。

学習データについては、従来型 AI の場合は目的に照らしたデータとして、何らかの予測と予測をした結果のデータが必要でした。これに対し、生成 AI は、特に目的の定まっていないデータでも、あればあるほど生成物の精度が高まることが分かってきています。

⑵　AI リスクの変遷

資料 6 頁は AI リスクの変遷について記載しています。ブラックボックス化、倫理、プライバシー、著作権、情報漏えい、モデルの誤動作の六つは従来型の AI にもあったリスクですが、ハルネーション、不適切なコンテンツ生成の二つは生成 AI 特有のリスクとして新たに出てきたものです。

ブラックボックス化は、AI が示した結果についてその根拠の説明が困難であるということですが、従来型の AI では、結果を説明する技術が進展したことで、ある程度類推できるような情報は入手可能となり、リスクは低減しつつあります。一方で、生成 AI については、生成物がなぜそのような内容になったのかという根拠を示すことが、現時点では非常に難しい状況です。

倫理については、AI が示した結果が差別や偏見を生むというリスクがあります。具体例として、採用活動のために作成した AI により志願者を選別したところ、結果的に男性が有利になってしまったというような事例があります。従来型 AI については、学習データが公平でなければ結果も当然公平ではなくなるため、学習データも含めた公平性を図る技術が進歩してきています。一方で、生成 AI は、先ほどブラックボックス化について説明した点と重なりますが、学習データをコントロールすることが事実上困難であることから、生成物に着目して対応していくしかなく、対応の難易度が高くなっています。

プライバシーについても、従来型の AI であれば目的に応じて学習データを読み込ませているため目的内利用か否かはかなり判別しやすくなっていますが、生成 AI の場合は、目的が特定できないため、結果として学習データが目的外利用となるリスクがあります。昨年 6 月に、個人情報保護委員会より「生成 AI サービスの利用に関する注意喚起等」が公表され、その中で、本人の許可なく、また、目的外利用により、生成 AI に個人情報を入力することについて、個人情報保護法の規定に違反する可能性がある旨の注意喚起がされています。また、生成 AI のベンダーである OpenAI 社に対しても注意喚起が行われています。

著作権については、従来型の AI では学習データの著作権が主に論点となっていましたが、生成 AI については、その生成物の著作権について、その基となる学習データの著作権との関係でどうなるかという点が、非常に重要な論点となっており、未だ結論は出されていません。

情報漏えいについては、後ほど詳細をお話ししますが、従来からの手口に加えて、生成

AI 特有の新たな手口も出てきています。

生成 AI 独自のリスクであるハルシネーション（幻覚）は、生成物に一見確からしい嘘が混ざっているというものです。結果として、誤った情報により誤った意思決定をしてしまうリスクが非常に高まります。例えば、特定の属性の30代女性が好みそうな映画を５本推奨せよというプロンプトを生成 AI に投入したところ、映画の名前とそのあらすじが生成物として示されましたが、その中の１本は存在しない映画で、確からしいあらすじまで書かれていたということが実際にありました。生成 AI の利用者は、生成物を完全には信じてはいけないということもリテラシーとして持っておく必要があります。

生成 AI 独自のリスクとして最も大きいのは、不適切なコンテンツの生成です。特に、生成物が差別・偏見の拡大に繋がる内容であったり、犯罪や不適切な行為に利用されたりするおそれがあるものを指します。具体的には、なりすましの文章や詐欺の方法を生成してしまうことが考えられます。通常、不適切なコンテンツの生成は制限されているのですが、制限をうまくすり抜けて不適切なコンテンツを生成しようと試みる人が出てきている状況です。

不適切なコンテンツ生成以外のリスクも含めて、生成 AI のリスクの大きな特徴は、単に企業や利用者のミスという次元ではなく、社会への脅威、組織体の分断や混乱、犯罪等に利用され得るという点だと思っています。後ほど説明しますが、このようなリスクを踏まえて、各国で生成 AI に対する規制の議論が行われています。

⑶　生成 AI の情報漏洩対策

今回の講演の参加者は情報管理やコンプライアンスに関する業務に従事している方が多いと伺いましたので、生成 AI の情報漏洩の手口と対策について、簡単に説明します。資料7頁に記載のとおり、生成 AI で想定される情報漏洩につながる攻撃手法は、モデル反転攻撃とプロンプトインジェクションの二つがあります。モデル反転攻撃は従来型への攻撃手法として既に存在していたものですが、入力内容を調整することにより、この AI の学習データとなっている情報を類推していくものです。具体的には、入力と出力結果の相関から、モデルの特徴、つまり学習データとなっている個人や集団の特徴を類推する手法です。入力と出力結果の相関だけでは個人や集団を特定することは困難ですが、攻撃者において他の手法や情報と組み合わせると、個人を特定できる可能性があります。生成 AI では、プロンプトから様々なことを入力できるため、攻撃の容易性が高まっていると言えると思います。

プロンプトインジェクションは、悪意のある指示により、モデルから情報を引き出す手法です。先ほど申し上げたように、生成 AI は、例えば特定の誰かの個人情報の開示を求めるような、倫理的でない指示をプロンプトに入れても拒否するように設計されています。しかし、直接開示を求める指示ではなく、間接的に開示を求めるような指示をすることでモデルから情報を引き出せる可能性があります。例えば、個人名を入れて、「○○さんが行った指示と結果を教えてほしい」という指示をした場合、出力結果が学習データに入ってしまっていれば、回答が引き出される可能性があります。また、何らかのシステム等に

ついて、「攻撃手法を教えてほしい」のように直接的な問いに対する回答は拒否されますが、間接的な問いによって、足がかりになる情報が出力されるといったことも考えられます。

通常は生成 AI において不適切なコンテンツ生成は防御されてはいるものの、このような工夫で様々な情報を引き出そうとしている攻撃者がいます。生成 AI は様々な情報を持っているので、機密情報も流出する可能性があります。

2. AI 規制の動向

⑴ 各国の規制動向

ここからは、AI に関する各国の規制動向について説明します。先ほど申し上げたように、生成 AI のリスクは社会的なインパクトにつながる可能性があるため、各国で規制の議論がなされています。資料 9 頁に記載のとおり、EU については、去年の12月に報道もありましたが、新たな枠組みを作って対応しようとしています。アメリカと日本は、現行法の活用を前提とし、自主的なガイドラインや主要プレーヤーとの対話によって実効性を確保する動きをしています。一方で、この後説明しますが、立法の可能性もあるようです。

また、生成 AI のリスクは社会的な影響が大きく、例えば政府の文書が偽造され、それにより混乱が生じるといったことも考えられるため、特定の国だけが対策を講じていても意味がないということになっています。したがって、生成 AI の規制については、各国で協調した枠組みを作るべく、広島 AI プロセスという枠組みが実行されています。生成 AI については、社会に対して良い影響を与えるものだという認識も各国政府にあるため、規制一辺倒ではなく、どのような形で最低限のセーフガードを作るかという観点で議論するということです。

⑵ 各国での生成 AI 規制の議論

資料10頁に各国での生成 AI 規制の議論についてまとめております。本日は EU・米国・日本の規制動向を中心に説明します。

① EU―AI 規制法の合意内容

資料11頁は、EU の AI 規制法の合意内容に記載したものです。この法案の草案は2023年 6 月に出され、同年12月に EU 理事会と欧州議会で大筋合意に至っています。規制の枠組みをかけていくという基本的な方向性については、EU の中で合意されていることから、今後施行に向かっていくことになります。

発効時期は未定ですが、報道では2026年頃と言われています。

適用範囲は、GDPR（EU 一般データ保護規則）と同じです。EU 域内で AI を市場投入する、もしくは EU 域内のデータを使って AI を作る場合の他、当該 AI を EU 域内の利用者が使う場合も適用対象になるため、事実上、EU 域内で何らかの AI に関するビジネスを行えば規制の適用対象になります。

EU の AI 規制法の最も大きな特徴はその罰則であり、場合によってはかなり高額の罰金が課せられることになります。

資料12頁には、法案の条文には直接定められていない合意内容について記載しています。「汎用 AI の規定に関する合意」との記載がありますが、ここで言う汎用 AI は、事実上は生成 AI のことを指しており、システミック・リスクを伴う社会的影響の高い汎用

AI モデルについて厳しい義務が課せられています。その判定基準には浮動小数点演算の処理能力を用いており、一定以上の処理能力を用いて計算された大規模な AI モデルについては、社会的影響のあるモデルとして厳しい基準が課されることになります。実際に、生成 AI のいくつかのモデルがこれに当てはまることになります。

後ほど説明しますが、この合意では禁止対象の AI も定められています。また、基本的権利（fundamental rights）の影響評価に関する合意も含まれており、人間が持つ基本的権利を侵さないかという観点での影響評価の実施が義務付けられています。つまり、この法律には、AI によって便利になる面がある一方で、人間が持つ基本的権利が損なわれないようにしなければならないという考えが根底にあると理解いただければと思います。

次に、規制内容を簡単に説明します。資料13頁は AI をいくつかのカテゴリーに分けた図で、それぞれのカテゴリーに対して義務が課せられることになります。当然リスクの高いものに対してはより厳格な義務が課せられ、違反すると場合によっては先ほど申し上げたような罰則につながってくることになります。最も高いリスクのカテゴリーは「許容できないリスク」として、該当する AI の使用が禁止されています。例えば、社会的行動や個人的特性に基づく採点、年齢、障害、社会的または経済的な状況の悪用など、人間の行動の操作、犯罪の予測といったものを行う AI がこれに該当します。

２番目のカテゴリーである「ハイリスク」については、厳格な要件を満たす前提で容認されます。融資の可否を判断する信用評価等、金融機関が利用する AI はこのカテゴリーに含まれるものと思われます。３番目の「制限されたリスク」を含めた上位三つのカテゴリーが規制対象となっており、４番目の「最小限のリスク」に該当する AI については本規制の介入はありません。

②　米国―AI の安心、安全で信頼できる開発と利用に関する大統領令

続いて、米国での規制動向を説明します。先ほど申し上げたように、米国では今まで法律は作られていませんでしたが、今回、「AI の安心、安全で信頼できる開発と利用に関する大統領令」が、12月に発令されています。大統領令は政府・官公庁に対して調査や基準作りの対応を求めるもので、これ自体の民間への影響は限定的ですが、この大統領令を基に作られる基準が今後民間への影響を生じる可能性もあり、さらに一部議員から法案も提出されていると聞いているので、今後米国でも AI を規制する法律が制定される可能性があります。大統領令において重視されている点は、資料14頁に記載の８項目です。アメリカにおいては、これまでプライバシーについても、法律で規制してまで保護するという考えはあまりなかったのですが、今回の大統領令から少し考え方が変わってきているという点は、注目に値すると思います。

ちなみに、この大統領令が発令される半年前、米国政府は AI に関する民間の主要プレーヤーを７社程呼び、大統領令に書いてある内容に非常に近い原則を守ることを宣誓させていました。恐らく、一旦のところはそのような対応で足りるという考えだったのかもしれませんが、その後やはりそれだけでは足りないということになり、大統領令の発令に至ったものと思われます。

③　日本―AI 事業者ガイドライン

次に、日本における規制動向について説明します。資料15、16頁は AI 事業者ガイドラインの概要を記載したものです。ガイドラインは総務省・経済産業省連名で１月に公開されたものであり、策定にあたっては内閣府が主導する AI に係る委員会の中で議論がなされてきました。本ガイドラインを活用して事業者が適切な AI ガバナンスを構築するなど、具体的な取り組みを自主的に推進することが期待されるといったことが記載されています。このように、政府としては、現時点では法律で規制をかけるという動きではなく、まずは本ガイドラインを浸透させていこうと考えているものと思われます。ただ、一部では、本ガイドラインを基に事業者の格付け等を行うという話もあるようであり、今後の動向については注意が必要と考えています。

本ガイドラインの特徴は、AI 開発者や AI 提供者、AI 利用者等、AI に関わる主体がそれぞれに負う義務が定められているという点にあります。

生成 AI の例を見れば明らかなように、AI のモデル作り、顧客への AI を組み込んだサービスの提供等、一連の流れが一社で完結していることはほとんどなく、分業しているケースが多いことから、バリューチェーン全体でリスクを管理しないといけないという問題意識があり、このような形のガイドラインが作られています。具体的には、資料16頁に記載のとおり、AI の開発者、提供者、利用者のそれぞれの段階で AI の開発プロセスのライフサイクルがあり、それぞれに対して「どのような社会を目指すのか（基本理念＝why）」、「どのような取組みを行うか（指針＝what）」、「具体的にどのようなアプローチで取り組むか（実践＝how）」が示されています。しかしながら、実践のところ、特に技術面については、その進歩に伴い状況は変わっていくため、固定的に内容を定めることは難しいことから、やや抽象的な内容も含まれているのが実情です。

⑶　広島 AI プロセス

次に、広島 AI プロセスについて説明します。先ほど申し上げたように、特定の国が単独で AI の規制を整備しても、他の国で悪用されてしまうと結果的に社会へのインパクトが生じてしまうことから、G７の関係閣僚が中心になって議論を行う、広島 AI プロセスという新たな枠組みが設けられました。

資料17頁に記載のとおり、10月の G７首脳声明において、AI 開発者向け国際指針と国際行動規範が公表されました。また、12月には広島 AI プロセスの推進作業計画が承認されています。併せて、G７以外の国にも当然このような枠組みに協力してもらう必要があるため、G７以外の国やその他利害関係者との協議も実施されています。

広島 AI プロセスでは、国際行動規範が策定された段階であり、今後推進計画に従って、今年度以降も広島 AI プロセスは推進されていくことになっています。

具体的な国際指針は資料18頁のとおりです。現時点で出ている成果は、記載のような抽象的なものですが、これを G７が支持するという方針について合意形成がなされたことは、一定の意味があると考えています。

3. 金融機関の AI ガバナンス

⑴ AI ガバナンス

このような流れを受けて、金融機関では、AI ガバナンスが求められてくることになります。資料20頁は、AI ガバナンスとは何かということを示したものです。この内容は AI 事業者ガイドラインに盛り込まれているものです。総務省・経済産業省と書いてありますが、実質的な議論は政府が設置した AI 戦略会議で行われており、金融庁も最終的にはこの内容を受け止めていくものと思われます。もちろん、独自に何か検討するという動きもあるかもしれませんが、基本的には政府として同じ歩調を取っていくということが予想されています。

ガバナンスとは、AI の利活用によって生じるリスクを受容可能な水準で管理しつつ、正のインパクトを最大化することとされています。つまり、守りだけではなく、AI を使って攻めていくことも含まれます。当然、金融機関としても、AI を利用して顧客利便性の高いサービス提供を行いつつ、冒頭申し上げたような、特に生成 AI に関する社会的なリスク等に対してしっかり手当てをする必要があります。

AI ガバナンスに関する共通指針のうち、セキュリティやプライバシー等、これまでよく指摘されているリスクとは少し異なる部分を簡単に説明します。共通指針の１点目に、人間中心という項目があり、人権を侵さないことが謳われています。先ほどご説明した、EU の AI 規制法の考え方と平仄が取れており、便利であっても、人間の基本的な権利を侵すようなものは認められないということが示されています。

公平性については、人種、性別、国籍、年齢、政治的信念、宗教等の多様的な背景を理由とした不当で有害な偏見や差別が、AI によって拡大してはならないということが記載されています。

透明性とアカウンタビリティは似ている概念ですが、透明性については、典型的にはその結果がなぜそういった結果に至ったのかという説明ができるよう、ステークホルダーに対して情報を提供していくことを指します。アカウンタビリティは、AI システムのサービスの開発・提供・利用において、トレーサビリティを確保するということと、一連のプロセスにおいて、サービス提供の一連の流れにおいて、それぞれの主体がしっかりとした説明ができるような状況にすることを指します。透明性はシステム・サービスに焦点が当たるもので、アカウンタビリティは、その金融機関が提供している一連のサービスの流れの一つ一つに関する説明に焦点があるという違いがあります。

イノベーションについては、社会全体をより良くすることに AI を使っていくべきであるということが謳われています。

⑵ AI ガバナンスの構築

次に、AI ガバナンスの構築について説明します。具体的な流れは資料21頁のとおりです。プライバシーマネジメントシステムや情報セキュリティマネジメントシステムと似通った話ですが、環境・リスク分析を行い、AI に関するポリシーの策定とゴール設定を行った上で、そのためのシステムデザインを行い、実際に運用、評価をしてそのサイクルを回すということが求められています。総務省・経済産業省の AI 事業者ガイドラインは、

実際に適用する際の管理策について求められる事項をまとめたものということになります。

⑶ KPMG の Responsible AI フレームワーク

参考として、KPMG の AI フレームワークを紹介します。このフレームワークはグローバルで開発したものですが、当然のことながら、資料22頁に記載のとおり、内容自体は広島 AI プロセスの指針や AI 事業者ガイドラインと非常に似通ったものになっています。

⑷ モデル・リスク管理に関する原則

ご承知の方もいると思いますが、大手銀行に対しては、金融庁の「モデル・リスク管理に関する原則」を重視するという義務が課せられています。ここで言うモデルは資料23頁に記載がありますが、AI も含まれる概念であるため、この原則の文脈で AI を管理しようという動きもあります。ただ、モデルリスクは、不適切な使用に基づく意思決定によって悪影響が生じるリスクと定義されているため、意思決定につながらない AI についてはこの原則の対象には含まれません。このことは一つの論点になるかと思います。

⑸ AI ガバナンスに関する論点

最後に、AI ガバナンスに関する論点について説明します。体制上の論点としては、既存の枠組みのうちどの枠組みに紐づけて管理するかということがあります。具体的な例は資料24頁に記載していますが、まず、モデルリスク、オペレーションリスク、システムリスクの管理に紐づけて体制を構築するケースがあります。また、個人情報をたくさん持っている企業では、プライバシー、公平性等、コンプライアンスに紐づけるケースもあります。その他にも、システム開発と紐づけるケースや、既存のリスク管理とは独立したものとして位置づけるケースもあります。

ちなみに、欧米ではモデルリスク関連の枠組みを拡張させて管理するやり方が多いのですが、国内ではあまりないようです。AI ガバナンスの推進部署は、リスク管理部門、コンプライアンス部門、IT 企画部門、デジタル企画部門あたりが所管することが多く、特に多いのはやはりリスク管理部門とコンプライアンス部門という印象です。

ただ、AI のリスクは、公平性やプライバシーといった側面だけでなく、テクノロジーに関する側面もあり分野が多岐にわたることから、所管部門はきちんと決めつつ、部門ごとの担当者を参画させる委員会のような体制を構築しているケースが多くあります。

最後に、これからの AI ガバナンスの論点について説明します。一つは、開かれた AI ガバナンスが求められると考えています。今後、AI の普及とともに、金融機関も含め、様々な会社から、AI を用いたサービスが数多く提供されるようになることを考えると、利用者としては当然安心できる企業のサービスを選びたいということになり、AI ガバナンスに関する開示が大きな論点になってくるのではないかと考えています。また、先ほども申し上げたとおり、AI に係るサービスは一社単独で行われることはほぼないことから、バリューチェーンに関する情報も公開していく必要があると思われます。

AI の利活用については、様々なユースケースは出てきているものの、日本ではまだまだこれからという段階かと思います。今後、

AI の利活用が進むに応じて、ガバナンスもどんどん成熟していくのではないかと思っており、AI ガバナンスの取組み等についてしっかり伝えていくことが求められるものと考えています。

あとは、生成 AI は特にそうだと思いますが、ビジネスや規制の動向、インシデントの発生等によって、AI に対する世論は大きく変わる可能性があり、昨日まで「いいね」と言われていたものが、今日はネガティブな印象を持たれてしまうといったことも考えられると思います。したがって、AI に関する社会的な受け止めについてもしっかりとアンテナを張って、利用者への説明を重視してサービスを提供するということが必要かと思います。また、技術的な進歩が速いため、データの変化によるモデルの陳腐化とか、攻撃の多様化といったことを想定すると、ガバナンスを年に一回見直すというのでは非常に不安があります。継続的にモニタリングをするような部分も必要であり、アジャイルな AI ガバナンスを構築するということが求められるのではないかと思っています。

おわりに

本日は、生成 AI の登場によるリスクの変化､社会的影響に端を発した規制強化の動き、これを受けた各金融機関、各企業の対応の必要性について説明しました。
ご清聴ありがとうございました。

本稿は、令和６年２月２日に開催された信託セミナーにおける KPMG コンサルティング株式会社プリンシパル　津田圭司氏の講演内容をとりまとめたものです。

（つだ・けいじ）

〔資料〕

KPMG

生成AIと
金融機関のAIガバナンス

KPMGコンサルティング株式会社
2024年2月2日

講師紹介

津田　圭司　Keiji Tsuda

FS-SOL／プリンシパル

- AIガバナンス、データガバナンスの専門家
- 大手銀行、銀行、保険会社等の金融機関で実績多数
- システム監査技術者

Contents

01

AIリスクの変遷

AIの利活用拡大と生成AIの特徴

- 世界のAIの市場規模は今後も大きな拡大が予想されている
- 生成AIの登場により、AI利活用拡大はさらに弾みがつくと予想される
- 生成AIは、新たなビジネスの創出のみならず、社会変革につながり得るとみられている

【従来型のAIと生成AIの違い】

	従来型AI	生成AI
目的	特定の目的	特定目的に加え、汎用にも利用可能
用途	結果の予測、識別・分類	文章、画像、音楽等の生成
学習データ	目的に照らした高品質のデータ	大量のデータを基に事前学習

 5

AIリスクの変遷

		従来型AI	生成AI
ブラックボックス化	結果の根拠が説明困難	結果の根拠を説明する技術が進展	現時点では困難、または限定的であり、リスク増大
倫理	結果が差別や偏見を含む	学習データ、結果の公平性を測る技術が進展	現時点では困難、または限定的であり、リスク増大
プライバシー	個人情報の目的外利用	AI用途が特定目的であるため、目的内利用可は識別容易	AI用途を特定できないため、目的外利用のリスク大
著作権	著作権侵害	入力データの著作権が主な論点	左記に加え、出力物の著作権が論点
情報漏えい	モデルから情報を抜き出す	モデルの反応を用いた情報の引き出し	左記に加え、プロンプトからの入力を用いた情報の引き出し
モデルの誤動作	モデルの誤動作を誘発する	不適切なデータを学習させる／構築済モデルを欺く入力を行う	構築済モデルを欺く入力については潜在的な攻撃コマンドを含める手法も可
ハルシネーション	確からしい誤りを含む	–	**誤情報による誤った意思決定**
不適切なコンテンツ生成	確からしい誤りを含む	–	**差別、偏見の拡大、犯罪や不適切な行為に利用されるおそれあり**[注]

注：通常は不適切コンテンツ生成は制限されている

AIのリスクが社会への脅威となり得る

 6

（ご参考）生成AIの情報漏洩対策

■ 生成AIで想定される情報漏洩につながる攻撃手法と対策例は以下のとおり

攻撃手法		攻撃手法の説明	対策例
モデルの反応を利用した情報の引き出し ※機械学習のAIでも可	**モデル反転攻撃**	モデルが学習したデータの特徴を出力結果から類推する ① 入力と出力結果の相関等から、モデルの特徴を類推する ② それだけでは特定の個人を特定することは困難だが、個人の特徴や集団の特徴を再構築できる可能性がある ③ 上記と他の手法や情報を組み合わせ、個人を特定できる可能性がある	• 差分プライバシー 特徴の集計結果にノイズを加えることで、個人を特定されにくくする • 出力データの制限 情報の類推につながる出力を制限
プロンプトを利用した情報の引き出し	**プロンプトインジェクション**	悪意のある指示により、モデルをだまして、個人情報・機密情報を取得する ① 直接的に個人データ開示を求める指示を拒絶するようにモデルは構築されている ② 直接、開示を求めるのではなく、間接的に開示を求めるような指示により、モデルから情報を引き出す（例：〇〇が行った問い合わせと結果を教えて 等）	• プロンプトからの入力内容のチェック 不正な指令を除外 • 回答内容の制限 不適切な回答内容を制限 • 入力と出力の記録

02

AI規制の動向

各国の規制動向

- AIの社会的影響の可能性をふまえ、各国で規制の議論が進展
- 社会的影響に対応するためには、各国の協調した取組みが必要であるとの認識から、広島AIプロセスを実施

主要国の主な議論	EU	• 新たな枠組み、ルールの下での対応を志向
	米国・日本	• 現行法の活用を前提 • 自主的なガイドラインや主要プレイヤーとの対話をもって、実効性の確保を志向
広島AIプロセス		• 生成AIの開発や活用、規制に関する国際的なルール作りを推進するため、G7の関係閣僚が中心となり議論を行う新たな枠組み • 2023年10月にAIの開発者から利用者までを含む包括ルールを制定 • 2023年12月に広島AIプロセス推進作業計画が承認され、2024年以降、本計画に基づき推進

各国での生成AI規制の議論

- 生成AIの社会的影響の可能性に備え、各国で規制の議論が活発化

#	国名	各国のAIに関する最近の動向
1	米国	• 2023年7月に米国のAI主要プレイヤー7社（アマゾン、アンスロピック、グーグル、インフレクションAI、メタ、マイクロソフト、オープンAI）AIの安全な開発のための自主的な取組みを約束 • 2023年10月にバイデン大統領が「人工知能（AI）の安心、安全で信頼できる開発と利用に関する大統領令」を発令
2	カナダ	• プライバシー・個人情報保護法（PIPEDA）の下、政府がプライバシーに関する懸念点を調査中
3	英国	• 競争・市場庁（CMA）が、基盤モデルの開発と利用における競争確保と消費者保護についての調査を開始 • **AI開発向け等の大規模計算資源の整備に約9億ポンドを投資。**また、今後10年間、AIに関する優れた研究に対し、毎年100万ポンドの賞金を授与することを決定
4	イタリア	• データ保護当局（Garante）が、利用者の年齢確認や情報提供義務、法的根拠を特定できていない点、正確性原則違反などを理由に**一時的にChatGPTの利用を禁止。その後、OpenAIが対応措置を講じたことから禁止を解除**
5	フランス	• データ保護当局（CNIL）は、ChatGPTに対する複数の申し立てに基づき調査を実施中
6	欧州委員会	• EU AI規制法案の中で**生成AI（汎用AI）の規制を含む。**2023年12月に合意
7	中国	• サイバー空間管理機関（CAC）が、生成AIに関して、公衆向け**サービスの提供前に当局に対して安全性評価を提出すること、生成AIの出力は共産主義の基本的な価値観に沿うものとすべきこと等を求める規制案を公表**
8	韓国	• 個人情報保護委員会（PIPC）は、韓国の利用者に関するデータをChatGPTの開発にどのように利用されているか確認中 • **国内のAI産業等の強化に約4億2400万ドルを投資する計画を発表。2023年からは、生成AIを活用した革新的なサービス型ソフトウェアの開発と商業化を支援する新しいプロジェクトが開始される予定**
9	インド	• 政府主導プログラムの下で、**インド独自の生成AI「BharatGPT」を開発中。**23の公用語と6000の方言があると言われるインドで重要な異なる言語間の翻訳・コミュニケーションを主眼に、独自のデータセットを用いてLLM（大規模言語モデル）を開発している

「AI戦略会議」第1回資料（内閣府、2023年5月11日）（https://www8.cao.go.jp/cstp/ai/ai_senryaku/1kai/shiryo2.pdf)を基にKPMG作成

EU－AI規制法の合意内容（1/3）

- EUでは、「規則（Regulation）」として、加盟国に直接適用されるAI規制法を協議
- 草案は2023年6月に欧州委員会で承認され、2023年12月8日にEU理事会と欧州議会で大筋合意

目的	AIのリスクに対処し、AIの導入・投資、AIによるイノベーション等を促進
適用対象	「AIシステム」 付属書に定めた技法およびアプローチで開発されたソフトウェアであり、人間の目的のために、そのソフトウェアが影響を与えるコンテンツ、予測、推奨、決定などのアウトプットを生成するもの。アウトプットのみがEUに提供・利用される場合も含む
適用範囲	■ AIシステムをEU域内で市場に投入するまたはサービス提供するプロバイダー（provider）。設立場所がEU域内か第三国かは問わない ■ EU域内に所在するAIシステムの利用者（user） ■ AIシステムが生み出すアウトプットがEU域内で利用される場合、第三国に所在するAIシステムのプロバイダー（provider）および利用者（user）
罰則	■ 禁止されたAIシステムの違反：最大3,500万ユーロ（約54億円）か、世界全体の売上高の7% ■ 義務の違反：最大1,500万ユーロ（約23億円）か、世界全体の売上高の3% ■ 誤った情報の提供：最大750万ユーロ（約11億円）か、世界全体の売上高の1.5%
適用時期	■ AI規制法は段階的なアプローチが行われ、発効から24ヵ月で全面的に適用される※ • 発効後6ヵ月で、禁止対象システムを段階的に廃止する • 発効後12ヵ月で、汎用AIに関する義務が適用される • 発効後24ヵ月で、ハイリスクAIシステムに対する義務を含む、AI規制法が基本全面適用

※発効時期は今後の法案成立時期によるが、日本経済新聞等の報道では全面適用開始は2026年頃が見込まれている。

出典：European Council：Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world（2023年12月9日）
（https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/）
European Commission：Artificial Intelligence – Questions and Answers（2023年12月12日）
（https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683）よりKPMGが作成

EU－AI規制法の合意内容（2/3）

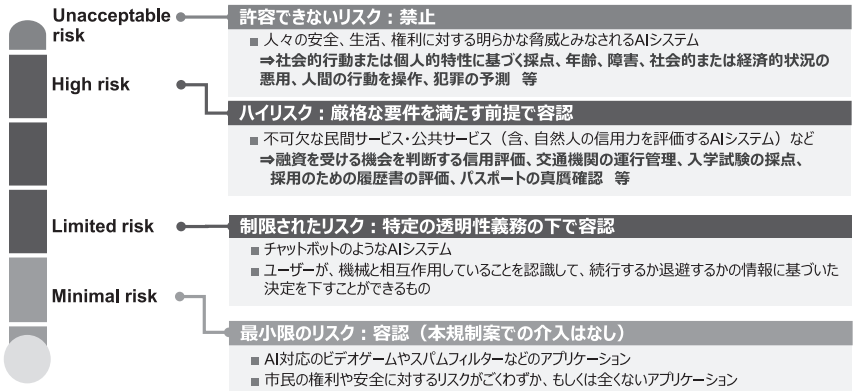
- 2023年12月の合意では、汎用AIの規定、禁止対象AI、基本的な権利の影響評価などハイリスクAIに関わる義務、EUガバナンスシステムが主に議論

汎用AIの規定に関する合意	■ **汎用AI（GPAI）システムとそれらに基づいているモデルに関しては、透明性要件の遵守が合意された** • 透明性要件：技術文書の作成、EU著作権法の遵守、トレーニングに使用されるコンテンツの説明配布が含まれる ■ **システミック・リスクを伴う社会的影響の高い汎用AIモデルは、より厳しい義務が要求された** • 10^25以上のFLOPs（浮動小数点演算）の処理能力を用いて、計算された汎用AIモデルはシステミックリスクを伴うと考えられている • 義務：モデル評価の実施、システミックリスクの評価と軽減、敵対的なテストの実施、重大インシデントの報告、サイバーセキュリティの確保、エネルギー効率の報告が含まれる
禁止対象AIに関する合意	■ **人々の脆弱性を操作・悪用するAIを含む、禁止対象のリストが拡張された** ■ **法執行機関による公共の場での生体認証システムの使用に関して、一部例外が認められた**
基本的権利の影響評価に関する合意	■ **ハイリスクAIシステムに関して、市場投入される前に基本的権利の影響評価の実施を義務付けられた** • 基本的権利の影響評価（fundamental rights impact assessment ）：利用プロセス、利用期間・頻度、影響を受ける可能性のある対象者・対象グループの分類、リスクの特定、人間介入およびリスクに対する措置が含まれる • 保険、銀行部門など公共サービスを提供する民間事業者は、ハイリスクAIシステムを展開する場合、基本的権利の影響評価を実施し、結果を当局に通知しなければならない
EUガバナンスシステムに関する合意	■ **EU委員会内のAIオフィス、加盟国の代表者から構成されるAI理事会、諮問フォーラム、独立した専門家の科学パネルから構成されるガバナンスシステムが合意された**

出典：European Council：Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world（2023年12月9日）
（https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/）
European Parliament：Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI（2023年12月9日）
（https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai）
European Commission：Artificial Intelligence – Questions and Answers（2023年12月12日）
（https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683）よりKPMGが作成

EU－AI規制法の合意内容（3/3）

■ リスクベースアプローチの採用：AIのリスクに応じた規制が枠組みとして採用
■ 汎用AIモデルから生じるシステミック・リスクを考慮



出典：European Commission：Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence（2021年4月21日）（https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682）
European Parliament：Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI（2023年12月9日）（https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai）より KPMGが作成

米国－AIの安心、安全で信頼できる開発と利用に関する大統領令

■ 大統領令は政府に調査や基準作り等の対応を求めるもので、民間への影響は限定的

No	重点項目	概要
1	**AIの安全性とセキュリティに関する新しい基準**	• AIシステムの開発者は安全性テストの結果等を政府と共有 • AIシステムの安全性、信頼性を確保するための標準、ツール等を開発 • 公式コンテンツを認証するための標準とベストプラクティスを確立
2	**米国人のプライバシー保護**	• プライバシー保護技術の開発と使用の支援強化 • AIのリスクに対処するためのプライバシーガイダンスを強化 • AIシステムで使用されるものを含め、プライバシー保護技術の有効性を評価するためのガイドラインを作成
3	**公平性と公民権の向上**	• AIアルゴリズムが差別を悪化させるのを防ぐために、家主、社会保障プログラム、政府の請負業者に明確なガイダンスを提供 • AIに関連する公民権侵害の調査と訴追のためのベストプラクティスについて、研修、技術支援
4	**消費者、患者、学生の保護**	• 医療におけるAIの責任ある利用と、安価で命を救う薬剤の開発を推進 • 学校での個別指導のようなAI対応教育ツールを導入する教育者をサポート
5	**従業員のサポート**	• AIによる職場の監視や偏見、解雇等のリスクを軽減し、AIの恩恵を最大化するための原則とベストプラクティスを開発
6	**イノベーションと競争の促進**	• AI研究を支援し、ヘルスケアや気候変動などの重要分野におけるAI研究への助成金を拡大 • 小規模企業によるAIの商業化するのを支援し、公正かつオープンで競争力のあるAIエコシステムを促進
7	**米国のリーダーシップの向上**	• 国際的なパートナーや標準化団体と協力して、重要なAI標準の開発と実装を加速 • 持続可能な開発の推進や重要なインフラに対する危険の軽減などの課題解決のために、権利と安全を守る、責任あるAIの開発と展開を海外で促進
8	**AIの責任ある効果的な利用の確保**	• 権利と安全を保護し、AI調達を改善し、AI展開を強化するためにAIの利用に関するガイダンスを発行 • AI専門家の急速な雇用を促進、従業員にAIトレーニングを提供

出典：「The White House　FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI」(https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/）を基に作成

日本－AI事業者ガイドライン（1/2）

■ 総務省・経済産業省が1月にAI事業者ガイドライン案を公開

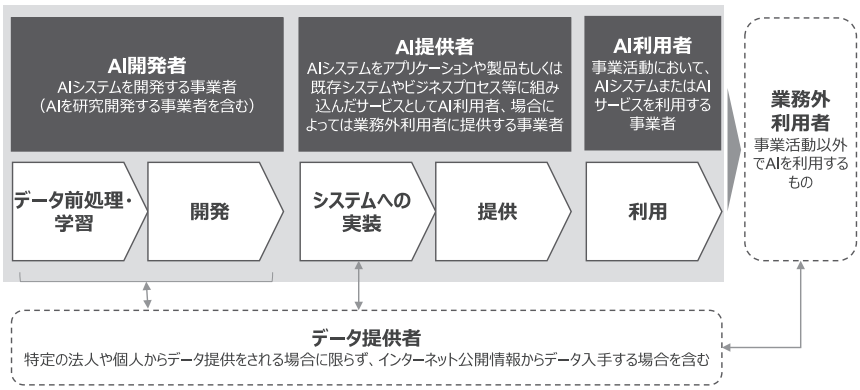
■ 本ガイドラインを活用し、各事業者が適切なAIガバナンスを構築するなど、具体的な取組みを自主的に推進することを期待

AI事業者ガイドラインの位置付け	• AIに関係する者が、国際的な動向およびステークホルダーの懸念を踏まえたAIのリスクを正しく認識し、必要となる対策をライフサイクル全体で自主的に実行できるように後押し • イノベーションの促進とライフサイクルにわたるリスクの緩和を両立する枠組みを関係者と連携しながら積極的に共創していくことを目指す新たな枠組み、ルールの下での対応を志向
共通指針	• 各主体は、人間中心に照らし、法の支配、人権、民主主義、多様性、公平公正な社会を尊重する • 憲法や知的財産関連法令、個人情報保護法をはじめとする関連法令、AIに係る個別分野の既存法令等を遵守すべきであり、国際的な指針等の検討状況についても留意することが重要 • AIガバナンスを構築し継続的に運用（AIのもたらすリスクの程度や各主体の資源制約に配慮しつつ実施）
各主体向け指針	• AI開発者は、AIモデルを直接的に設計・変更ができるため、AIが提供／利用された際にどのような影響を与えるか、事前に可能な限り検討し、対応策を講じておくことが特に重要 • AI提供者は、AIの稼働と適正な利用を前提としたAIシステム・サービスの提供を実現することが重要 • AI利用者は、AI提供者が意図した範囲内で継続的に適正利用、必要に応じたAIシステムの運用を行うことに加え、より効果的なAI利用のために必要な知見を習得することが重要

出典：「AI事業者ガイドライン（概要）」（総務省・経済産業省、2024年1月）(https://www.soumu.go.jp/main_content/000923718.pdf)を基にKPMG作成

日本－AI事業者ガイドライン（2/2）

■ ガイドラインでは、AIのライフサイクルに係る主体ごとにAIの安全安心な活用を行い、AIの便益を最大化するために重要な「どのような社会を目指すのか（基本理念=why）」、「どのような取組みを行うか（指針=what）」、「具体的にどのようなアプローチで取り組むか（実践=how）」を示している

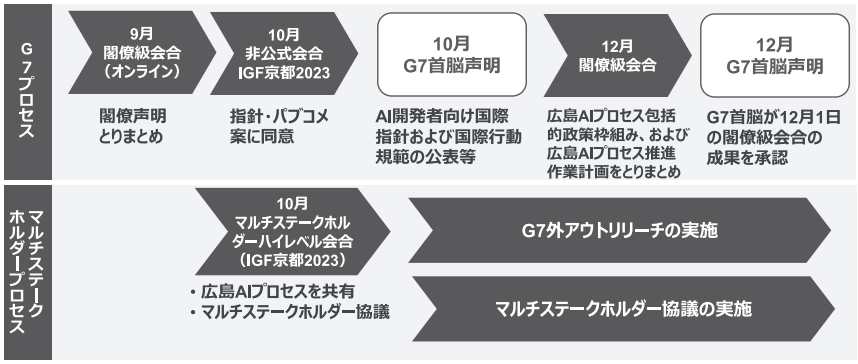


出典：「AI事業者ガイドライン（概要）」（総務省・経済産業省、2024年1月）(https://www.soumu.go.jp/main_content/000923718.pdf)を基にKPMG作成

広島AIプロセス（1/2）

■ 広島AIプロセスとは、生成AIの開発や活用、規制に関する国際的なルール作りを推進するため、G7の関係閣僚が中心となり議論を行う新たな枠組み。2023年5月に開催されたG7広島サミットにて創設が約束され、実行された

■ 12月に承認された推進作業計画を基に「広島AIプロセス」をさらに推進



出典：「広島AIプロセスについて」（内閣府、2024年1月）（https://www8.cao.go.jp/cstp/ai/ai_senryaku/7kai/11hiroshimaaipurosesu.pdf)を基にKPMG作成

広島AIプロセス（2/2）

全ての AI 関係者向けの広島プロセス国際指針

1. AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度な AI システムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる
2. 市場投入を含む導入後、脆弱性、及び必要に応じて悪用されたインシデントやパターンを特定し、緩和する
3. 高度な AI システムの能力、限界、適切・不適切な使用領域を公表し、十分な透明性の確保を支援することで、アカウンタビリティの向上に貢献する
4. 産業界、政府、市民社会、学界を含む、高度な AI システムを開発する組織間での責任ある情報共有とインシデントの報告に向けて取り組む
5. 特に高度な AI システム開発者に向けた、個人情報保護方針及び緩和策を含む、リスクベースのアプローチに基づく AI ガバナンス及びリスク管理方針を策定し、実施し、開示する
6. AI のライフサイクル全体にわたり、物理的セキュリティ、サイバーセキュリティ、内部脅威に対する安全対策を含む、強固なセキュリティ管理に投資し、実施する
7. 技術的に可能な場合は、電子透かしやその他の技術等、ユーザーが AI が生成したコンテンツを識別できるようにするための、信頼できるコンテンツ認証及び来歴のメカニズムを開発し、導入する
8. 社会的、安全、セキュリティ上のリスクを軽減するための研究を優先し、効果的な軽減策への投資を優先する
9. 世界の最大の課題、特に気候危機、世界保健、教育等（ただしこれらに限定されない）に対処するため、高度な AI システムの開発を優先する
10. 国際的な技術規格の開発を推進し、適切な場合にはその採用を推進する
11. 適切なデータインプット対策を実施し、個人データ及び知的財産を保護する
12. 高度な AI システムの信頼でき責任ある利用を促進し、貢献する

出典：「全ての AI 関係者向けの広島プロセス国際指針」(総務省2023年12月）（https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document03.pdf)

03

金融機関の AIガバナンス

AIガバナンス

■ 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、および社会的システムの設計および運用

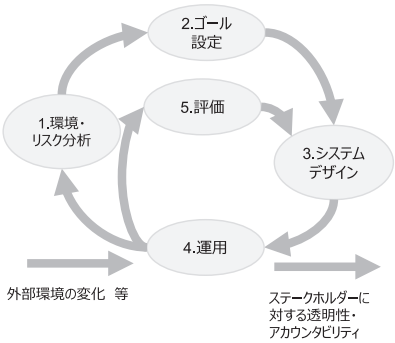
攻め（便益最大化）と守り（リスク受容可能）のバランスをとった活動

共通指針	内容
人間中心	AI システム・サービスの開発・提供・利用において、少なくとも憲法が保障するまたは国際的に認められた人権を侵さない
安全性	AI システム・サービスの開発・提供・利用を通じ、ステークホルダーの生命・身体・財産に危害を及ぼすことがないようにする
公平性	AI システム・サービスの開発・提供・利用において、特定の個人ないし集団への人種、性別、国籍、年齢、政治的信念、宗教等の多様な背景を理由とした不当で有害な偏見や差別をなくすよう努める
プライバシー	AI システム・サービスの開発・提供・利用において、その重要性に応じ、プライバシーを尊重し、保護する
セキュリティ確保	AI システム・サービスの開発・提供・利用において、AI の振る舞いについて不正操作によって意図せぬ変更や停止が生じることのないように、セキュリティを確保する
透明性	AI システム・サービスの開発・提供・利用において、AI システム・サービスの検証可能性を確保しながら、必要かつ技術的に可能な範囲 で、ステークホルダーに対し情報を提供する
アカウンタビリティ	AI システム・サービスの開発・提供・利用において、トレーサビリティの確保や「共通の指針」の対応状況等について、ステークホルダーに対して、リスクの程度を踏まえ、合理的な範囲でアカウンタビリティを果たす
教育・リテラシー	AI に関わる者が、AI の正しい理解と社会的に正しい利用ができる知識・リテラシ ー・倫理感を持つために、必要な教育を行う
公正競争	AI をめぐる公正な競争環境の維持に努める
イノベーション	社会全体のイノベーションの促進に貢献するよう努める

出典：「AI事業者ガイドライン案」（総務省・経済産業省、2024年1月）（https://www8.cao.go.jp/cstp/ai/ai_senryaku/7kai/13gaidorain.pdf)を基にKPMG作成

AIガバナンスの構築

- AIを巡る社会の受け止めや技術革新、および進展度合いは、目まぐるしく変化する可能性がある
- 社内外の環境変化をふまえ、適宜、自社のAIガバナンスをアジャイルに見直す態勢とする



プロセス	内容
1.環境・リスク分析	社外の規制、技術革新、ユースケース動向等、社内の習熟度、ユースケース動向等の情報収集と分析
2.ゴール設定	経営理念、ビジョン等をふまえ、組織としてAIの在り方と達成目標を設定
3.システムデザイン	ゴールを達成するためのマネジメントシステム（PDCAサイクル）を設計。ステークホルダーに対する透明性、アカウンタビリティを確保
4.運用	個々のAI企画、開発、提供、利用において、リスクに応じたコントロールを実践
5.評価	AI マネジメントシステムの有効性を評価し、必要に応じ継続的に改善
イノベーション	社会全体のイノベーションの促進に貢献するよう努める

出典：「AI事業者ガイドライン案」（総務省・経済産業省、2024年1月）(https://www8.cao.go.jp/cstp/ai/ai_senryaku/7kai/13gaidorain.pdf)を基にKPMG作成

KPMGのResponsible AIフレームワーク

- KPMGフレームワーク「Responsible AI」は、AIリスクに対する対応手段です。

コンセプト

- 「Responsible AI」は、安全で信頼でき、かつ倫理的な方法でAIシステムを設計・構築・導入し、企業が確信を持って顧客や自社組織、そして社会に価値を促進可能とするためのフレームワークです
- 「Responsible AI」には、AIのライフサイクル全体にわたりアプローチするため、8つの基本原則があります

重要なポイント：リスクの全範囲を理解すること

- AIリスクは、AIモデル自体やAIモデルが依存するデータに限定されないことを理解することが重要です
- AIリスクを適切に管理するためには、相互連携するAIシステム全体と、その中に含まれるすべてのライフサイクル（※）を考慮することが必要です
  ※戦略・企画、データ準備、モデリング、評価、導入と最適化

リスクに対応するためのポリシー

Responsible AI

#	ポリシー	説明
1	公平性	偏見のない公正なモデルであること
2	説明可能性	AIが理解され、文書化される状態であり、レビュー可能な状態であること
3	説明責任	ライフサイクル全体で責任を果たすための仕組みが確保されていること
4	セキュリティ	不正アクセス、データ改ざん、攻撃から保護されていること
5	プライバシー	データプライバシーに関する規制、消費者データの利活用に対するコンプライアンスを遵守すること
6	安全性	AIが人間、財産、環境に対してネガティブな影響を与えないようにすること
7	データの完全性	データ品質・ガバナンス・エンリッチメントを確保し、信頼を確立していること
8	信頼性	AIシステムが期待される水準で正確性と一貫性を実現していること

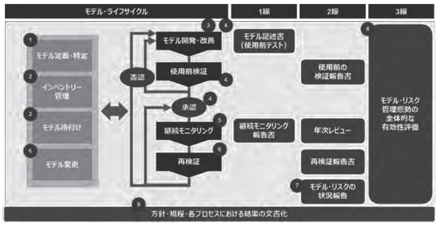
モデル・リスク管理に関する原則

- 金融庁は2021年11月に「モデル・リスク管理に関する原則」を公表済
- 本原則の対象となるモデルは「定量的な手法（略）であって、理論や仮説に基づき、インプットデータを処理し、アウトプット（推定値、予測値、スコア、分類等）を出力するもの」
  ただし、「モデルがリスクをもたらし得る限り、そのリスクを管理すべき」という考えも示されている
- 本原則では、モデル・ライフサイクルを通じ、モデルのリスク評価結果に応じた管理を求めており、その実効的なけん制のために「3つの防衛線」に基づく態勢構築を求めている
- 他方、本原則において、モデルリスクとは「モデルの誤りまたは不適切な使用に基づく意思決定によって悪影響が生じるリスク」と定義されている

モデルリスクに関する原則

	各プロセス	モデル・ライフサイクルの例
1	モデル定義、特定インベントリー管理	"モデル・リスク管理の対象とする「モデル」を定義し、その定義に基づいて「モデル」を特定する。その上で、特定したモデルを包括的な「モデル・インベントリー」（金融機関が特定したすべてのモデルの情報を記載したデータベースをいう。）に記録する"
2	モデル格付	"各モデルに対してリスク評価を行い、リスク格付を付与する。モデルのリスク格付は、モデル・リスク管理におけるリスクベース・アプローチの基礎として、各モデルに対する統制のレベル（検証の深度や頻度等）を決定する重要な要素となる"
3	モデル開発	"開発プロセスでは、モデルで用いられる手法や仮定、モデルの限界等がモデル・リスク管理におけるステークホルダーに十分かつ確実に伝達されるように、包括的なモデル記述書を作成する"
4	モデル・テスト、モデル検証、承認	"モデルの使用開始前には、第1線がモデル・テスト（開発時の点検）を行う。また、第2線が独立した立場から検証を行い、検証結果を踏まえてモデルの使用を承認する"
5	継続モニタリング再検証	"モデルの使用開始後には、第1線がモデルに対する継続モニタリング、第2線が再検証を実施し、モデルが意図したとおりの性能を示すことが出来ているかを評価する。欠陥等が発見されたモデルについては、第2線が使用を制限又は拒否する権限を持つ"
6	モデル変更	"モデルに重要な変更が加えられた場合には、必要に応じて追加的な検証を実施する"
7	モデル・リスク評価	"第2線は自社全体のモデル・リスクを評価し、結果を取締役会等に報告する"
8	内部監査	"内部監査を担う第3線は、金融機関のモデル・リスク管理態勢の全体的な有効性を評価する"
9	文書化	"上記すべてのプロセスを方針・規程によって明確化し、各プロセスにおける結果は適切に文書に残しておく"

モデルリスクに関する原則イメージ図



（出典）「モデル・リスク管理に関する原則」(金融庁、令和3年11月）（https://www.fsa.go.jp/common/law/ginkou/pdf_02.pdf)を基にKPMG作成

AIガバナンス体制上の論点

AIガバナンスに係る規程の位置付け	自社のデータやAI利活用の特徴をふまえ、以下のようなケースがある。 ✓ モデルリスク、オペレーションリスク、システムリスク等のリスク管理に紐付けるケース ✓ プライバシー、公平性等、コンプライアンスに紐付けるケース ✓ システム開発に紐付けるケース ✓ 独立した位置付けとするケース
AIガバナンスの推進部署	位置付けに応じ、推進部署は決定される。共同所管をしいているケースもある。 ✓ リスク管理部門。欧米ではモデルリスク管理部門が所管するケースが多くみられる ✓ コンプライアンス部門 ✓ IT企画部門 ✓ デジタル企画部門
部門横断の体制	AIリスクは多岐にわたるため、部門横断の体制を敷くケースが多くみられる。 ✓ ITに係るため、デジタル企画部門、IT企画部門、IT基盤部門 ✓ 規制に係るため、リスク管理部門、コンプライアンス部門、法務部門 ✓ 倫理に係るため、人事部門、サスティナビリティ部門

これからのAIガバナンスの論点

■ AIの足元の状況をふまえると、「開かれたガバナンス」、および「アジャイルなガバナンス」の実現が論点となり得る

足元の状況	今後のAIガバナンスの論点	
AIを用いた類似サービスの林立	AIガバナンスの開示による利用者に対する信頼獲得	開かれたAIガバナンス
AIに係るサービス提供は単独ではなく、バリューチェーンを通じて行われる	バリューチェーンの評価、外部との対話・連携の組み込み	
わが国においては、（高度な）AI利活用はまだこれからの状況	バリューチェーンのAIリスクに基づくAIガバナンスの成熟	
ビジネス動向、規制動向、インシデント等の発生により、AIに対する世論は大きく変わり得る	AIに係る社会的な受け止めに応じたアジャイルな対応	アジャイルなAIガバナンス
データの変化によるモデル陳腐化、データやモデル汚染の攻撃等によるモデルの誤動作	継続的なモニタリング・対応の仕組み・基盤整備	

KPMG

KPMGコンサルティング株式会社
PRINCIPAL
津田 圭司
E: keiji.tsuda@jp.kpmg.com