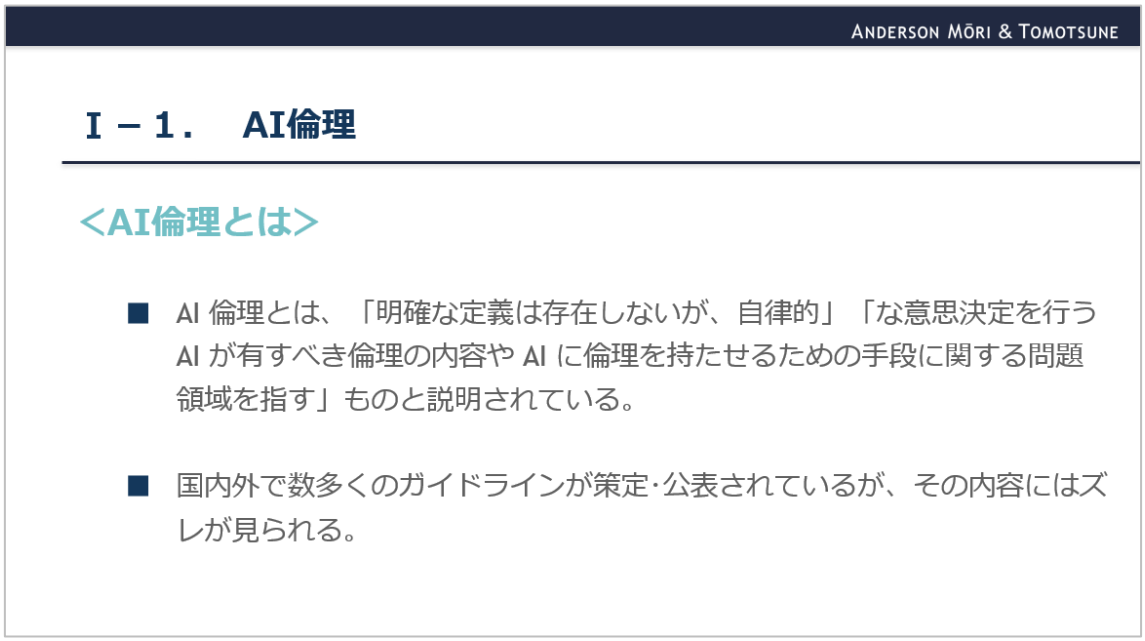
第 9 回　AI 倫理・AI ガバナンスと医療の邂逅
(アンダーソン・毛利・友常法律事務所外国法共同事業　弁護士　中崎　尚)

1. AI 倫理とは

AI 倫理について、生成 AI が中心に議論をされることが多いが、AI ガバナンスに関しては生成 AI に限らず、AI 全体を幅広くに議論されることが多いです。



AI 倫理は、医療関係や生命倫理は異なる独自の概念です。明確な定義はないが、「AI が有すべき倫理的な内容や AI に倫理を持たせるための手段に関する領域を指す」と説明されています。

AI 倫理に基づいて、例えばヨーロッパの「trustworthy」のガイドラインや日本国内で総務省が発表した AI 倫理のガイドラインなど、様々な国で策定しています。

ANDERSON MŌRI & TOMOTSUNE

Ⅰ－１．　AI倫理

＜国内のAI倫理関連のガイドライン・原則＞

- 「国際的な議論のための AI 開発ガイドライン案」（通称：AI 開発ガイドライン）（総務省 AI ネットワーク社会推進会議　2017年）
- 「人間中心の AI 社会原則」（内閣府　2019年）
- 「AI 利活用ガイドライン」（総務省 AI ネットワーク社会推進会議　2019年）
- 「AI 事業者ガイドライン（第1.0版）」（総務省・経済産業省　2024年４月）

2017 年に総務省が AI 倫理に関するガイドラインを発表して以来、内閣府や経産省からも複数のガイドラインが公表されています。2024 年の 4 月に、これらを統合するため、「AI 事業者ガイドライン（第 1.0 版）」が策定・公表されました。

ANDERSON MŌRI & TOMOTSUNE

Ⅰ－１．　AI倫理

＜国外のAI倫理関連のガイドライン・原則＞

- 「Tenets」（Partnership on AI　2016年）
- 「アシロマ AI 原則（Asilomer AI Principles）」（Future of Life 2017年）
- 「信頼できる AI のための倫理ガイドライン（Ethics guidelines for trustworthy AI）」（EU　2018年）
- 「人工知能に関する理事会勧告（Recommendation of the Council on Artificial Intelligence）」（OECD　2019年）

海外でも、様々な研究機関、NPO、学会等で多数のガイドラインが作られています。その中でも特に注目されるのが、EU の「信頼できる AI のための倫理ガイドライン（Ethics guidelines for trustworthy AI)」です。これは、2024 の 7 月にできた EU AI Act の内容に反映されており、一つのデファクトスタンダートと位置づけられています。

同様に広く普及しているのは、OECD が出している AI に関する勧告です。

一つは「女性差別」の例です。大手の通販会社が AI を使って人材雇用を試みたところ、男性偏重の結果が出てしまい、システムの採用を断念しました。意図的に性別差別を行うためではなく、過去の採用例を学習させた結果、男性ばかり採用していたため、AI が「女性」というキーワードが含まれる履歴書を自動的に除外した事案でした。

ANDERSON MŌRI & TOMOTSUNE

Ⅰ－２．　AI倫理が問題となった事例

＜事例１：「女性差別」（公平性･バイアス）＞

■　（続き）あるいは「女子テニス部の副部長」といった経歴が記されていると評価が下がる傾向が確認されたのである。このようなエラーが発生した原因としては、AI の学習モデルに、同社に過去10年間にわたって提出された履歴書のパターンを学習させたところ、実際には技術職のほとんどが男性からの応募であった結果、AI は技術職には男性を採用するのが好ましいと学習したのであろうと推測されている。

このように不公平や差別を AI に学習させることが良いのかは難しい問題です。差別の再生産は避けるべきとする議論がある一方で、AI に現実社会の状況を変える役割を求めるべきかという声もあります。この「公平性・バイアス」の問題はまだ解決していません。

ANDERSON MŌRI & TOMOTSUNE

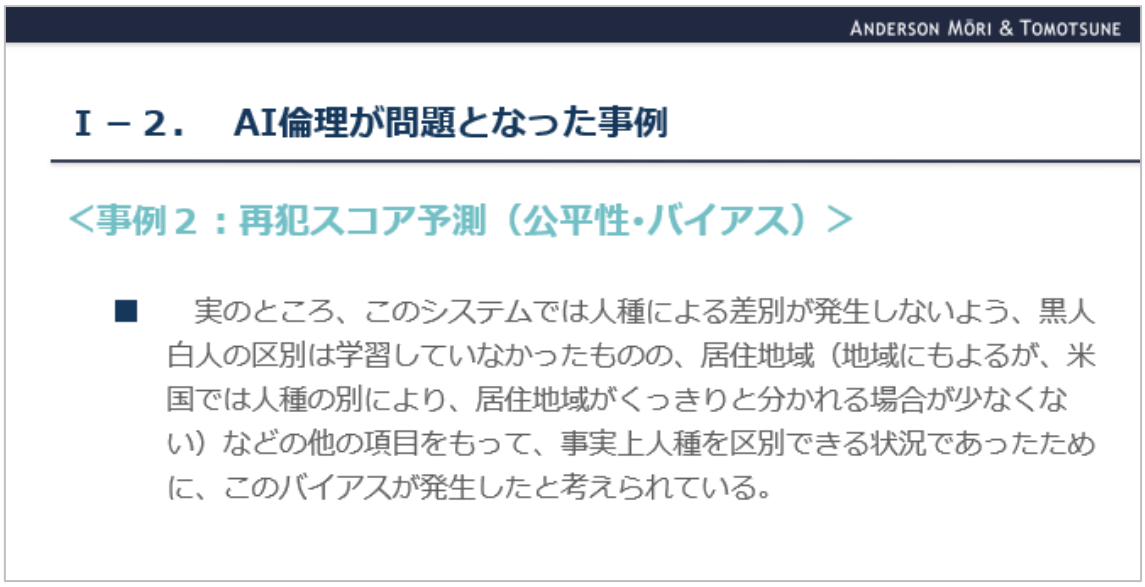
Ⅰ－２．　AI倫理が問題となった事例

＜事例２：再犯スコア予測（公平性･バイアス）＞

■　米国の裁判所において使用されていた COMPAS という事案管理・判断予測支援のシステムでは、刑事被告人の再犯スコアを出力する機能が備わっていた。裁判所によっては、この再犯スコアを勘案して判断を下していた。

■　2016年、このシステムによる再犯スコアの算出において、高リスクと判断していたのに実際は再犯がなかったケースが黒人に高い割合で見られ、逆に低リスクと判断していたのに実際には再犯してしまったケースが白人に高い割合で見られる傾向が確認されたことが報じられ、問題となった。

「再犯スコア予測（公平性･バイアス）」も、アメリカの司法制度で問題になっています。システム導入後、算出した再犯スコアと実際の再犯率を照合すると、うまくいっていないことが判明し、このシステムは本当に適切なのかという問題になった事案です。



また、人種差別を避けるために、黒人や白人という区別を学習させていないものの、居住地域の情報は登録していたことが仇になったケースもあります。アメリカでは都市の中心部に黒人や低収入者が住み、周辺部に高収入者や白人が住む傾向があります。これが暗黙のうちに反映され、結果的に人種を区別できるデータを学習してしまい、バイアスが発生したと指摘されています。

ANDERSON MŌRI & TOMOTSUNE

Ⅰ－２．　AI倫理が問題となった事例

＜事例３：顔認識技術による誤認逮捕（公平性・バイアス）＞

■　いわゆる、顔認識の「人種バイアス」問題に由来する事例である。

■　2023年８月、米ミシガン州デトロイト市で当時妊娠８カ月だった黒人女性が、AI を用いた顔認識システムによって強盗犯と誤って判断され、逮捕されてしまい、後日、女性が不当逮捕だったとして市を提訴したことが報じられた。　デトロイト市警察は、監視カメラに映った強盗犯の男女の身元を割り出す過程で、AI 顔認識技術を使用したころ、ある黒人女性の８年前の写真と一致するという結果が得られ、実際に逮捕に至ったものの、後日、誤認逮捕であることが判明したものである。

「顔認識の『人種バイアス』」も問題になります。監視カメラに映った強盗犯を特定するために AI を使用したところ、誤認逮捕が発生した事案です。特に、誤認逮捕された女性が妊婦だったことが問題になりました。

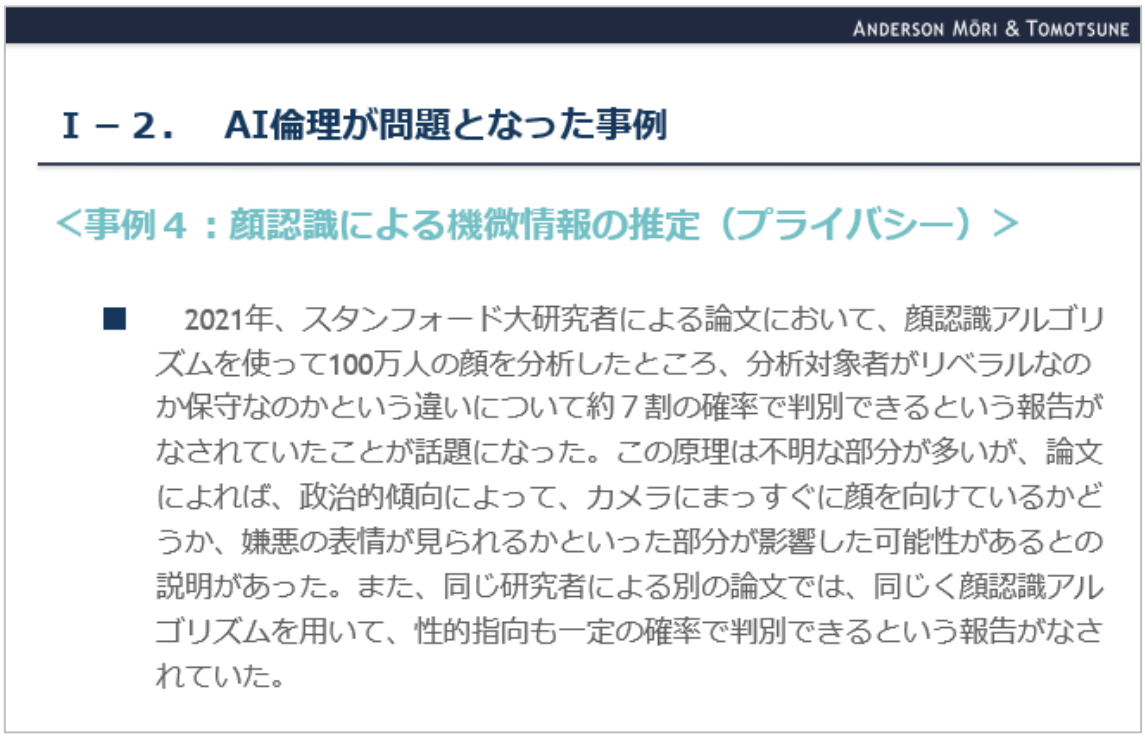
ANDERSON MŌRI & TOMOTSUNE

Ⅰ－２．　AI倫理が問題となった事例

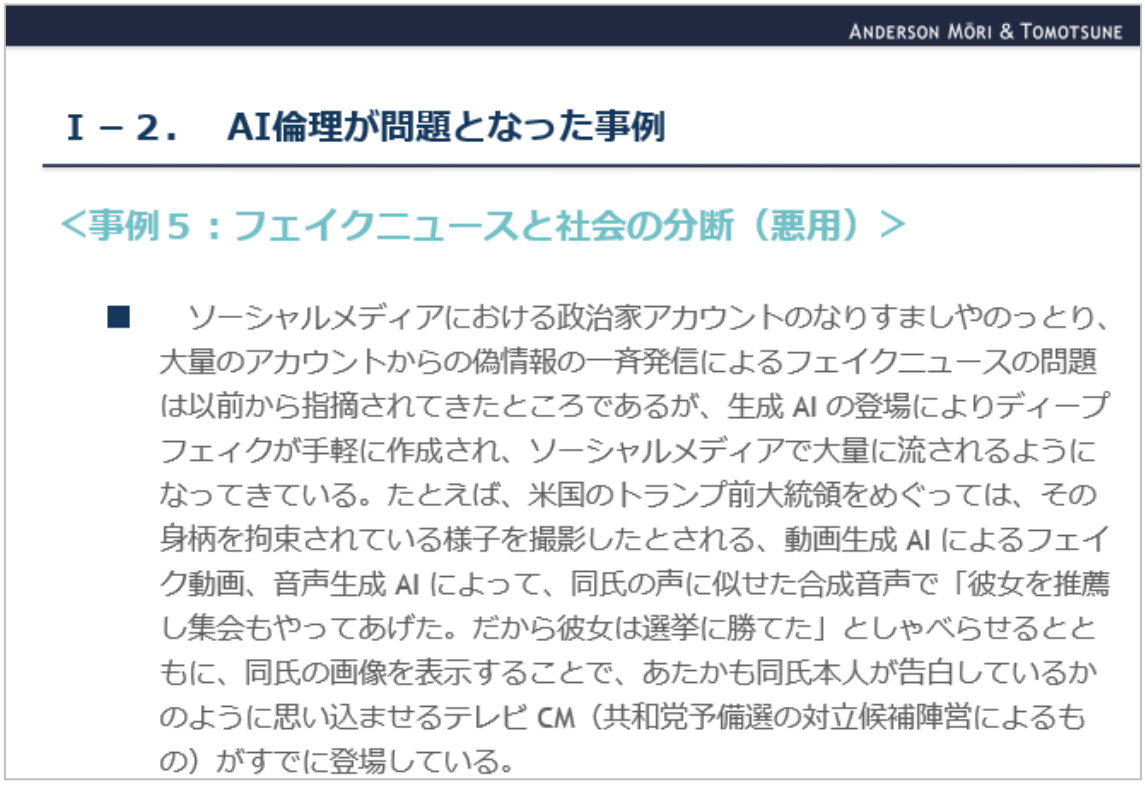
＜事例３：顔認識技術による誤認逮捕（公平性・バイアス）＞

■　警察による AI 顔認識システムの欠陥が明らかになったのはこれが初めてではない。とくにデトロイト市警察はこの技術によって2019～20年に黒人男性３人を誤認逮捕しており、AI による顔認識のアルゴリズムにはバイアスがあり、警察による人種差別を助長しているという問題の根深さが改めて浮き彫りになったとも指摘されている。このような悲劇を避けるべく、米国では、2022年９月、連邦議会において AI の顔認識アルゴリズムの透明性を規制する初の試みとして「2022年顔認識法」案が提出されている。

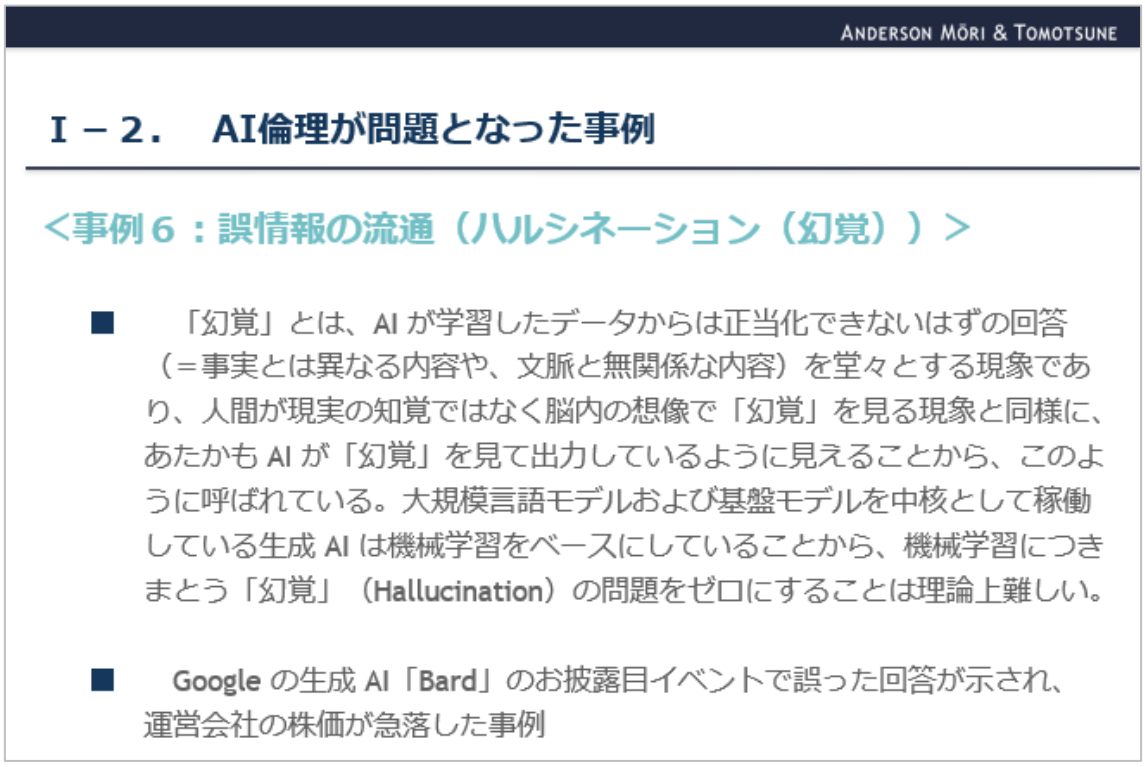
結果的に、黒人が誤認逮捕されやすい状況が AI の顔認識アルゴリズムで発生し、人種差別を助長、公平性やバイアスの原則を侵害する問題になりました。これを受けて 2022 年 9 月に連邦議会で規制法案が提出されたものの、法律としてはまだ成立していません。現在、顔認識技術は世界中で広く使われ、特に捜査機関で多用されているが、ルールなしでの使用が問題視されています。特にヨーロッパでは、顔認識技術の無茶な使用を防ぐ動きが以前からあり、EU AI 規則でも規制するなど、AI 倫理から法律の規制につながった具体例が見られます。



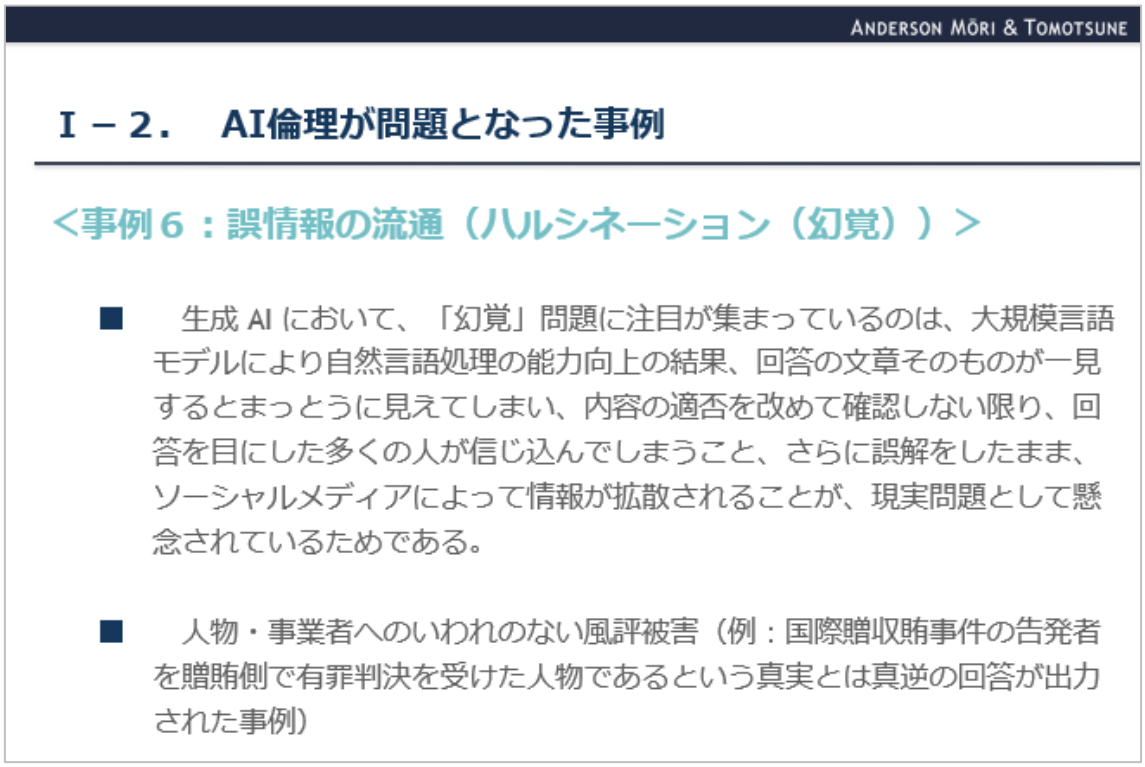
同じ顔認識でも、個人の政治的な信条や性的な趣味などを推定できるとスタンフォードの研究者が発表した事案がありました。このケースでは、このような推論に AI を使うこと自体が大きく問題になりました。



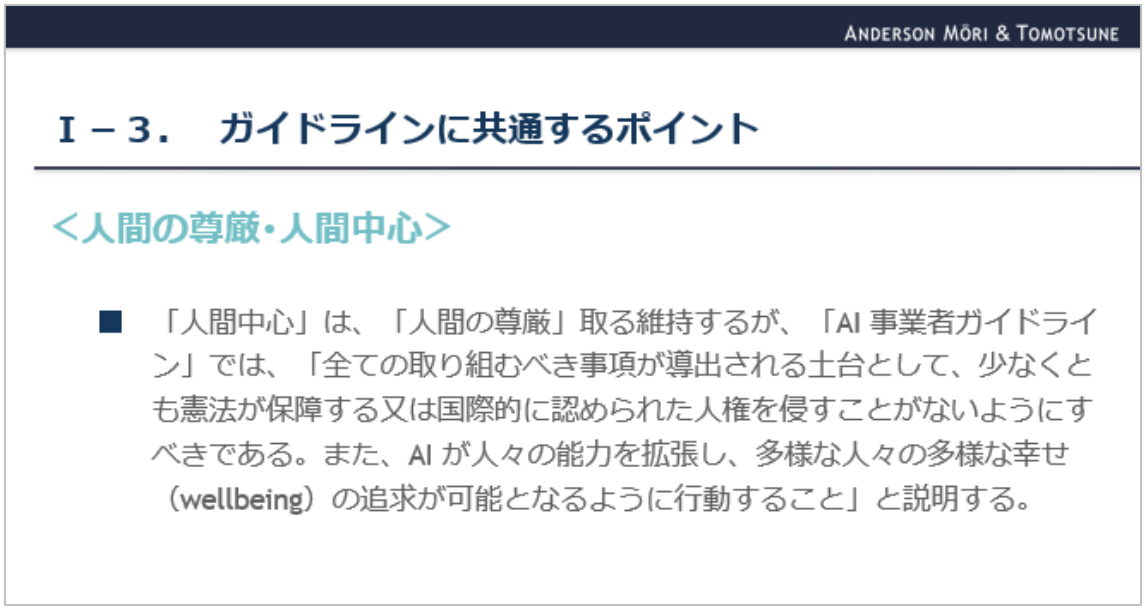
次は「フェイクニュース」の事案です。2024 年のアメリカ大統領選で、トランプ前大統領やハリス候補の顔や声を無断で使用し、虚偽の動画や音声を作成してテレビやインターネットで流す行為が一般的に行われました。AI が社会の分断のために悪用されることが問題になっています。



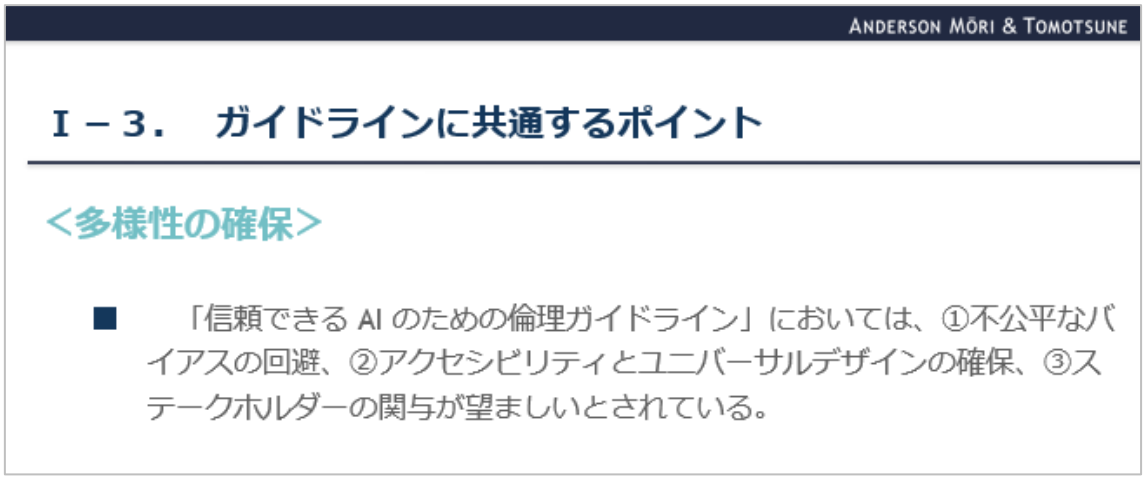
最後の事例は「ハルシネーション」の事案です。AI が事実と無関係な内容や文脈を勝手に回答してしまうことを指します。AI は確率論でそのような回答をするのが得意で、誤りが普通に起きます。例えば、Google の生成 AI のお披露目イベントで、回答が大きく間違っており、株価が下落したケースが実際にありました。



特に生成 AI においては、ユーザーリテラシーが重要です。EU の AI 規則でも「AI リテラシー」が取り上げられており（第 4 条）、リテラシーの教育が不十分な中で、AI の出力情報を無断で拡散することが大問題になっています。例えば、国際的な贈収賄事件で無実の人物が有罪と誤報されたケースがあります。AI は情報の正誤を判断せずに出力するため、開発者は注意を払い、学習用データの質を高めるべきという議論が行われています。特に高度なリスクを抱える AI には、より一層注意を払うべきだということが、G7 の国際会議で議論され、共同宣言が出されました。また、EU AI 規則でも厳しい規制が課せられています。このように、AI ガバナンスに法規制をかけ、研究開発サイドにもその姿勢を共有し、問題の発生を防ぐことが求められています。



ガイドラインに共通するポイントの一つ目は、「人間の尊厳」や「人間中心の原則」です。人権の侵害の防止、AI を人々の手助けツールとして能力を拡張すること、多様な幸福の追求が挙げられます。



共通するポイントの二つ目は「多様性の確保」です。EU の「信頼できる AI のための倫理ガイドライン」では、アクセシビリティやユニバーサル・デザインの確保が重要視されています。AI の多様性を確保することが強く求められています。

ANDERSON MŌRI & TOMOTSUNE

Ⅰ－３．　ガイドラインに共通するポイント

＜安全性＞

■　「AI 事業者ガイドライン」においては、「ステークホルダー（事業外利用者等及び第三者を含む）の生命・心身・財産、精神及び環境に危害を及ぼすことがないよう努めること」と説明されている。

共通するポイントの三つ目は「安全性」です。昨年 11 月、イギリスのブレッチリー・パークで、AI の安全性に関するサミットが行われ、各国が共同で AI の安全性の確保することが強調されました。国連の諮問機関も AI の安全性が重要性を述べ、各国で「AI Safety Institute」が設立されています。

AI はステークホルダー（利害関係者）の生命・身体・財産、精神、環境等を害する懸念があるため、このような機関の設立が進んでいます。

ANDERSON MŌRI & TOMOTSUNE

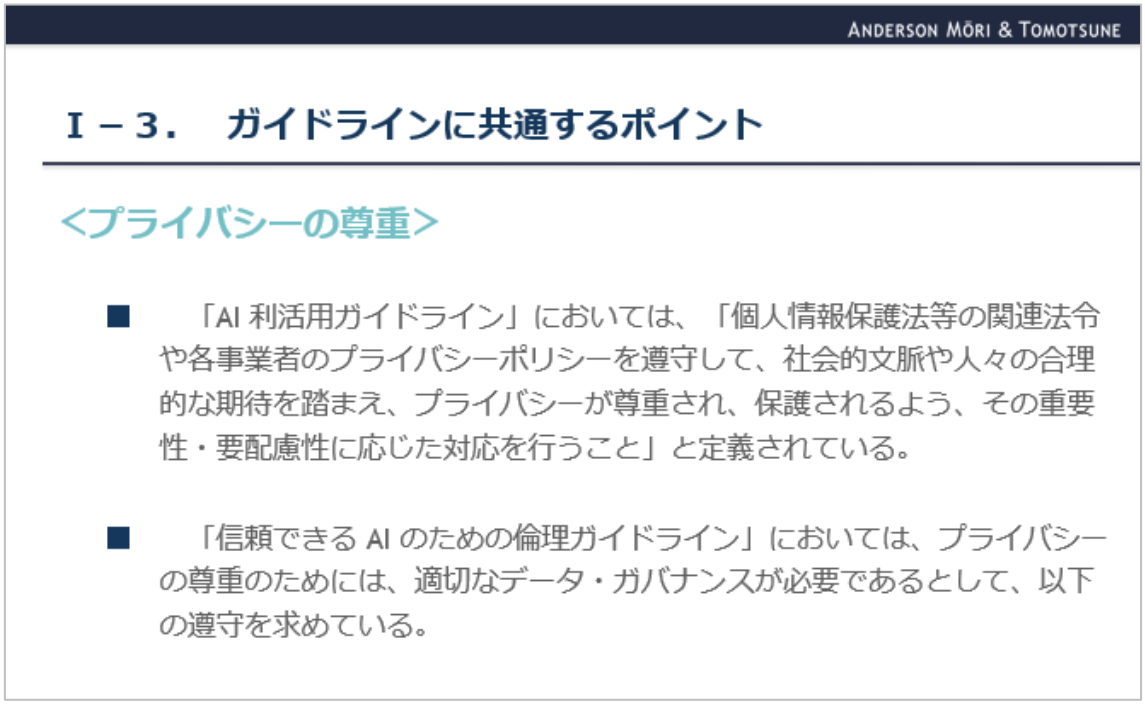
Ⅰ－３．　ガイドラインに共通するポイント

＜セキュリティ＞

■　「AI 事業者ガイドライン」においては、「不正操作によって AI の振る舞いに意図せぬ変更又は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こと」と説明されている。

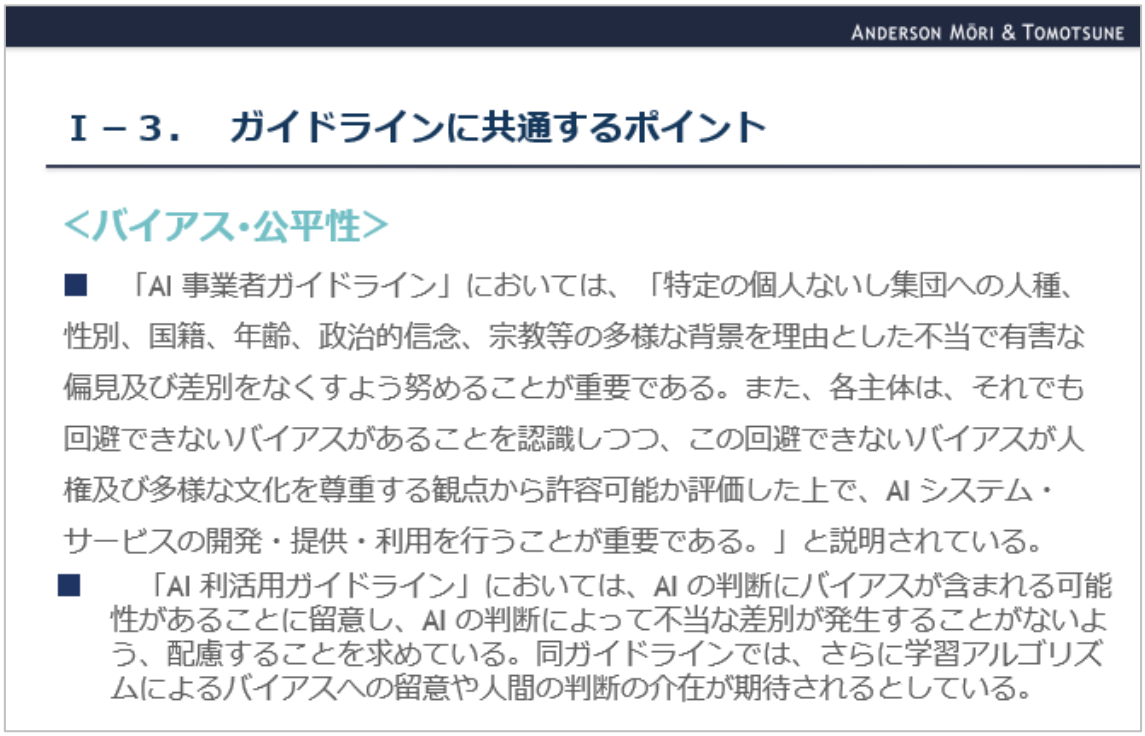
■　「AI 利活用ガイドライン」では、①セキュリティ対策の実施、②セキュリティ対策のためのサービス提供等、③ AI の学習モデルに対するセキュリティ脆弱性への留意を期待するとしている。

四つ目は、「セキュリティ」です。これは安全性と区別され、第三者からの攻撃への対応が議論されます。ハッキングされると偏った回答結果が生成し、人を傷つける恐れがあります。そうした事態を防ぐために、AI 事業者、特に開発やサービスの提供事業者には対策が要求されています。

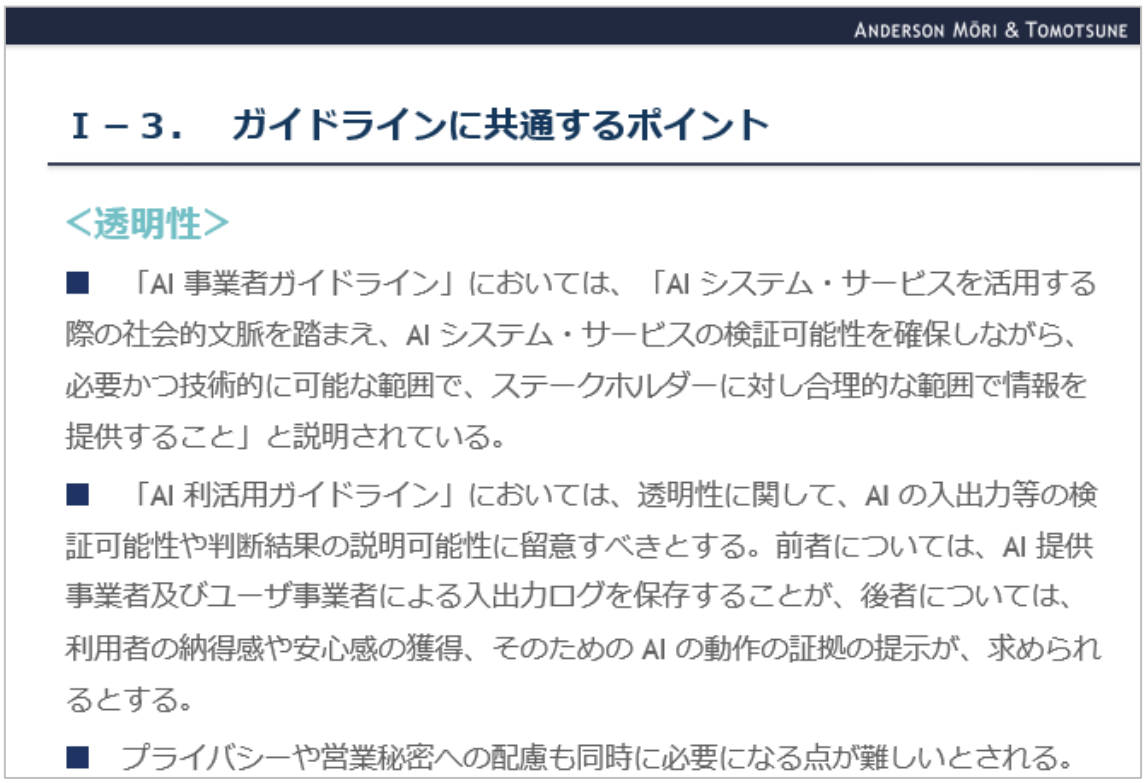


五つ目は「プライバシーの尊重」です。個人情報保護法等の文脈や各事業者のプライバシーポリシーがあります。データガバナンスの必要性が EU から強調されています。

プライバシーの尊重について、①は、違法目的や差別目的の利用をしないことが大事です。②は、収集されたデータを学習させる前にバイアス等を訂正することです。しかし、現在の大規模言語モデルではバイアス訂正は難しいです。③は、アクセス権の問題です。誰が、どのような状況でアクセスできるか明確に定めることが必要です。



六つ目は「バイアス・公平性」です。「AI 事業者ガイドライン」では、人種、性別、国籍、政治的信条や宗教等による偏見や差別をなくすことが重要です。また、回避できないバイアスがあることを認識しつつ、人権や多様な文化を尊重する観点から許容可能なのかを評価し、AI システム・サービスの提供や開発に活かしていくことも重要と説明されています。また、「学習アルゴリズムによるバイアスへの留意」として、差別の再生産を防ぐために人間の判断を介在させることが期待されています。



七つ目は「透明性」です。「AI 事業者ガイドライン」では、AI システム・サービスの検証可能性を確保しながら、合理的な範囲で情報を提供すべきと説明されています。「AI 利活用ガイドライン」も、AI の入出力等の検証可能性や判断結果の説明可能性に留意すべきとしています。特に入出力のログに関しては、利用者が納得できるよう証拠の提示が求められているものの、事業者にとって営業秘密に当たるため、全部開示するのが厳しいとされます。

ANDERSON MŌRI & TOMOTSUNE

Ⅰ－３．　ガイドラインに共通するポイント

＜アカウンタビリティ＞

■　「AI 事業者ガイドライン」においては、トレーサビリティの確保、他の 10大原則の対応状況等について、「ステークホルダーに対して、各主体の役割及び開発・提供・利用する AI システム・サービスのもたらす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ことが重要である」と説明されている。

■　「AI 利活用ガイドライン」においては、アカウンタビリティを果たす努力、AI に関する利用方針の通知・公表が求められている。

■「信頼できる AI のための倫理ガイドライン」においては、「①監査可能性、②悪影響の最小化と報告、③適切なトレードオフ、④適切な是正」が挙げられている。

八つ目は「アカウンタビリティ」です。この言葉は論者によって、異なる意味合いで使われがちです。EU の「信頼できる AI のための倫理ガイドライン」では、監査可能性や悪影響を最小化することなどが挙げられています。

ANDERSON MŌRI & TOMOTSUNE

Ⅰ－３．　ガイドラインに共通するポイント

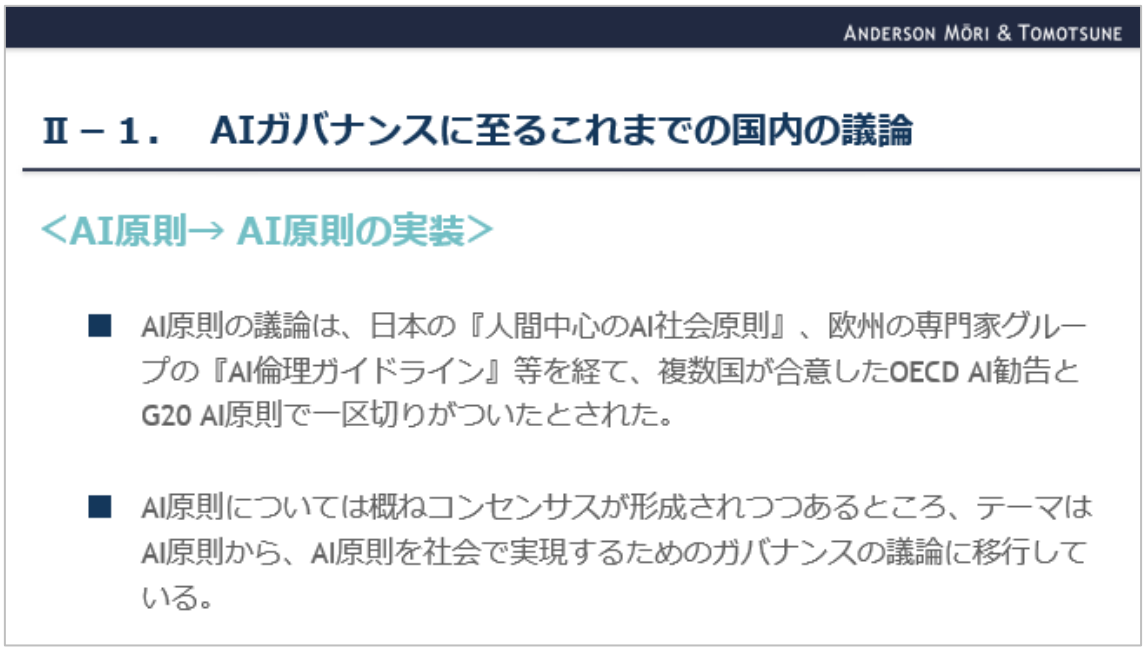
＜アカウンタビリティ＞

■　アカウンタビリティについては、論者による理解が大きく異なるが、中心となる要素としては「説明可能性」「理解可能性」「追跡可能性」「答責性」「透明性」「説明責任」「補償の枠組み、救済」が挙げられている。

中心要素としては「説明可能性」、「追跡可能性」、「補償の枠組み、救済」などが挙げられています。

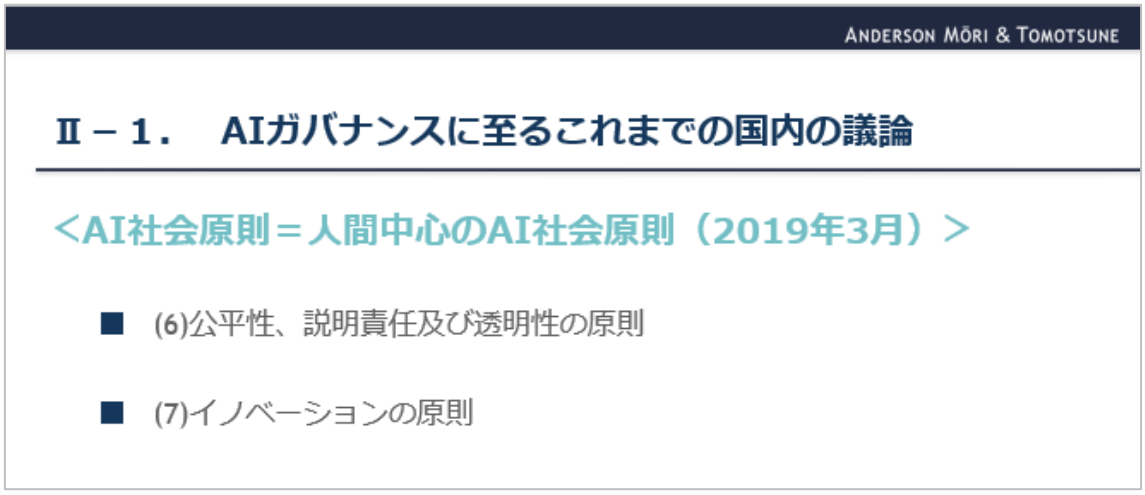
2. AI ガバナンスとは

AI 原則は、様々なガイドラインで述べられたが、これを文章化するために作成されたのが「AI 事業者ガイドライン」です。

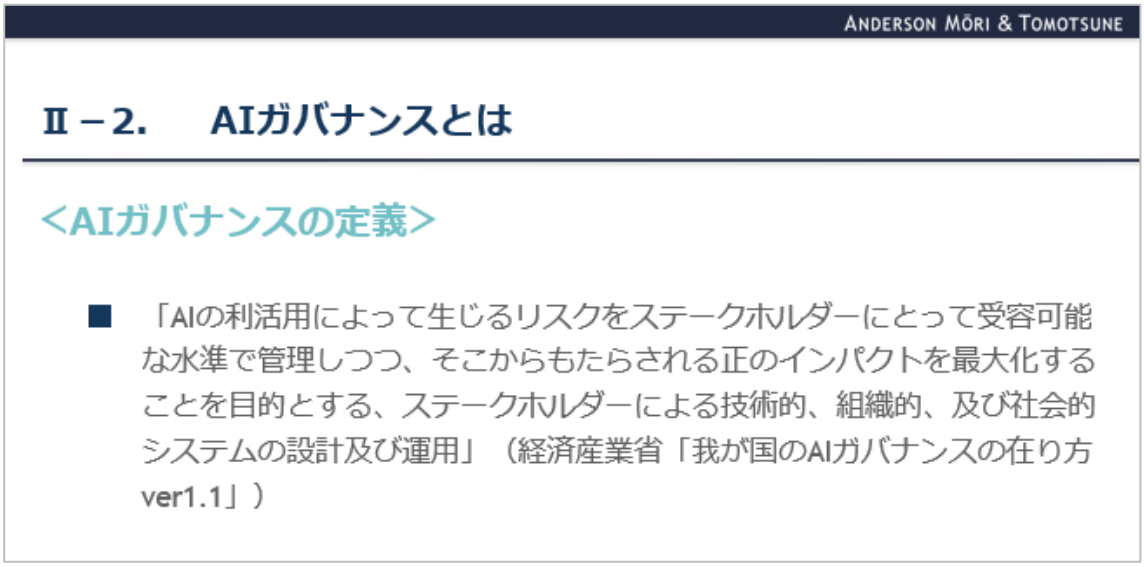


「原則」からその「実装」へと移行する過程は国によって異なります。ヨーロッパはガバナンスをハードローに落とし込み、日本はソフトロー路線を採用しました。アメリカは 2023 年 10 月に当時のバイデン大統領が大統領令を発表したほか、大手 AI 事業者から協力を取り付けることで合意していました。日本では、2024 年の 7 月に AI 戦略会議の下に新たに AI 制度研究会が設立されました。特に、「安全性」の問題について、日本はソフトロー一辺倒では不十分という議論と見直しが行われています。

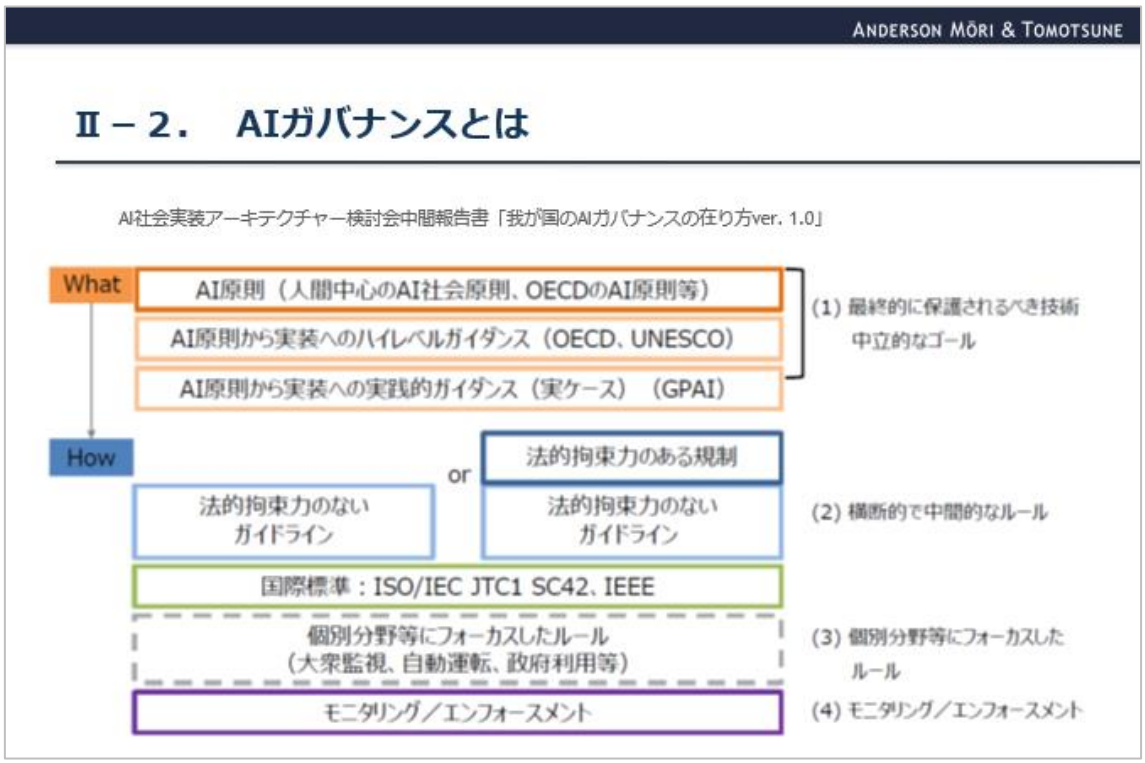
現状、ガバナンスについては、「人間の尊厳が尊重される社会（Dignity）」「多様な拝啓を持つ人々が多様な幸せを追求できる社会（Diversity & Inclusion）」「持続性ある社会（Sustainability）」という三つの基本理念を中心に、日本国内で議論されています。



2019 年 3 月に出された「人間中心の AI 社会原則」の中に、「人間中心」、「教育・リテラシー」、「プライバシー」、「セキュリティ」、「公正競争」、「公平性、説明責任及び透明性」、「イノベーション」という 7 つの原則あります。

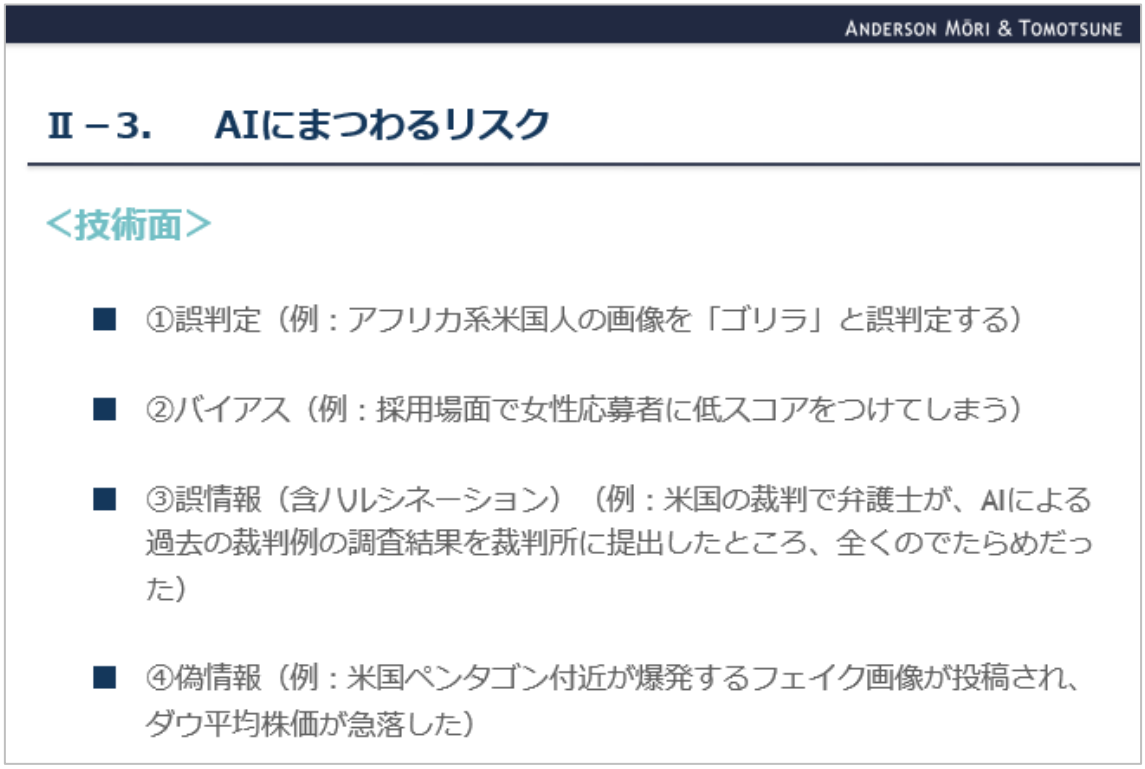


AI ガバナンスの定義は、「正のインパクトを最大化する」ことです。AI 使用に伴うリスクを適切かつ許容可能なレベルに抑えつつ、メリットが最大化するために、技術的、組織的、社会的に検討していくことが、AI ガバナンスだと定義されています。



国によって、法的拘束力のある規制を設ける場合とガイドラインにとどめる場合があります。アメリカの NIST と日本、欧州の機関が合同でその在り方を議論していま

す。また、AI を使った自動運転などの個別分野にフォーカスしたルールも発展しています。



具体的に、AI にまつわるリスクは「技術面」と「社会面」に分けて分類しています。

「技術面」では、「①誤判定」、「②バイアス」、「③誤情報」、「④偽情報」が含まれます。「③誤情報」については、アメリカの裁判所での事例があります。弁護士が生成 AI のチャット GPT に問合わせた結果、ハルシネーションで誤った情報を生成し、それを裁判所に提出して懲戒を受けました。

「④偽情報」については、例えば株価に影響を与えるために、ホワイトハウスやペンタゴンが空爆された偽映像を作成し拡散することで、一瞬株価が下落し、その後に偽情報だと判明すると再び上がるため、短期間で大きな利益を得ることが可能です。

ANDERSON MŌRI & TOMOTSUNE

Ⅱ－3.　AIにまつわるリスク

＜技術面＞

- ⑤安全性（例：自動運転など、AI制御のシステムが物理的事故を起こす）
- ⑥セキュリティ（例：チャットボットTayに、悪意ある会話データを学習させることで、問題発言を頻発させるに至った）

「⑥セキュリティ」に関しては、チャットボットに悪意ある会話データを学習させ、問題発言を頻発させた事案が実際にありました。その結果、サービスの休止に追い込まれました。エンドユーザーがチャットボットと会話する中で問題発言を繰り返し、その内容を学習させることで、チャットボット側も問題発言がありました。

ANDERSON MŌRI & TOMOTSUNE

Ⅱ－3.　AIにまつわるリスク

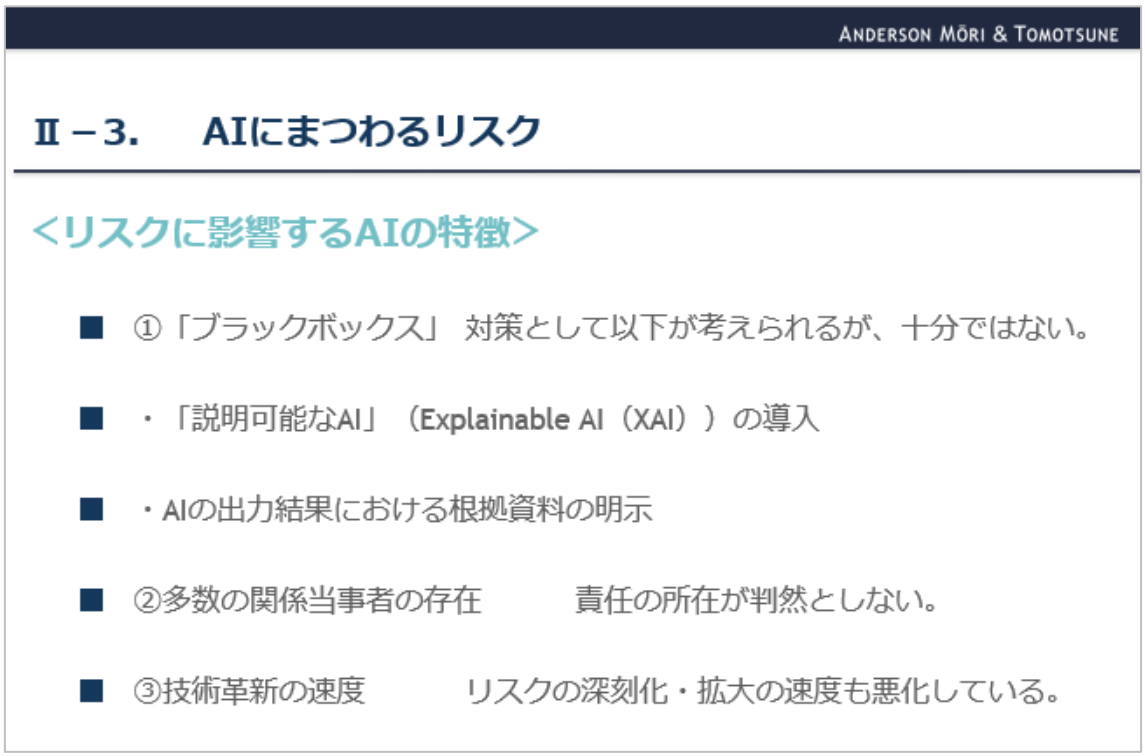
＜社会面＞

- ①プライバシー（プロファイリング（例：米国の大手スーパー「ターゲット」が購買行動を分析して、妊娠予測スコアを算出し、クーポンを送付したところ、女子高生が含まれていた）、行動の抑制・誘導（例：映画「マイノリティ・レポート」の世界）、スコアリング（例：リクナビ、Yahoo!スコア））
- ②民主主義への悪影響（例：ソーシャルメディアの表示AIアルゴリズムによって発生する「フィルターバブル」による世論の分断）

「社会面」について、「①プライバシー」の問題 AI によるプロファイリングのリスクや行動の抑制・誘導があります。日本では「リクナビ」のように、無駄にスコアリングを行われることが問題になります。

「②民主主義への悪影響」は、分断の話題として、特に「フィルターバブル」が問題視されています。

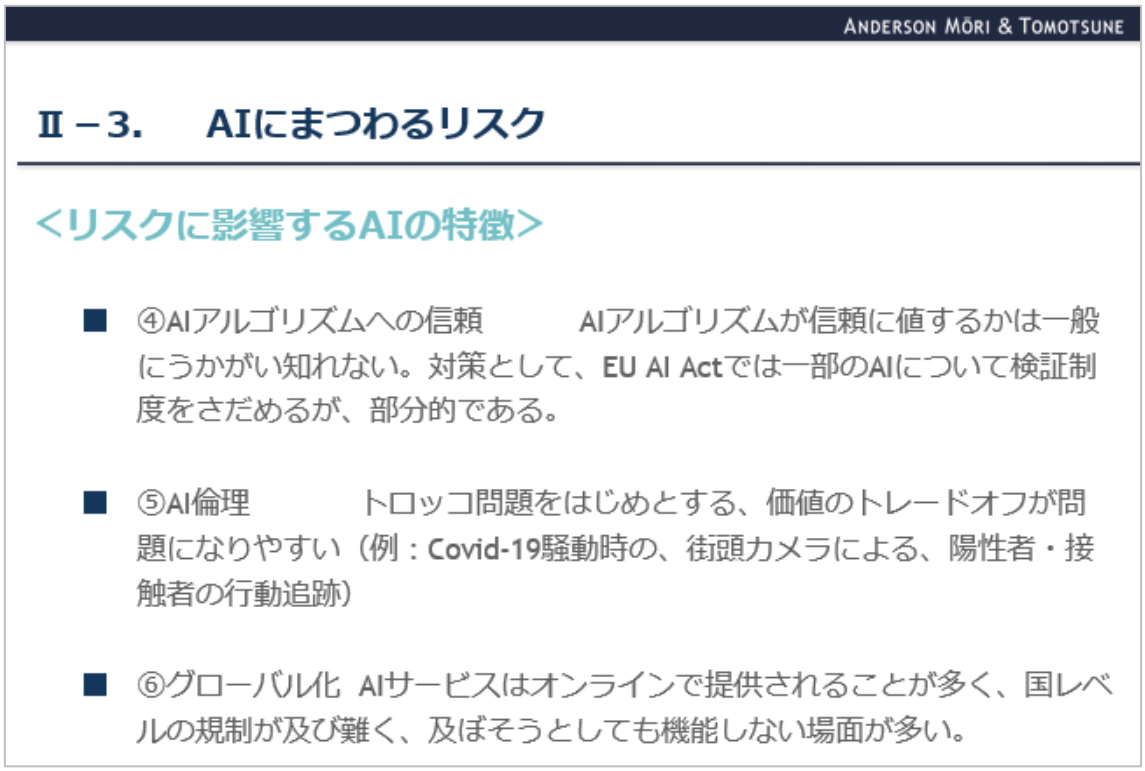
「③不正利用・攻撃利用」と「④経済への悪影響」があります。「④経済への悪影響」は雇用の喪失や GAFAM に資源が過度に集中することが挙げられます。「⑥電力問題（環境への悪影響）」は WHO のレポートで強調され、最終的には環境への悪影響が人の生命や健康を害する可能なため、大きな社会問題です。



リスクに影響する AI の特徴について、「①ブラックボックス」が問題です。数年前に機械学習で AI が注目を浴びた際に、AI がどのようにアルゴリズムを使っているかが開発者にも分からないまま、性能だけが向上する現象が見られました。人間は AI コントロールできないリスクが発生する懸念があります。対策として「説明可能な AI の導入」が重要視されています。AI 出力結果における「根拠資料の明示」が重要です。

「②多数の関係当事者の存在」があり、責任の所在が判断しにくくなり、権利の帰属が不明確なため、紛争になりやすいです。

「③技術革新の速度」は、迅速すぎて制度が追いつかない点も問題視されています。



「④AI アルゴリズムへの信頼」です。人間側は、AI が信頼に値するかは判断しにくいため EU AI Act の一部は開示制度や検証制度を定めました。しかし、本当に危険なものに限られたため、問題は十分に解消されないのではと言われています。

「⑤AI 倫理」です。これは「トロッコ問題」のように、人によって回答が異なる場面が多く存在します。例えばコロナの時に、陽性者の行動追跡を社会的に許容されるかどうかは人によって意見が分かれました。特に AI に関する場合は、人の価値観によって結論が分かれることが多いです。

「⑥グローバル化」です。AI サービスはオンラインで提供されることが多いため、特定の国や法域で規制を強化しても、追いつきません。危険な部分に対して法律的な規制をかけても、規制の緩い国が存在する限り、完全にカバーしきれない問題が指摘されています。

ANDERSON MŌRI & TOMOTSUNE

Ⅱ－3.　AIにまつわるリスク

<リスクに影響するAIの特徴>

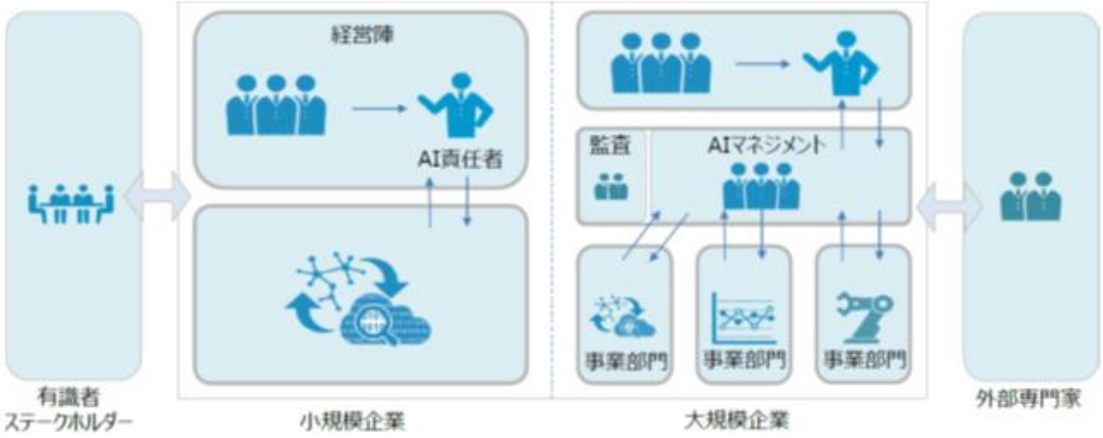
■　⑦汎用人工知能（AGI）の登場　　シンギュラリティの懸念。

7つ目は「シンギュラリティ」です。

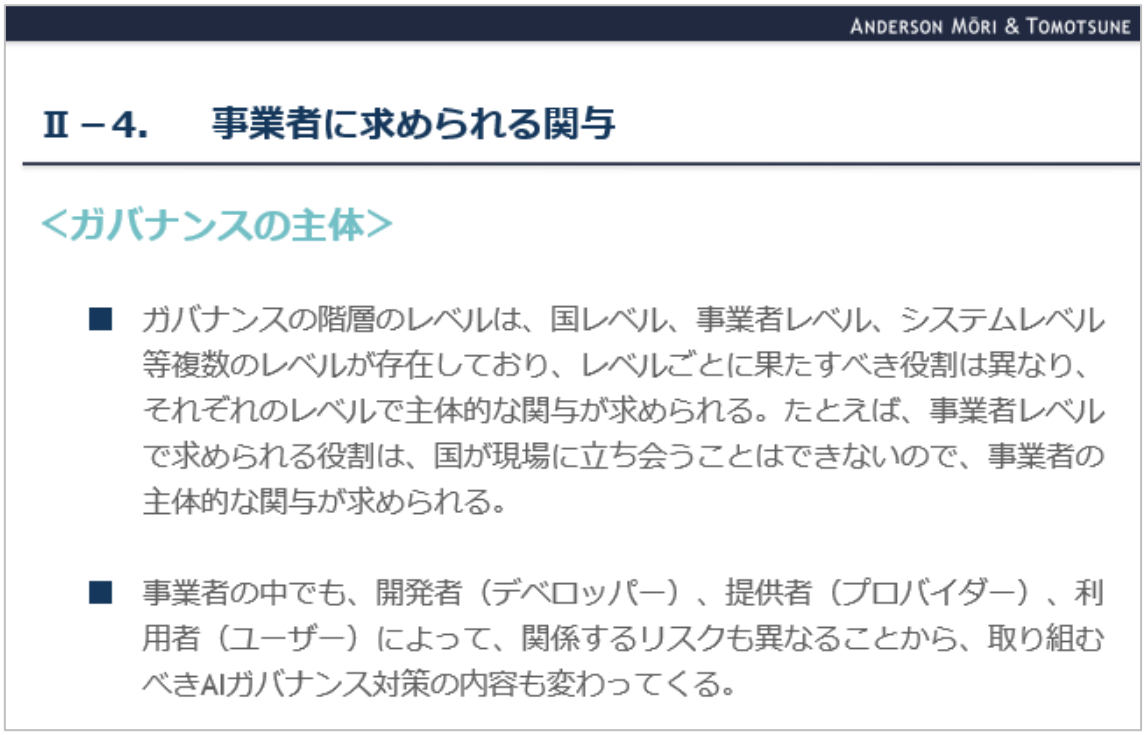
ANDERSON MŌRI & TOMOTSUNE

Ⅱ－4.　事業者に求められる関与

AI社会実装アーキテクチャー検討会中間報告書「我が国のAIガバナンスの在り方ver. 1.0」



AI事業者がどのようにAIを運用していくべきかについては、「AI事業者ガイドライン」が注意を喚起しています。ここでは事業者の規模によって異なることが指摘されています。例えば、医療分野においても、医療機関と町の小さな診療所では状況が大きく異なります。



「ガバナンスの主体」については、国や事業者など様々なレベルがあり、国がすべてを実行するのは難しいため、実際には事業者が主体的に対応する必要があります。

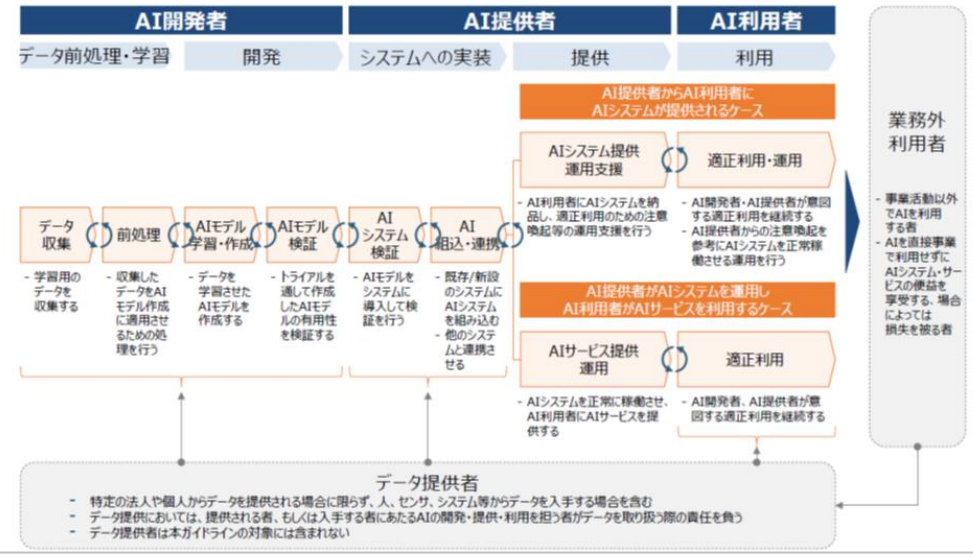
加えて、同じ事業者内でも担当者と経営陣レベルなど、対応や役割が異なります。

3. 事業者と AI

ANDERSON MŌRI & TOMOTSUNE

Ⅲ－2.　事業者のAI関与のあり方

AI事業者ガイドライン検討会（第3回）資料2-2「『AI事業者ガイドライン案』概要」



事業者の役割は、大きく「開発者」、「提供者」、「利用者」に分類できます。

ANDERSON MŌRI & TOMOTSUNE

Ⅲ－2.　事業者のAI関与のあり方

<AIに関与する事業者の分類>

- 業務外利用者：事業活動以外でAIを利用する者又はAIを直接事業で利用せずにAIシステム・サービスの便益を享受する、場合によっては損失を被る者

- データ提供者：AIの開発者・提供者・利用者にデータを提供する者（センサ、システムからデータを入手する場合を含む）。

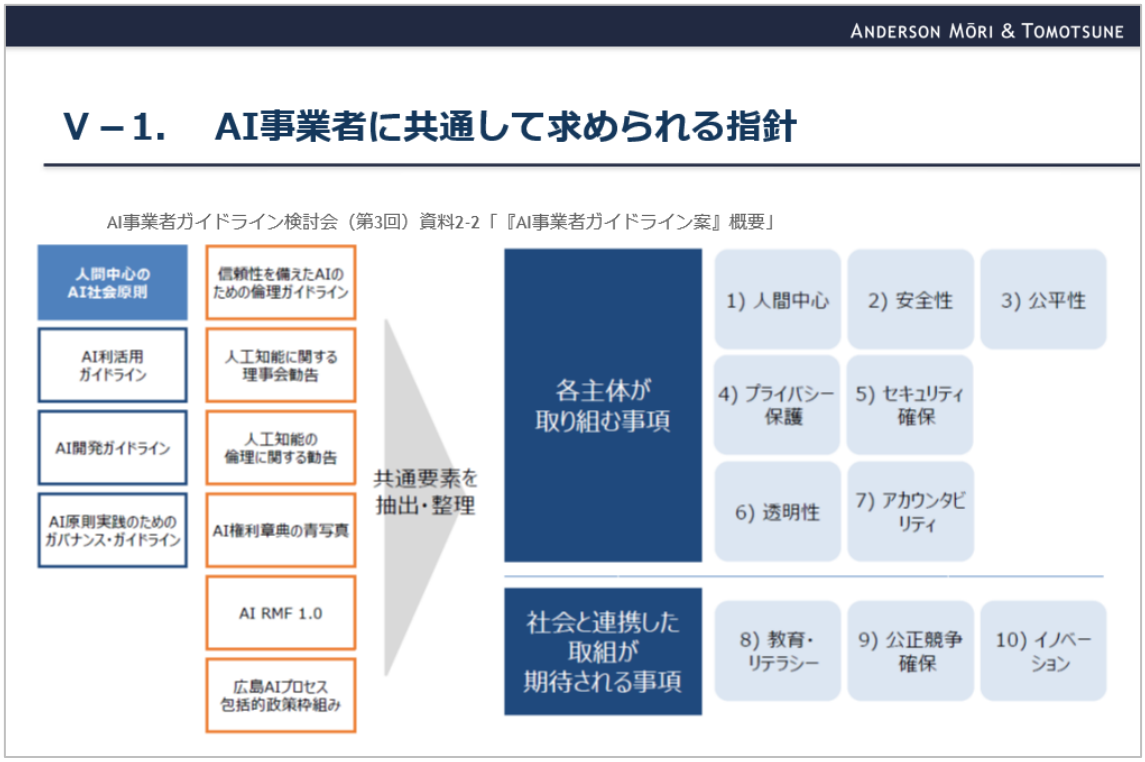
ガイドラインでは適用対象外とされているものの、ビジネスの実務を考えると、「業務外利用者」や「データ提供者」を挙げています。特に、医療の分野では、将来的に

は患者さんのデータを学習に使うことが想定されるため、「データ提供者」に該当する病院や診療所の位置づけが非常に重要です。

4. AI 事業者ガイドラインの全体像

ガイドラインの構成本編と付属資料で大きく二つに分かれています。

5. AI 事業者ガイドラインの概要



「AI 事業者に共通して求められる指針」は、「人間中心」から「アカウンタビリティ」までの部分は各主体が取り組むべき事項として記載されています。残りの「教育・リテラシー」「公正競争確保」「イノベーション」は社会と連携して取り組む事項については、政府側が中心となることが想定されています。

各主体が取り組むべき事項は 1 から 10 まで挙げられています。

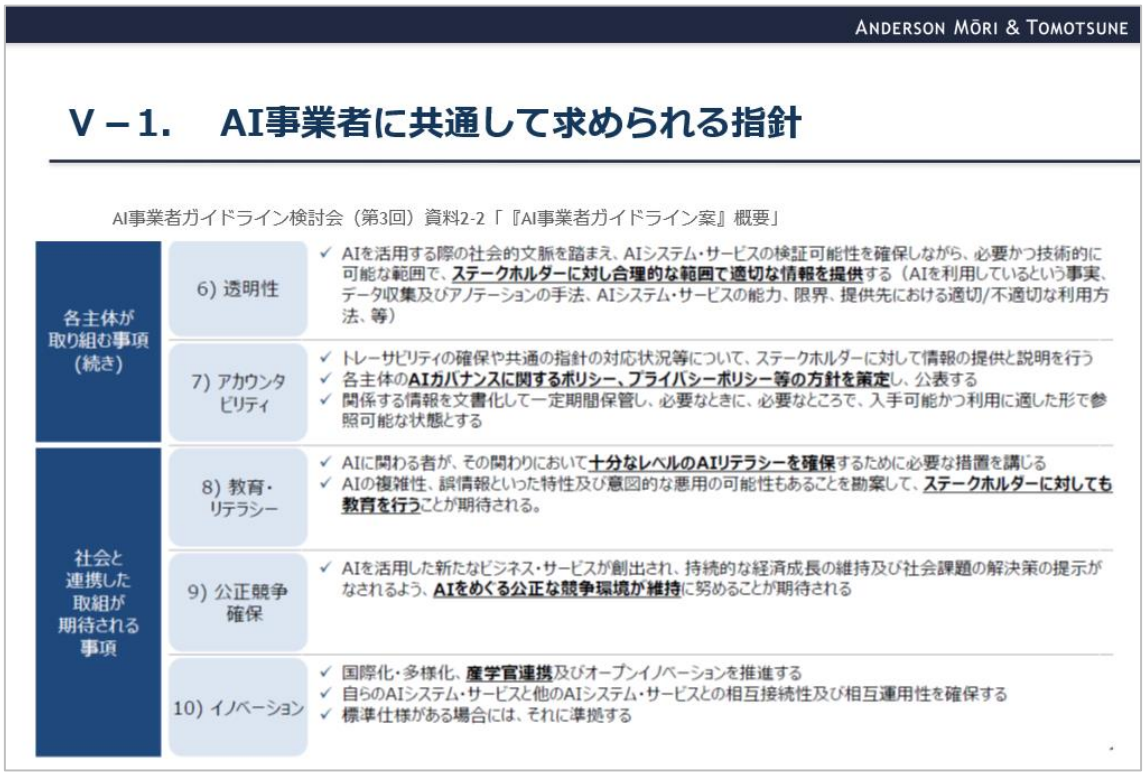
1 番の「人間中心」では、誤情報や偽情報の対策が必要とされています。また、「社会的弱者による AI の活用」は AI が情報弱者にも利用可能であるべきとされています。

2 番の「安全性」では、「学習等に用いるデータの正確性」の確保が求められています。また、データの透明性などを確保する枠組みを作ることが重要です。データの品質が低下すると、AI モデルの性能が低下し、ハルシネーションを引き起こす確率が上がるため、適切なタイミングと方法で更新することが重要です。

3 番の「公平性」では、AI の出力結果が公平性を欠かないように、人間の判断を介在させることが求められています。GDPR の人間に影響を与える決定を機械任せにしないという考えと同様です。

4 番の「プライバシー保護」は、事業者はプライバシーポリシー等を公開することが求められます。

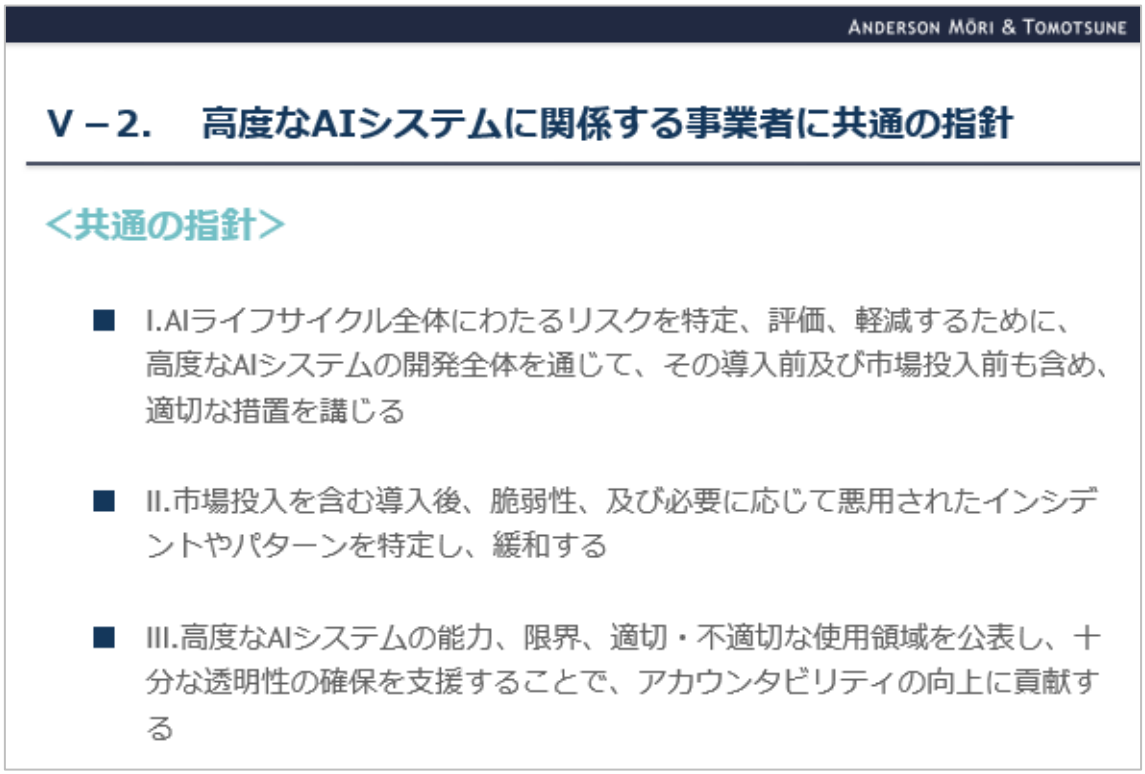
5 番の「セキュリティ確保」は、一般的なセキュリティの話です。



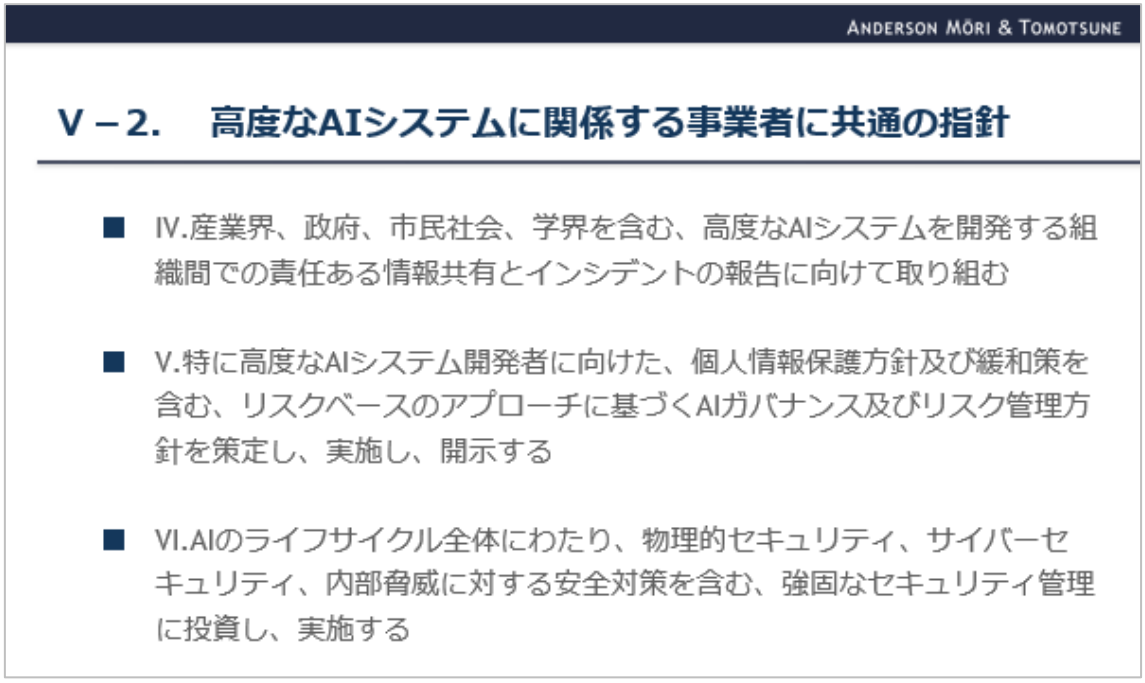
6 番の「透明性」では、営業秘密やプライバシーとのバランスを考慮しつつ、ステークホルダーに対し合理的な範囲で適切な情報を提供する」ことが求められています。その中には、AI の利用の事実、アノテーション方法、AI システムの能力限界などの情報提供が含まれます。

7 番の「アカウンタビリティ」では、関係する情報を文書化して一定期間保管することで、エンドユーザーや末端の事業者が正しく判断できることが求められています。

8 番の「リテラシーの確保」について、医療関係における AI 情報提供など、事業者側が対応する場面も想定されます。

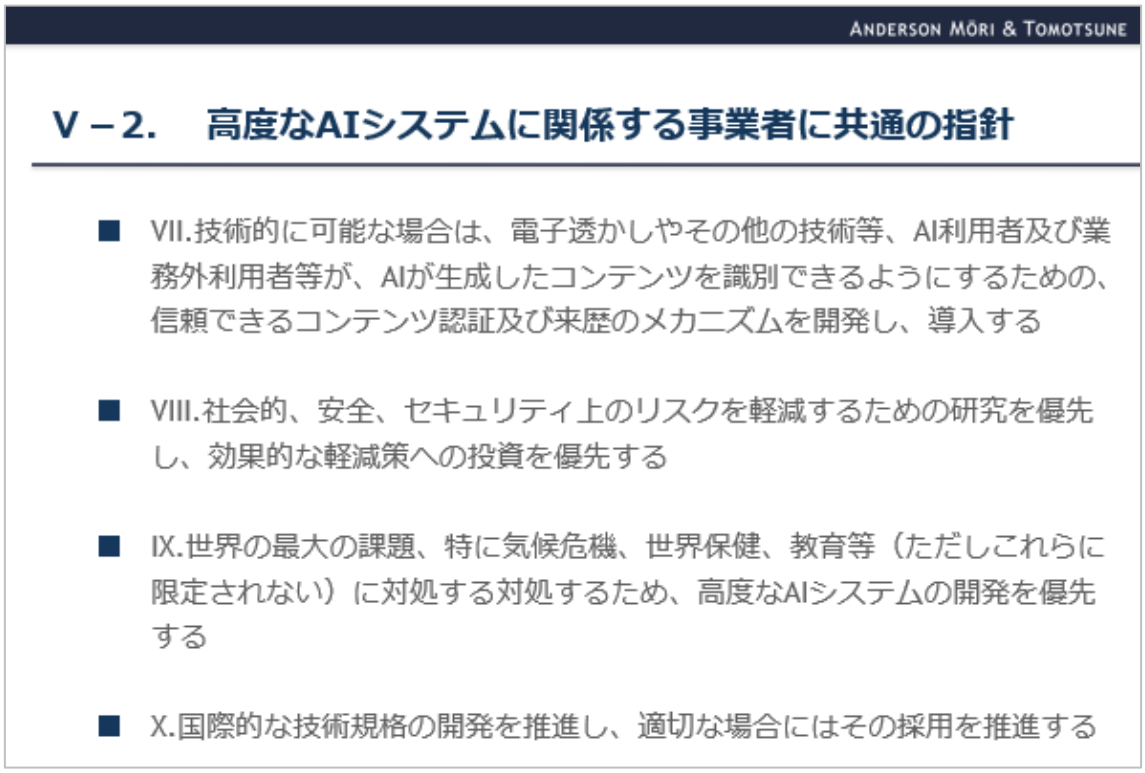


広島での G7 会合の合意に基づいて作られた「高度な AI システムに関係する事業者に共通の指針」は AI ライフサイクル全体にわたって適切な管理を求めています。特に、透明性の確保とアカウンタビリティの向上が強調されています。



トラブルが発生した場合には、インシデントの報告と情報共有システムの構築が求められています。不祥事や事故、トラブルの情報を業界全体や政府と共有し、対応します。

VI では強固なセキュリティ管理を実施することが記載されています。

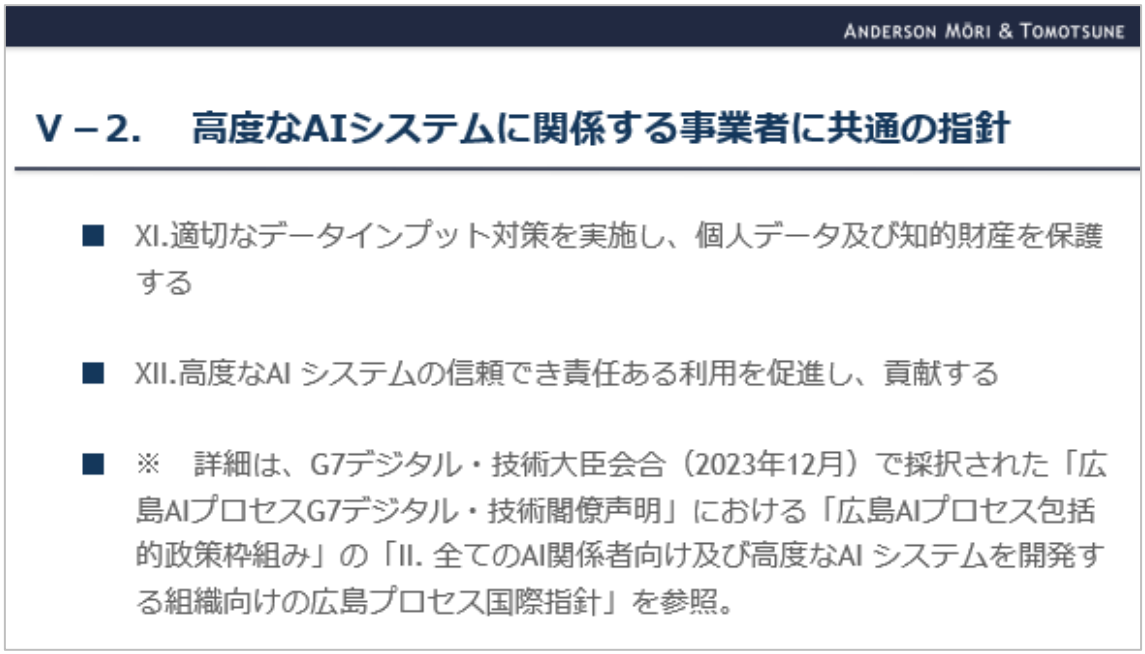


VII では電子透かし等を導入して AI 生成物であることが識別できるようにすることを求められています。これは、権利者やクリエイターの権利を守るために有効とされています。

VIII ではセキュリティ上のリスクを軽減するための研究を優先することになります。

IX は気候危機などの気候変動への対応を優先するというものです。

X は、国際的な、標準的な技術規格を作っていこうということで、実際、安全性に関しては既に進められているかと思います。



XI ではデータインプット対策で、個人情報や著作権を保護することが求められています。

6. AI 開発者

「開発者特有の留意事項」として、データ前処理の段階では、プライバシー・バイ・デザインを採用し、個人情報等の適切な法措置を実施することが重要です。また、AI システムのデータの質を管理し、不公正なバイアスがないか確認することも求められています。

リスクを最小限に抑えることも重要で、問題が発生した場合には、事後検証のための記録を保存することが必要です。

D-7「生じるリスク等の情報提供と説明」では、開発者から提供者に対して求められています。D-10 では、研究開発等において、品質や信頼性等を確保するための措置が求められています。

7. AI 提供者

提供者に関しても「特有の留意事項」を説明されています

P-2 では、AI 開発者が想定している利用環境と実際の利用環境と齟齬がないかを、提供者が確認することが求められています。

P-3 では、バイアスの検討に加え、AI モデルの判断根拠を定期的に評価し、バイアスをチェックすることも求められています。

P-6 では、システムアーキテクチャやデータ処理プロセス等、AI システムの意思決定に影響を与える部分については、AI 提供者がきちんと文書化することが求められています。

Ⅶ－１．　AI提供者特有の留意事項

AI事業者ガイドライン検討会（第3回）資料2-2「『AI事業者ガイドライン案』概要」

AIシステム・サービス提供後	P-2) ii.	適正利用に資する提供	- **適切な目的**でAIシステム・サービスが利用されているかを定期的に検証する
	P-4) ii.	プライバシー侵害への対策	- AIシステム・サービスにおけるプライバシー侵害に関して**適宜情報収集し、再発防止**を検討する
	P-5) ii.	脆弱性への対応	- 最新のリスクに対応するために提供の各工程で気を付けるべき点の動向を確認し、**脆弱性に対応することを検討する**
	P-6) ii.	関連するステークホルダーへの情報提供	- AIシステムの技術的特性、予見可能なリスク、緩和策、出力又はプログラムの変化の可能性、不具合の原因、対応状況、インシデント事例、学習データの収集ポリシー、その学習方法、実施体制等に関する情報を説明できるようにする - AIの性質及び利用目的等に照らして、**AIを利用しているという事実や適切/不適切な使用方法、更新内容とその理由等の情報提供や説明の実施**
	P-7) i.	AI利用者への共通の指針の対応状況の説明	- AI利用者に対して、**適正利用を促し**、正確性・必要に応じて最新性等が担保されたデータの利用やコンテキスト内学習による不適切なモデルの学習に対する注意喚起、**個人情報を入力する際の留意点についての情報を提供**する - AIシステム・サービスへの個人情報の不適切入力について注意喚起する
	P-7) ii.	サービス規約等の文書化	- AI利用者に向けた**サービス規約を作成するとともにプライバシーポリシーを明示**する

適正利用について。提供者は、利用者が AI システム・サービスを適正な目的で使用しているかを定期的に検証することが求められています。例えば、現在の生成 AI では、人が隠したいセンシティブな情報を工夫して引き出すこと、たとえば、芸能人が病気を隠している場合に、それを興味本位で知りたい人が AI から情報を引き出すことが可能です。このような不適切な使用を防ぐために、提供者は適切な対策を講じる必要があります。

また、P-6 では、AI を利用している事実や使用方法、更新内容等の説明の実施を求められています。

ANDERSON MŌRI & TOMOTSUNE

Ⅶ－2.　高度な AI システムを取り扱うAI提供者特有の留意事項

<共通の指針>

- 高度な AI システムを取り扱うAI提供者は、「第２部D. 高度なAIシステムに関係する事業者に共通の指針」について以下のように対応する。
- I）～XI）　適切な範囲で遵守すべきである
- XII）　遵守すべきである

「高度な AI システム」については、I から XI を適切な範囲で遵守すべきだとされています。

8. AI 利用者

ANDERSON MŌRI & TOMOTSUNE

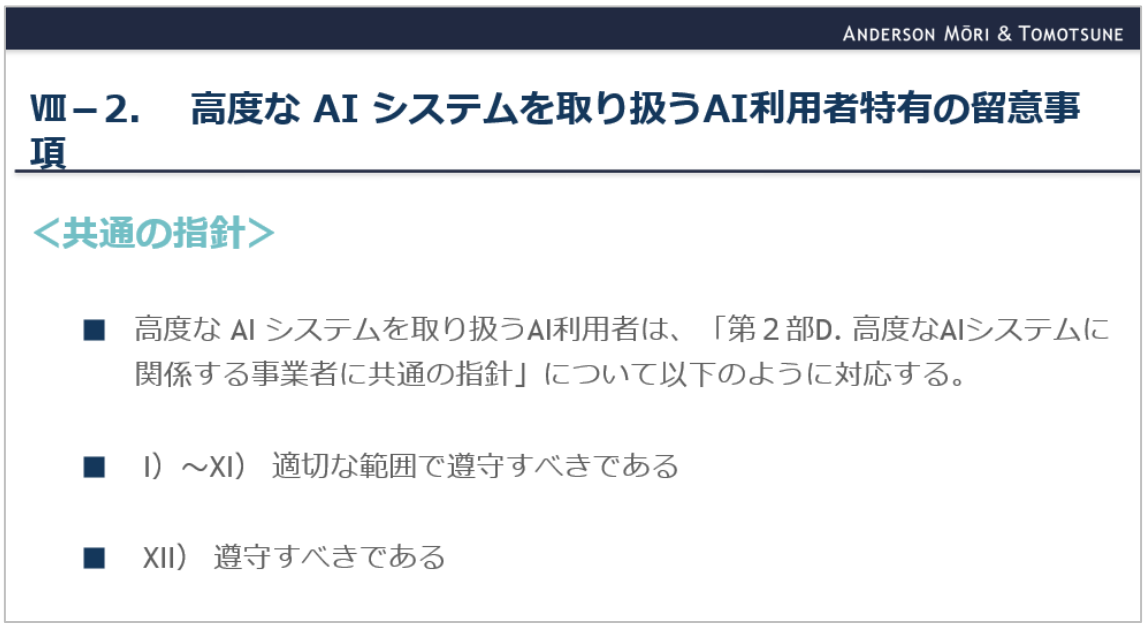
Ⅷ－１.　AI利用者特有の留意事項

AI事業者ガイドライン検討会（第3回）資料2-2「『AI事業者ガイドライン案』概要」

AIシステム サービス 利用時	U-2) i.	安全を考慮した適正利用	- AI提供者が定めた利用上の留意点を遵守して、**AI提供者が設計において想定した範囲内で利用**する - AIの出力について精度及びリスクの程度を理解し、**様々なリスク要因を確認した上で利用**する
	U-3) i.	入力データ、プロンプトに含まれるバイアスへの配慮	- 公平性が担保されたデータの入力を行い、プロンプトに含まれるバイアスに留意して、責任をもって**AI出力結果の事業利用判断を行う**
	U-4) i.	個人情報の不適切入力及びプライバシー侵害への対策	- AIシステム・サービスへ個人情報を不適切に入力しないよう注意を払う - AIシステム・サービスにおける**プライバシー侵害に関して適宜情報収集**し、防止を検討する
	U-5) i.	セキュリティ対策の実施	- AI提供者による**セキュリティ上の留意点を遵守**する - AIシステム・サービスに機密情報等を不適切に入力しないよう注意を払う
	U-6) i.	関連するステークホルダーへの情報提供	- 公平性が担保されたデータの入力を行い、プロンプトに含まれるバイアスに留意して、**出力結果を取得し、結果を事業判断に活用した際は、その結果が必要な関連するステークホルダーに周知**する
	U-7) i.	関連するステークホルダーへの説明	- AIの特性や用途、提供先との接点、プライバシーポリシー等を踏まえ、データ提供の手段、形式等について、あらかじめ**当該ステークホルダーに平易かつアクセスしやすい方法で情報提供**する - AIの出力結果を特定の個人又は集団に対する評価の参考とする場合には、合理的な範囲で人間による判断を行い、説明責任を果たす - **関連するステークホルダーからの問合せに対応する窓口を合理的な範囲で設置**し、AI提供者とも連携の上説明や要望の受付を行う
	U-7) ii.	提供された文書の活用と規約の遵守	- AI提供者から提供されたシステムについての**文書を保管・活用**する - AI提供者が定めた**サービス規約を遵守**する

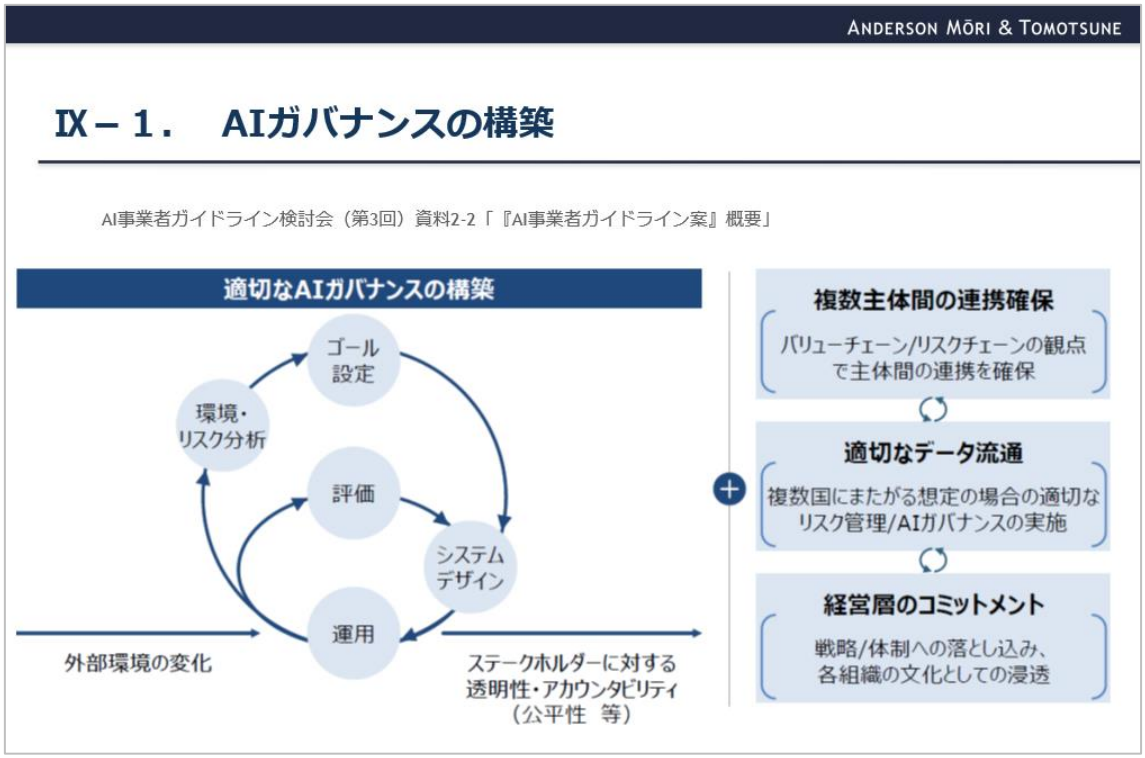
「AI 利用者特有の留意事項」については、自社のシステムやサービスに AI 組み込み、消費者に提供するケースや従業員が社内で利用するケースが該当します。

U-6 については、バイアスに注意した上で、「出力結果を取得し、結果を事業判断に活用した際は、その結果が必要な関連するステークホルダーに周知する」ことが求められています。同じステークホルダーへの説明については、アクセスしやすい方法で提供することが明確に求められています。



「高度な AI システム」に関しては、I から XI において「適切な範囲で」という表現が使われています。利用者側ができることが開発者より限られているため、このような書き方になっています。

9. AI ガバナンスの構築



AI ガバナンスの構築に関しては、継続的に改善していくプロセスが求められます。

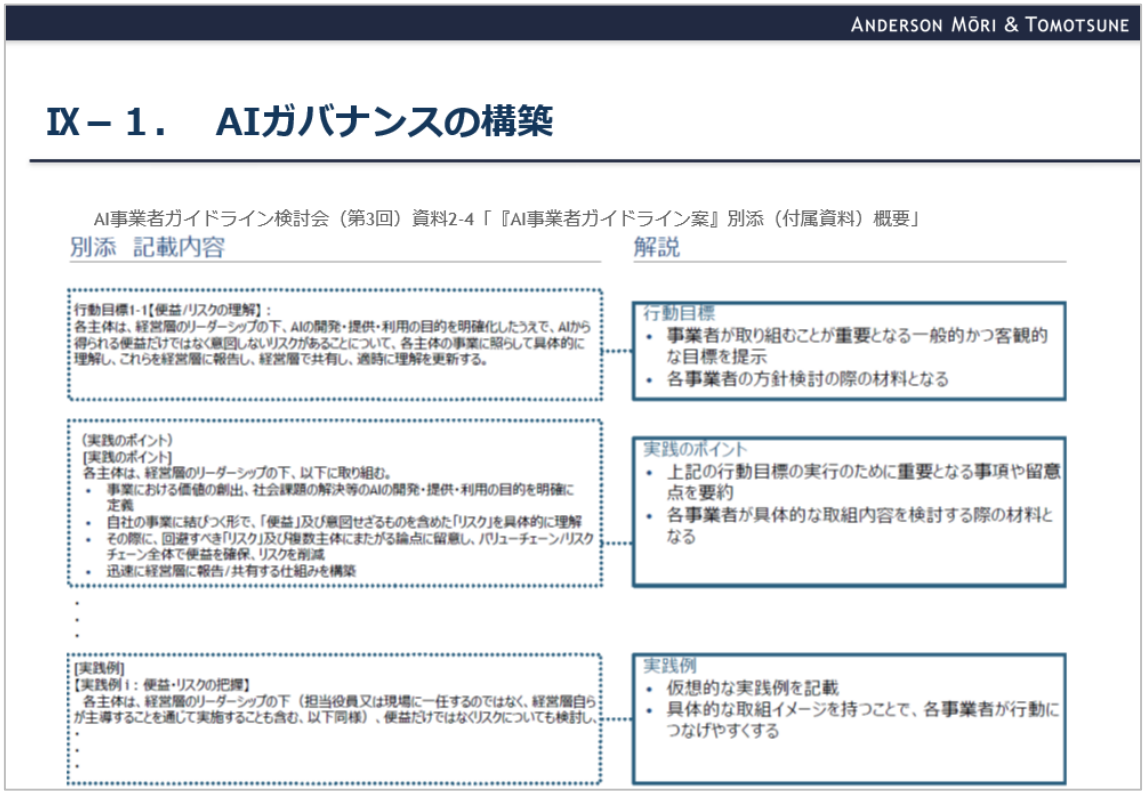
AI に関しては複数の事業者が関与することが多いため、バリューチェーンやリスクチェーンの観点から、連携を確保することが重要です。また、海外の事業者と連携する場合も多く、個人情報が含まれる場合には越境移転規制がかかる可能性があります。国によってデータの持ち出しやプログラムの輸出を制限する国もあり、規制を遵守しながらビジネスを進めるために、経営層が積極的にコミットすることが求められています。

右側が行動目標で、左側が分類です。

「環境・リスク分析」に関しては、「便益／リスクの理解」を十分に行った上で、自社の AI 理解度を把握し、それを踏まえてリスク分析を行うことが求められています。

「ゴール設定」に関しては、「AI ガバナンス・ゴール設定」が必要です。

「システムデザイン」では、現実とゴールの乖離、その評価をどのように行うか、また AI マネジメントをどのようにするのかが課題です。更に、「運用」「評価」「環境・リスクの再分析」を行うことが重要です。



更に、当事者や事業者が取り組むべき行動目標、実践のポイントと実践例も影響を受けると記載されています。

ANDERSON MŌRI & TOMOTSUNE

Ⅸ－2.　AIガバナンスとバリューチェーン

<バリューチェーン特有の留意事項>

- 複数主体にまたがる論点の例：AIリスク把握、品質の向上、各AIシステム・サービスが相互に繋がること（System of Systems）による新たな価値の創出、AI利用者又は業務外利用者のリテラシー向上等
- 主体間で整理が必要になりうる点の例：学習及び利用に用いるデータ・生成されたAIモデルに関する権利関係の契約等
- データの流通をはじめとしたリスクチェーンの明確化、並びに開発・提供・利用の各段階に適したリスク管理及び、AIガバナンス体制の構築を実施する

「バリューチェーン特有の留意事項」についての問題は、データの質や AI のリスクが複数の当事者に関与しているため、責任の所在が不明確になることが多い点です。仮に責任者が特定できたとしても、データや出力結果の権利帰属の問題が別途発生し、紛争の原因となる可能性があります。そのため、データの流通過程に着目し、ガバナンスの中でリスク管理を行うことが求められています。

ANDERSON MŌRI & TOMOTSUNE

Ⅸ－2.　AIガバナンスとバリューチェーン

- AI開発からサービス実施にわたるバリューチェーン/リスクチェーンが複数国にまたがることが想定される場合、データの自由な流通（Data Free Flow with Trust、以下、「DFFT」という）の確保のための適切なAIガバナンスに係る国際社会の検討状況の把握と、それを踏まえた相互運用性の確保（「標準」と及び「枠組み間の相互運用性」の二側面）

データの自由な流通や国境を超えた越境移転については、国によってはそれを制限する規制を設けている場合があります。このような規制の有無を含めて、ガバナンスの中で検討する必要があります。

こちらの表は、AI による便益の分析例を示しています。

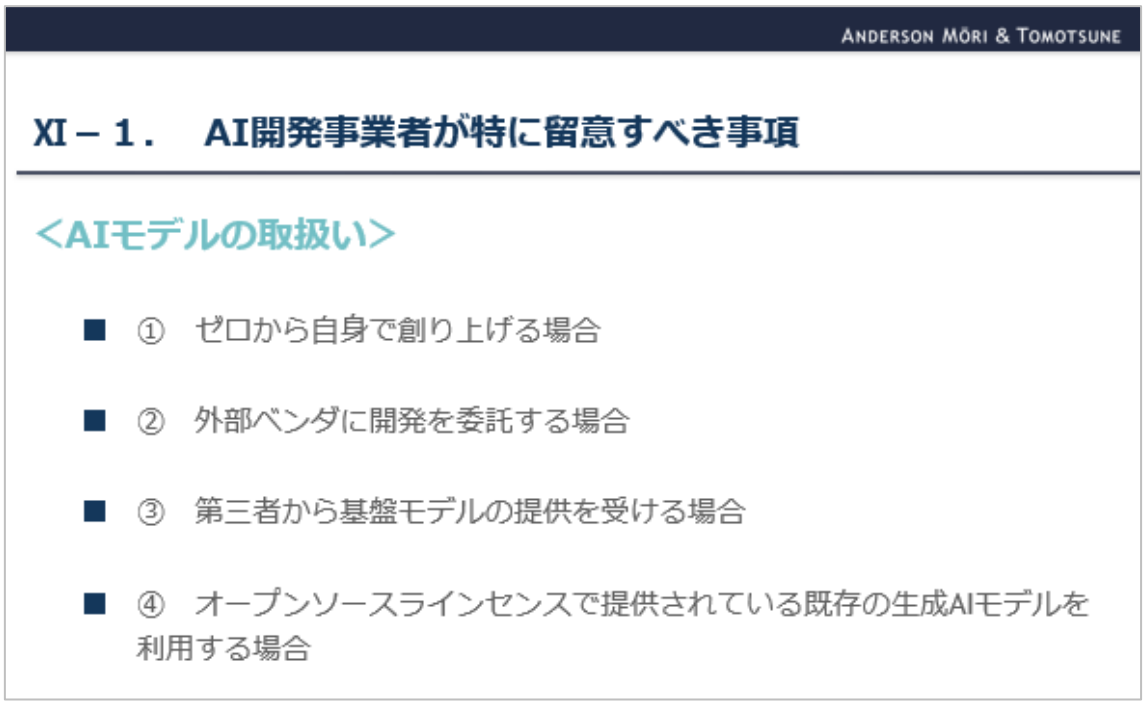
X－2.　AIによるリスクの分析例

AI事業者ガイドライン検討会（第3回）資料2-4「『AI事業者ガイドライン案』別添（付属資料）概要」

	リスク	事例	対応する「共通の指針」
従来型AIから存在するリスク	バイアスのある結果及び差別的な結果の出力	• IT企業が自社で開発したAI人材採用システムが女性を差別するという機械学習面の欠陥を持ち合わせていた	1) 人間中心 3) 公平性
	フィルターバブル及びエコーチェンバー現象	• SNS等によるレコメンドを通じた社会の分断が生じている	1) 人間中心
	多様性の喪失	• 社会全体が同じモデルを、同じ温度感で使った場合、導かれる意見及び回答がLLMによって収束してしまい、多様性が失われる可能性がある	1) 人間中心
	不適切な個人情報の取扱い	• 透明性を欠く個人情報の利用及び個人情報の政治利用も問題視されている	1) 人間中心 4) プライバシー保護
	生命、身体、財産の侵害	• AIが不適切な判断を下すことで、自動運転車が事故を引き起こし、生命や財産に深刻な損害を与える可能性がある • トリアージにおいては、AIが順位を決定する際に倫理的なバイアスを持つことで、公平性の喪失等が生じる可能性がある	2) 安全性 3) 公平性
	データ汚染攻撃	• AIの学習実施時及びサービス運用時には学習データへの不正データ混入、サービス運用時ではアプリケーション自体を狙ったサイバー攻撃等のリスクが存在する	5) セキュリティ確保
	ブラックボックス化、判断に関する説明の要求	• AIの判断のブラックボックス化に起因する問題も生じている • AIの判断に関する透明性を求める動きも上がっている	6) 透明性 7) アカウンタビリティ
	エネルギー使用量及び環境の負荷	• AIの利用拡大により、計算リソースの需要も拡大しており、結果として、データセンターが増大しエネルギー使用量の増加が懸念されている	1) 人間中心
生成AIで特に顕在化したリスク	悪用	• AIの詐欺目的での利用も問題視されている	2) 安全性 8) 教育・リテラシー
	機密情報の流出	• AIの利用においては、個人情報や機密情報がプロンプトとして入力され、そのAIからの出力等を通じて流出してしまうリスクがある	5) セキュリティ確保 8) 教育・リテラシー
	ハルシネーション	• 生成AIが事実と異なることをもっともらしく回答する「ハルシネーション」に関してはAI開発者・提供者への訴訟も起きている	2) 安全性 8) 教育・リテラシー
	偽情報、誤情報を鵜呑みにすること	• 生成AIが生み出す誤情報を鵜呑みにすることがリスクとなりうる • ディープフェイクは、各国で悪用例が相次いでいる	1) 人間中心 8) 教育・リテラシー
	著作権との関係	• 知的財産権の取扱いへの議論が提起されている	2) 安全性
	資格等との関係	• 生成AIの活用を通じた業法免許や資格等の侵害リスクも考えうる	2) 安全性
	バイアスの再生成	• 生成AIは既存の情報に基づいて回答を作るため既存の情報に含まれる偏見を増幅し、不公平や差別的な出力が継続/拡大する可能性がある	3) 公平性

こちらの表は、AI によるリスクの分析例を示しています。右端に「対応する『共通の指針』」と書かれており、これは AI 側が採用した 1 から 10 の指針に基づいています。

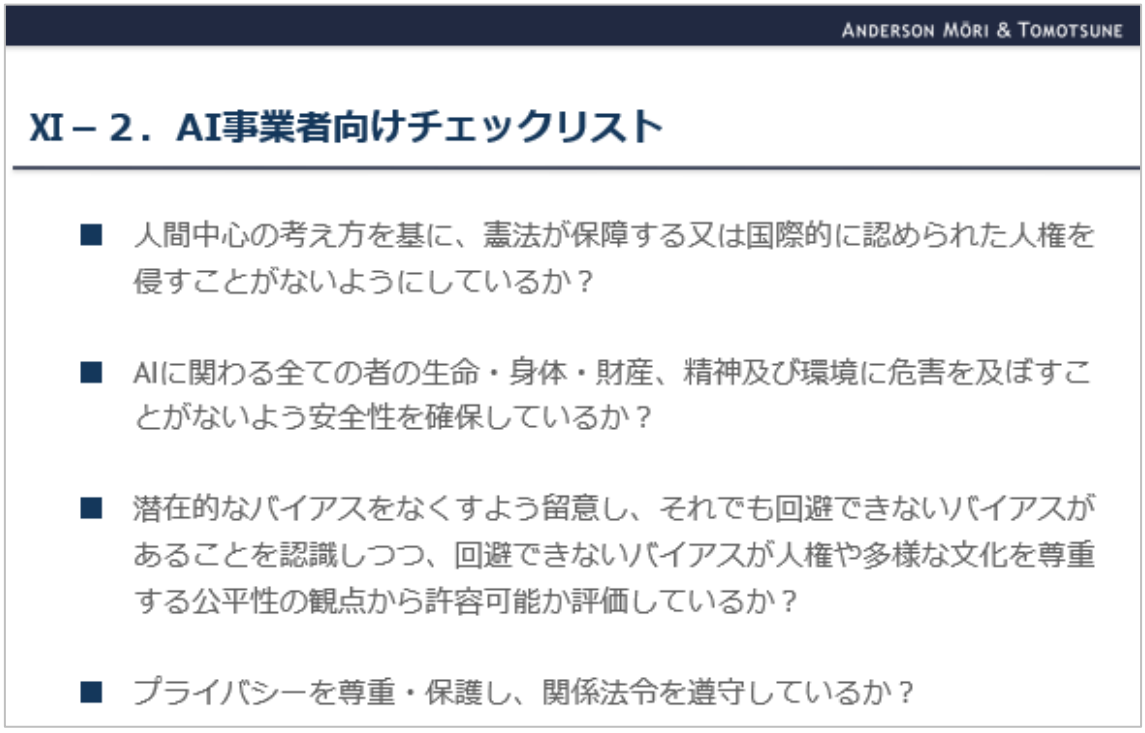
医療関係の分野では、メリットも相当大きい一方で、デメリットもある程度はどうしても避けられないかと思います。そのため、リスクと便益の分析は医療関係でも重要になります。



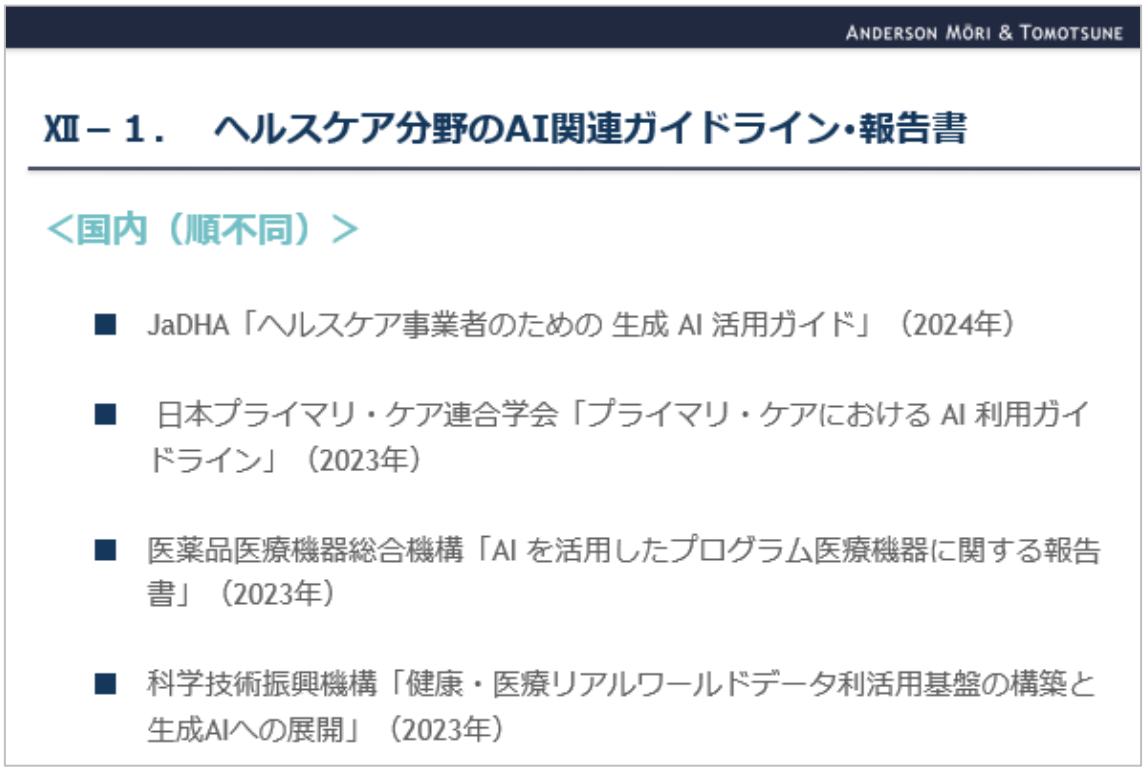
ここでは、「AI 開発事業者が特に留意すべき事項」の概要を記載しています。「AI モデルの取扱い」について、権利帰属は契約や契約書によって異なるため、適切な対応が難しいです。

また、「その他のポイント」としては、「生成 AI モデル開発」「学習用データセットの取り扱い」「リスクの棚卸し作業」があります。特に学習用データセットに関しては、

大きな会社からベンダーに AI の開発を依頼し、そのベンダーが会社のデータやカスタマーデータなどを使用して学習した成果を、依頼者である会社に帰属することを認めるべきかどうかがよく問題になっています。



そういったことをチェックリスト化したのが、これも、「AI 事業者ガイドライン」の別添資料の中で公表されているものです。これ自体は、非常に幅が広いというか緩やかな質問なので、だいたい、イエスなどになるかなというところです。



ここでは、まず、ヘルスケア分野の国内のガイドラインを挙げています。

ちょうど先々週だったでしょうか、東京財団で LLM や生成 AI とヘルスケアの関連ということで、シンポジウム的なことをされていました。そちらのホームページを見ると、三つの提言のようなことが出ています。そのようなものも含めると、結構な数のヘルスケアに特化した AI 関連のガイドラインのようなものが、いま既に世の中にあるかと思います。

国外に関しても、かなりの数の論文、記事等があります。その中でも重要性が高いのは、やはりこの WHO の 2024 年のドキュメントと、2021 年のドキュメント、加えて、OECD が 2024 年に出している文書です。あとは、先ほど触れた「高度な AI システムを開発する組織向けの国際行動規範」ですね。これが関係してくるところかと思います。

XII－２．　WHO　LMM報告書（2024）

WHO「AI Ethics and Governance Guidance for Large Multi-Modal Models operating in the Health Sector - Data Protection Considerations」

Table 1. Potential benefits and risks in various uses of LMMs in health care

Use	Potential or proposed benefits	Potential risks
Diagnosis and clinical care	Assist in managing complex cases and review of routine diagnoses Reduce the communication workload of health-care providers ("keyboard liberation") Provide novel insights and reports from various unstructured forms of health data	Inaccurate, incomplete or false responses Poor quality training data Bias (of training data and responses) Automation bias Degradation of skills (of health-care professionals) Informed consent (of patients)
Patient-guided use	Generate information to improve understanding of a medical condition (as a patient or as a caregiver) Virtual health assistant Clinical trial enrolment	Inaccurate, incomplete or false statements Manipulation Privacy Less interaction between clinicians and patients Epistemic injustice Risk of delivery of care outside the health system
Clerical and administrative tasks	Assist with paperwork and documentation required for clinical care Assist in language translation Completion of electronic health records Draft clinical notes after a patient visit	Inaccuracies and errors Inconsistent responses depending on prompts

今日はこの中で、今回のプロジェクトに一番関連性が高いと思われる、LLM でなく LMM に関する報告書をご紹介します。これは、2024 年に WHO から出されたものになります。内容自体は、ご関心があればあとで直接お読みいたくほうがいいかと思うのですが、簡単に概要だけご紹介します。

まず、Table 1.と書いてあるところです。LMM を医療分野で活用していくときに、どのような便益、メリットと、デメリットがあるのかが整理されています。

一番上の「Diagnosis and clinical care」では、LMM を利用すると、例えば不正確であったり、不完全であったり、虚偽の反応をしてしまうといったことや、学習用データの質が低いままバイアスを余計にかけてしまうようなことになるといったこと、あるいはインフォームド・コンセントが問題になり得る、というデメリットが書いてあります。

デメリットばかりではありません。日常的な診断を AI がサポートできるようになったら、それはそれで非常にありがたいことではあるので、メリットは大きいわけですね。そのようなメリットとデメリットをちゃんと比較検討していこうということが、この報告書の Table 1.では言われています。

今回のプロジェクトに関係がありそうなのは、「科学研究と医薬品開発」、一番最後の箇所になります。これに関しては、アクセスの格差を拡大する、あるいはアルゴリズムに対してコンテンツの責任を問うことはできないので、責任の問いようがなくなるという問題があると指摘されています。

XII－2． WHO　LMM報告書（2024）

WHO「AI Ethics and Governance Guidance for Large Multi-Modal Models operating in the Health Sector - Data Protection Considerations」

Table 2. Risks to health systems associated with use of LMMs in health care

Type of risk	Description
Overestimation of the benefits of LMMs	There may be a tendency to 'technological solutionism', or over-estimation of the benefits of LMMs while ignoring or downplaying challenges in its use, including its safety, efficacy and utility.
Accessibility and affordability	Equitable access to LMMs may be lacking for several reasons, including the "digital divide" and subscription fees to access LMMs.
System-wide biases	Use of ever-larger data sets could increase biases encoded in LMMs, which could be automated throughout a health-care system.
Impacts on labour	Use of LMMs could lead to job losses in some countries and require health workers to retrain and adjust to use of LMMs. Data annotation and filtering can lead to low wages and to untreated psychological distress.
Dependence of health systems on ill-suited LMMs	Dependence on LMMs could make health systems vulnerable if LMMs are not maintained or (in low- and middle-income countries) are updated only for use in high-income countries. Furthermore, lack of preservation and protection of privacy and confidentiality could undermine trust in health-care systems by people who are not confident that their privacy will be protected.
Cybersecurity risks	Malicious attacks or hacking could undermine safety and trust in the use of LMMs in health care.

続いて、Table 2.では、医療における LMM 使用に伴う医療システムのリスクが挙げられています。ここでは、LMM を過大評価してしまうのではないかという問題や、システム全体にバイアスがかかるのではないかと書かれています。より大規模なデータセットを使うと、LMM にエンコードされるバイアスを増加させる必要が出てきてしまうので、結局医療システム全体を通して、自動的にバイアスが出てしまう可能性があるということが指摘されています。

それ以外の、労働への影響やサイバーセキュリティへのリスクなどは、必ずしも AI 特有のことではないと思うので、本日は飛ばします。

XII－2．　WHO　LMM報告書（2024）

WHO「AI Ethics and Governance Guidance for Large Multi-Modal Models operating in the Health Sector - Data Protection Considerations」

Figure 2: Value chain of the development, provision and deployment of LMMs

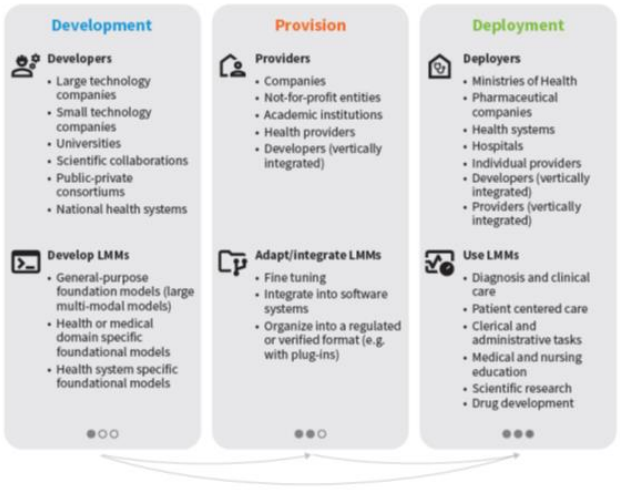
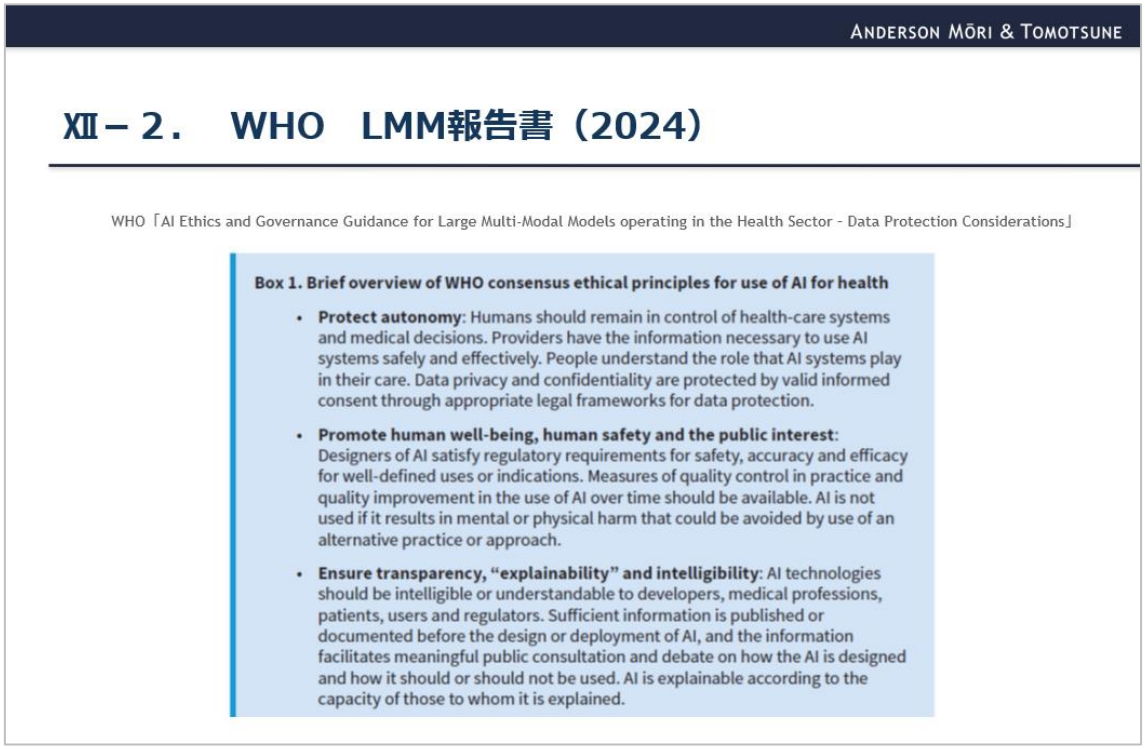


Figure 2.は、AI のバリューチェーンです。「開発者」「提供者」「利用者」は WHO のドキュメントでも、同じように「開発（Development）」「提供（Provision）」「展開（Deployment）」という 3 種類の分類をしています。

これは、LMM との関係での立場です。LMM の開発者は左側の青いところ、提供・統合するのは真ん中の橙色で、一番最後に使う人が出てくるということです。このようなバリューチェーンを踏まえて、WHO の報告書を読み込んでほしいという案内が、ここでされています。



この報告書では、「健康のために AI を利用する WHO コンセンサス倫理原則」の概観が説明されています。全部は読み上げられませんが、一つ目は「自律性の保護」、二つ目は「人間の幸福、幸せぶり」という話。あとは「透明性」と「説明可能性」の話が書かれています。

ANDERSON MŌRI & TOMOTSUNE

XII－2．　WHO　LMM報告書（2024）

WHO「AI Ethics and Governance Guidance for Large Multi-Modal Models operating in the Health Sector - Data Protection Considerations」

- **Foster responsibility and accountability** to ensure that AI is used under appropriate conditions and by appropriately trained people. Patients and clinicians evaluate development and deployment of AI. Regulatory principles are applied upstream and downstream of the algorithm by establishing points of human supervision. Appropriate mechanisms are available for questioning and for redress for individuals and groups that are adversely affected by decisions based on AI.
- **Ensure inclusiveness and equity**: AI is designed and shared to encourage the widest possible, appropriate, equitable use and access, irrespective of age, sex, gender identity, income, race, ethnicity, sexual orientation, ability or other characteristics. AI is available for use not only in high-income settings but also in low- and middle-income countries. AI does not encode biases to the disadvantage of identifiable groups. AI minimizes inevitable disparities in power. AI is monitored and evaluated to identify disproportionate effects on specific groups of people.
- **Promote AI that is responsive and sustainable**: AI technologies are consistent with the wider promotion of the sustainability of health systems, the environment and workplaces.

それから、「責任と説明責任の育成」、また「対応可能で持続可能な AI の推進」という話がここでは書かれています。

ANDERSON MŌRI & TOMOTSUNE

XII－2．　WHO　LMM報告書（2024）

＜診断と臨床治療におけるLMM使用のリスク＞

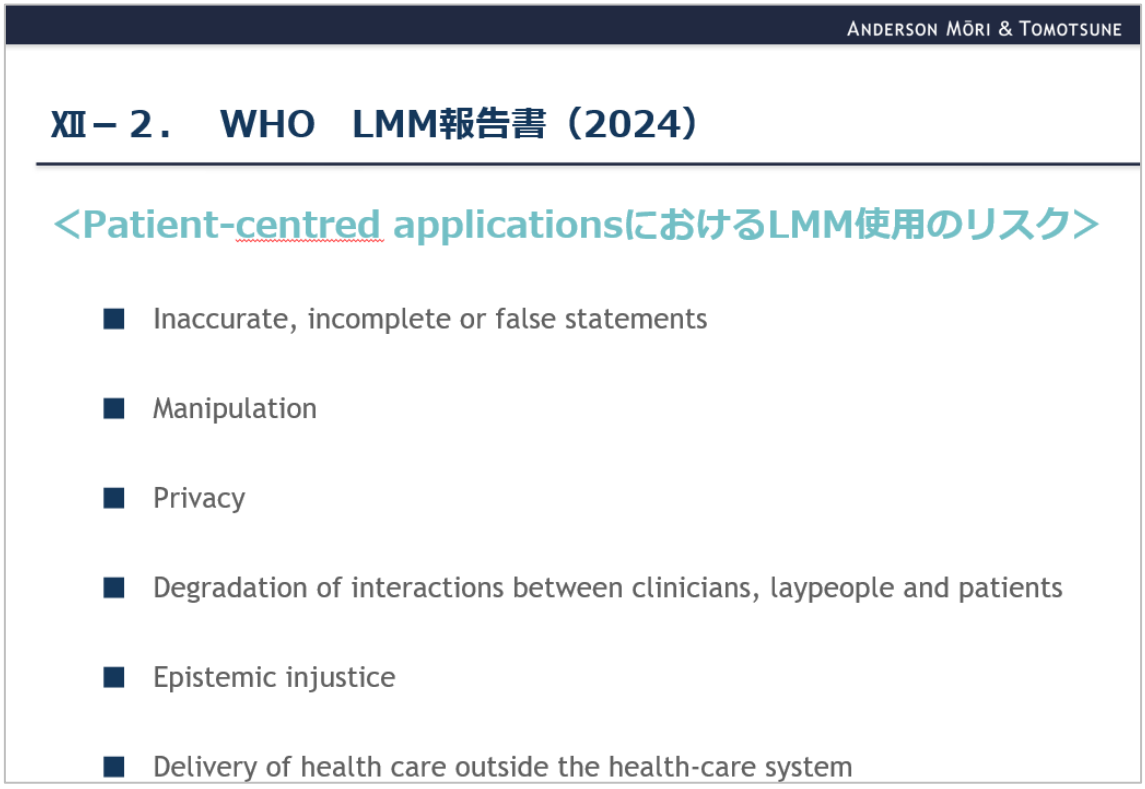
- Inaccurate, incomplete, biased or false responses
- Data quality and data bias
- Automation bias
- Skills degradation
- Informed consent

少し場面を進め、「診断と臨床治療における LMM 使用のリスク」についてご紹介します。

これは、ここで挙げたとおりなのですが、先ほどのハルシネーションの問題、不正や不誠実などの原因には、そういうものが働いている可能性もあるということです。データの質と偏りの問題も、当然、指摘されています。医療情報というのは、バイアスがかかりやすい種類の情報なので、そこにきちんと気を付けた上で学習させることが必要だということが指摘されています。

それ以外には、スキルが低下してしまうリスクなどが指摘されています。

また、インフォームド・コンセントについて簡単に申し上げると、LMM の利用が増えると、対面での利用はもちろんバーチャルな利用も増えるだろうということで、AI 技術が応答を補助したり、あるいは臨床医のフィードバックを真似した生成をする可能性があることを、患者に認識してもらう必要があるとしています。ここでは、そのような意味で、インフォームドという言い方をしているようです。



同じ話で、また少し場面が違って、これは患者を中心とするアプリケーションの場面において LMM を使用するときのリスクの一覧です。

二つ目に「Manipulation」と書いてあるところでは、ユーザーが AI を使っていくときに、変なことを教えたりしないようにということが、注意書きとして挙げられています。あとは、「Privacy」や「Degradation」などは、先ほども出てきた話かと思いますので飛ばします。

そして、分かりづらいのが「Epistemic injustice（認識論的不公正）」という言い方をしているところですが、これに関しては、いままで人間がしていた判断を AI に置き換えることにより、患者に認識論的な不公正がもたらされるということが指摘されています。

最後の「Delivery of health care」のところでは、医療システムの外で医療提供をすることについて、注意をするようにということが指摘されています。

Ⅻ－２．　WHO　LMM報告書（2024）

＜Scientific and medical research and drug におけるLMM使用のリスク＞

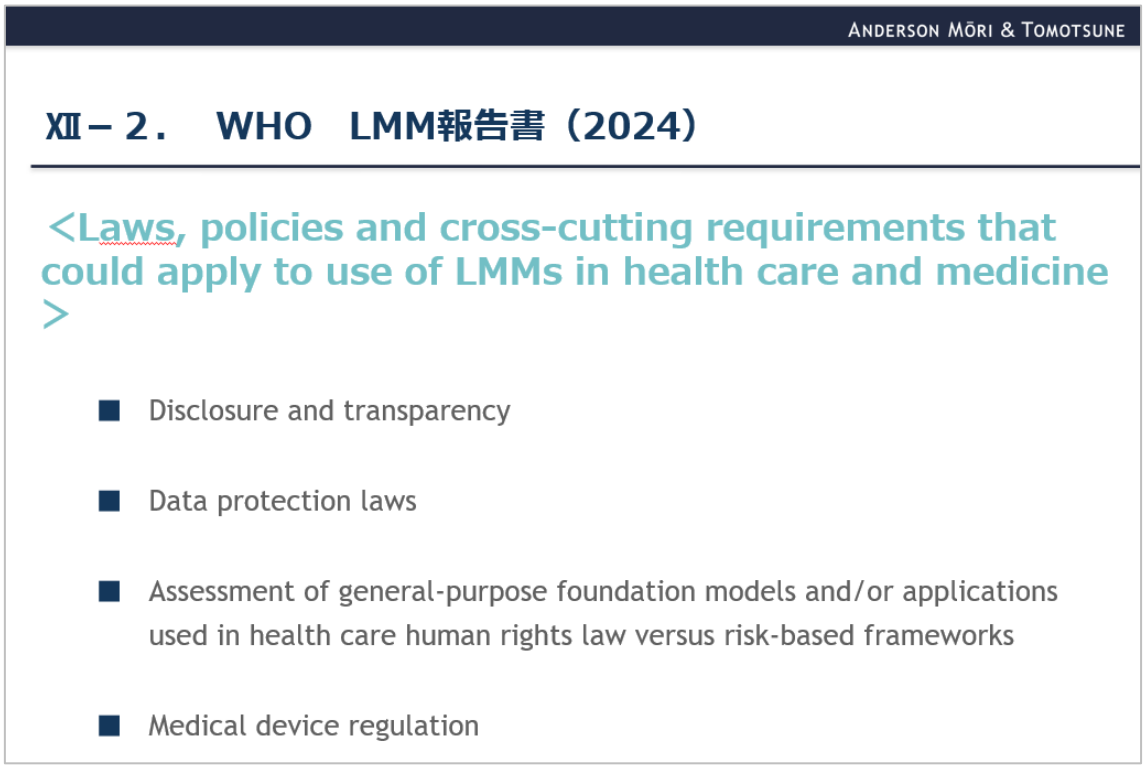
- Lack of accountability
- High-income country bias
- Hallucination and/or misinformation
- Undermining of trust
- Accessibility of LMMs and of knowledge generated by LMMs

「Scientific and medical research and drug」に関するリスクとしては、「説明責任の欠如」、「ハルシネーション」、「信頼の低下」や「LMM によって出力された情報へのアクセスビリティが不十分であること」等が挙げられています。

「Overestimating the benefits of LMMs and discounting risks」については、LMM を過大評価し、リスクを軽視してしまうことが指摘されています。LMM が新しい存在であるため、十分に理解されていないことが原因です。また、LMM が手軽に利用できるかどうか、システム全体のバイアス問題も重要です。データセットの偏りにより、女性、少数民族、田舎の人、高齢者ほか、社会的に弱い立場の人のデータが除外されやすい傾向があります。これが、医療において適切なのか疑問視されています。

労働と雇用への影響ということも書いてありますが、LMM が普及すると、最終的にホワイトカラーとかブルーカラーの一部に影響があると言われているということです。

もう一つ注意すべき点は、不適切な LMM に医療が依存するのは良くないという点です。医療従事者が不足している現状では、LMM の活躍が期待されるものの、導入後にその効果を冷静に見直すことも必要だと指摘されています。



「Laws, policies and cross-cutting requirements that could apply to use of LMMs in health care and medicine」では、法的及び政策運営の観点からリスクがあることが指摘されています。データの開示や透明性が問題となることも注意が必要です。また、「General-purpose foundation models」をヘルスケアで使用する際に、リスクベースのフレームワークで十分かどうかが議論されています。EU の AI 規則はリスクベースが、それを医療分野でも適応して良いかについても指摘されています。

ANDERSON MŌRI & TOMOTSUNE

XII－2．　WHO　LMM報告書（2024）

<Risks to be addressed when deploying a health-care service or application with a general-purpose foundation model (LMM) >

- inaccurate or false responses
- bias
- privacy of data entered into and put out by an LMM
- accessibility and affordability of an LMM

「general-purpose foundation model 汎用 AI のモデル」をヘルスケアの分野で使用する際に、不適切や不正確な応答、バイアス、プライバシー侵害が懸念されます。また、十分なアクセスビリティが適切なコストで確保されているかも重要です。

ANDERSON MŌRI & TOMOTSUNE

XII－2．　WHO　LMM報告書（2024）

<Risks to be addressed when deploying a health-care service or application with a general-purpose foundation model (LMM) >

- impacts on labour and employment
- automation bias and skills degradation
- the quality of interactions between health-care providers and patients