

# 我が国及び諸外国における生成AIに係る動向

2026/03/10 先進的AI利活用アドバイザーボード事務局

# 目次

## 1.生成 A I 等に関する政府・自治体等の動向

- AI事業者ガイドラインの令和7年度更新内容（案） p.4
- AIのセキュリティ確保のための技術的対策に係るガイドライン(案) p.6
- Chief AI Officer (CAIO) ガイドブック(案)及びCAIO設置・AIガバナンス実務マニュアル(案) p.7
- AIインシデントレスポンス・アプローチブック（エグゼクティブ・サマリ） p.8
- 生成AIの適切な利活用等に向けた知的財産の保護及び透明性に関する  
プリンシプル・コード（仮称）（案）概要 p.9

## 2.諸外国の A I に関する政策の動向

- 諸外国における政府のAI調達利用ルールとの比較 p.11
- 諸外国における政府のAI戦略における参考情報 p.12

# 1. 生成 A I 等に関する政府・自治体等の動向

# AI事業者ガイドラインの令和7年度更新内容（案）

## 1. AI事業者ガイドラインの令和7年度の更新の論点と更新方針

構成員・委員・事業者等からのご意見を踏まえ、令和7年度の更新の論点と更新方針を以下に整理

### 令和7年度更新の論点及び更新方針（案） 一覧

総務省検討会：AIネットワーク社会推進会議・AIガバナンス検討会  
経済産業省検討会：AI事業者ガイドライン検討会

| # | 更新の論点            | 主なご意見        | 更新方針   |
|---|------------------|--------------|--|
| 1 | AI技術の動向の反映       | 総務省検討会       | <b>AIエージェント、フィジカルAIに関する事項の追記</b><br>✓ AI事業者ガイドラインとしての定義の追加<br>✓ 便益の追加<br>✓ リスクに関する事項の追加<br>✓ 留意すべき事項の追加<br>✓ AIシステム・サービス例の追加 |
| 2 | AIによるリスクの記載の見直し  | 総務省検討会       | <b>AIによるリスクの記載の見直し</b><br>✓ リスクベースアプローチに関する内容の追記<br>✓ 一部リスクの分類見直し  |
| 3 | 主体区分の整理          | 経済産業省検討会     | <b>各主体区分の役割に関する補足の追加や図表の更新</b><br>✓ AI開発者の定義の補足<br>✓ 「一般的なAI活用の流れにおける主体の対応」の見直し<br>✓ 主体毎の役割の見直し                              |
| 4 | 特定単語の整理・見直し      | 経済産業省検討会     | <b>学習、データ種類等、多義的に捉えられる事項の記載</b><br>✓ 「学習」「推論」の定義・表現の見直し<br>✓ 「データ」の定義・表現の見直し   |
| 5 | ユーザビリティの改善       | 両省検討会        | <b>AI事業者ガイドラインの活用を支援する資料・ツールの検討</b><br>✓ 活用ガイドの検討<br>✓ チャットボットの検討  |
| 6 | AIガバナンスに関する動向の反映 | 両省検討会<br>事業者 | <b>AIガバナンスに関する国内外の最新動向や、事業者の取組事例の追記</b><br>✓ AI法や広島AIプロセスの動向等、国内外動向において注視すべき最新状況を追記<br>✓ 「AIガバナンスの構築に関する実際の取組事例」への事例追加等      |
| 7 | その他              | 両省検討会        | <b>脚注記載内容やリンクの更新</b>   |

# AI事業者ガイドラインの令和7年度更新内容（案）より抜粋

主な更新内容の一部として、AIエージェントの定義（案）の追加、リスクベースアプローチに資する内容の追記が挙げられている。

## 【更新内容】

### AI事業者ガイドラインとしての定義の追加

✓ AIエージェントの定義（案）

本ガイドラインにおけるAIエージェントとは、特定の目標を達成するために、環境を感知し自律的に行動するAIシステムとする。

✓ フィジカルAIの定義（案）

本ガイドラインにおけるフィジカルAIとは、センサ等によるセンシングを通じて物理環境の情報を取り込み、AIモデルによる処理を経て、設定された目的を達成するための最適な方策を自律的に推論・判断し、アクチュエータ（駆動系）等を介して物理的な行動へとつなげるシステムであり、サイバー空間での処理に留まらず、現実世界に対して直接的な働きかけ（移動、操作、加工など）を行うことを特徴とするものとする。

### リスクベースアプローチに資する内容の追記

✓ リスクの大きさ/発生可能性等を加味して対策の優先順位を検討するという考え方に基づくリスクベースアプローチの説明を追加

✓ 参考文献の追加

– AIガバナンス協会「AI時代の経営意思決定とガバナンス ～攻めのAIガバナンス実現のための戦略レポート」

– EU「Artificial Intelligence Act Annex III: High-Risk AI Systems Referred to in Article 6(2)」

# 「AIのセキュリティ確保のための技術的対策に係るガイドライン」(案)

ガイドライン案の本文は主に3つの項目で構成されており、2つの別紙、参考により構成されている。

## 本文目次（大項目と、一部項目抜粋）

本ガイドラインの策定の背景等

1 ガイドライン案のスコープ

2 脅威

2.1 対象とする主な脅威

2.1.1 プロンプトインジェクション攻撃

2.1.2 DoS攻撃（サービス拒否攻撃）

2.2 その他の脅威

3 脅威への対策

3.1 対策の位置付け

3.2 対策の概観

3.3 AI開発者における対策

3.4 AI提供者における対策

3.5 AI開発者・提供者に係るその他の基本的な対策等

3.6 AIサービスの想定事例に応じた分析

用語集

## 別紙目次（大項目と、一部項目抜粋）

別紙1 対策の詳細

1 AI開発者における対策

1.1 安全基準等の学習による不正な指示への耐性の向上

2 AI提供者における対策

2.1 システムプロンプトによる不正な指示への耐性の向上

2.2 ガードレール等による入出力や外部参照データの検証

2.3 オークストレータやRAG等の権限管理

別紙2 画像識別AI（CNN）に対する脅威と対策

参考 新たな脅威・対策に係る情報源の例

# Chief AI Officer (CAIO) ガイドブック(案) CAIO設置・AIガバナンス実務マニュアル(案)

本ガイドブックは、主に民間事業者がCAIO (Chief AI Officer) を設置・運用する際の標準的な実務指針を提供し、AIによる価値創出 (ROI最大化) と責任ある活用 (リスク低減・規制遵守) の両立を支援することを目的としている。CAIOの役割を明確化したうえで、組織設計、プロセス、評価・監督、教育、人材、調達などの要素を統合的に示し、各社が自社の文脈に合わせて着実にAI活用を進められるようにすることが狙いとされている

## CAIO設置の目的と役割

AISI Japan  
AI Safety  
Institute

CAIO: 価値創出とリスク統制を統合し、部門横断でAI活用を前進させる“司令塔”

- ◆ 生成AI/エージェント型AIが事業インフラの一部となる中で、価値創出と、権利・安全・公平性・プライバシー等の保護の両立が必須
  - ◆ 既存体制 (CIO、CDO等) だけでは「推進×統制」を継続運用しにくい場合がある
- ↓
- ◆ **AIに関する「単一の責任点」としてCAIOを設置**

### メモ

- CAIOガイド・マニュアルは、AIを“経営課題”として扱うための、戦略・ガバナンス・リスク・人材・調達の留意点を統合する実務的な参照資料
- 想定読者
  - CAIOガイドブック: CAIO本人、CAIOを任命する経営層
  - CAIO設置・AIガバナンス実務マニュアル: CAIO本人 + 実務担当スタッフ

### CAIOの役割領域 (5領域)

|         |   |
|---------|---|
| 戦略      | <ul style="list-style-type: none"><li>• ビジョン/ロードマップ</li><li>• 投資判断 (KPI/ROI)</li><li>• 重点ユースケース選定</li></ul>             |
| 企画～導入管理 | <ul style="list-style-type: none"><li>• 開発/調達/導入のゲート監理</li><li>• 技術・サービス選定</li><li>• データ戦略 (CIO/CDO) 連携</li></ul>       |
| ガバナンス   | <ul style="list-style-type: none"><li>• リスクマネジメント体制・意思決定</li><li>• 証跡 (台帳、AIIA等)</li><li>• 高リスクユースケースの停止・差し止め</li></ul> |
| 変革・人材   | <ul style="list-style-type: none"><li>• 業務再設計・定着</li><li>• 採用/育成/リスクリング</li><li>• 全社リテラシー向上</li></ul>                   |
| 内外連携    | <ul style="list-style-type: none"><li>• 経営層・取締役会報告</li><li>• 社内外ステークホルダー連携</li><li>• 対外説明</li></ul>                     |

## — エグゼクティブ・サマリ

### AIシステム向けインシデントレスポンスの枠組み「AI Incident Response System (AI-IRS)」について

近年、AIが意思決定や業務効率化など**事業の中核として活用**され始めてきています。一方で、仮に中核を担うAIの誤判断やサイバー攻撃による停止などが発生した場合、事業活動に**大きな被害をもたらす**可能性があります。特にAIでは、ブラックボックス性（**処理過程が不透明で、意思決定の内実を非常に把握しにくい性質**）や自律性（**目標達成に向けて人の介入なく自ら学習・判断・行動等を行う性質**）などの性質により、AIシステムではリスク対策を実施してインシデントの発生を未然に防ぐ**予防的統制**が、従来システムに比べて難易度が高いと見込まれます。AIシステムではゼロリスクの達成が困難である前提に立ち、インシデント発生時にも事業影響を最小化する取り組みが重要です。

これらを背景として、従来の情報システムにおけるインシデントレスポンスをAIシステムにも適用・拡張させる新たな枠組みであるAI-IRSを提示します。AI-IRSでは、**観測性（可視化）**と**制御性（封じ込め）**を中核に、運用中の挙動を把握し、問題箇所を切り離すことで**被害を最小化**する枠組みです。既存の開発・運用プロセスに容易に追加でき、運用負荷とインシデントの影響の双方を抑えられます。

また、AI-IRSは組織内部の取り組みに限定されず、サプライチェーンにまたがるAIライフサイクルの全体を観測・制御し、異常の検知・封じ込め・復旧を自律化する**共通基盤の整備**に向けたビジョンも示しています。

この枠組みにより、AIシステムを用いて事業活動を行う組織が、**AIを制御可能な戦略資産**として位置づけ、運用のレジリエンスを確立し、事業継続性を確保し、さらには**イノベーションの加速を促進**することが重要になります。組織や社会における整備・運用を進める際のきっかけや参考として、ぜひご活用ください。

# 生成AIの適切な利活用等に向けた知的財産の保護及び透明性に関するプリンシプル・コード（仮称）（案）概要

## 1. 目的

AI基本法の趣旨を踏まえ、技術の進歩の促進と知的財産権の適切な保護を両立し、利用者等が安心・安全に生成AIを活用できる環境を確保できるよう、EUにおけるAI法や、コーポレートガバナンスの分野におけるスチュワードシップ・コード等の取組も参考にしつつ、生成AI事業者が取るべき透明性の確保や知的財産権の保護の原則を定める。

具体的には、コンプライ・オア・エクスプレインの手法に基づき、プリンシプル・コードで定めた原則について、実施（コンプライ）するか、しない場合はその理由を外部に説明すること（エクスプレイン）を求めるものとする。

## 2. 対象

以下の生成AI開発者及び提供者（総称して生成AI事業者）を対象とする。

- 1.生成AI開発者：生成AIシステムを構築し、公衆に提供している事業者
- 2.生成AI提供者：開発された生成AIシステムをアプリケーション等に組み込み、生成AIサービスとして公衆に提供している者
- 3.海外企業であっても、日本向けに生成AIシステムやサービスを提供している場合には対象とする。

## 3. 原則

### 【原則 1：透明性と知的財産権保護のための概要開示】

生成AI事業者は、自社のウェブサイト等において、開示対象事項の概要を公開し、誰でも閲覧できるようにすること。

### 【原則 2：権利侵害を主張する者からのURL等の開示要求への対応】

生成AI事業者は、訴訟等の法的手続を行う者等から、当該者のコンテンツが存在するURLを示して、当該URLが学習データとして用いられているか否か等について開示の求めがあった場合には、その者に対して、当該事項に関する回答を行うものとする。

### 【原則 3：生成AI生成物と類似するコンテンツが存在する場合の開示】

生成AI事業者は、生成AIを用いて生成物を創出した者から、当該生成物と同一又は類似するコンテンツが存在するURLを示して、当該URLが学習データとして用いられているか否か等について開示の求めがあった場合には、その者に対して、当該事項に関する回答を行うものとする。

## 4. 受入れに係る手続

- ・プリンシプル・コードを受け入れる事業者は内閣府知的財産戦略推進事務局に届け出るとともに、自社のウェブサイト等でその旨を公表。
- ・内閣府において、参考となる様式や届出事業者一覧等を公表予定。

## 2. 諸外国の規制・ガバナンス動向

# 諸外国における政府のAI調達利用ルールとの比較 (公開情報に基づきデジタル庁にて調査・作成)

| 評価軸                             | 国名  |   |  |   |  |   |
|---------------------------------|---|---|--|---|--|---|
|                                 | 日本  | アメリカ  | イギリス   | カナダ   | オーストラリア  | シンガポール  |
| 【組織体制】<br>政府横断のガバナンス等           | 各省ごとにCAIOを置き、利用する生成AIを把握。<br>各省は、リスクの高い生成AIの利用についてアドバイザリーボードと協議の上判断       | 各省庁でCAIOの任命・省庁AIガバナンスボードの設置・CAIO協議会の開催の定めがある。<br>CAIOを核として省庁内連携による効果的なAI調達の促進やAI調達に係る情報や知見を省庁間で共有することが求められている | 共同で問題を解決する場として、既存の政府横断AIコミュニティを活用することを推奨。省庁間連携、ベストプラクティスの共有を実施。                            | 政府全体での知識共有や進捗モニタリングの仕組みを設けるとともに、全省庁がAI戦略の実行に関与し、既存AI活用事例の報告、課題やニーズの特定、インフラ・人材の確保などを担うことが求められている                   | 各省にてAIに関して、Accountable Officials(CAIO相当の役職)を設置。高リスク案件の報告、最新動向の把握や省庁間の窓口を担う。政府全体でAIユースケース台帳、戦略・運用方針・透明性声明・職員研修等の実施を義務化      | 政府横断のAI推進体制やガバナンスが推進され、リスク対応・サプライチェーン全体での責任の明確化・各規制当局のベストプラクティス共有が進められる |
| 【企画時】<br>生成AIのリスクについての規定・整理     | 各省は、リスクレベルについて、ガイドラインが示す利用範囲・扱うデータの機密性等の観点から判断                            | 各府省はHigh-Impact AIかどうかの判定をする。その出力が、法的、重要な、拘束力のある、または権利や行動に重大な影響を及ぼす決定や行動の主要な根拠となる場合、AIは影響力が大きいと判断             | AI影響度評価によるリスクの特定・軽減を実施。<br>重要決定時や設計変更時にも再評価  | プロジェクトのリスクを4段階で評価する基準(AIA)を整備。当該評価をプロジェクトの設計フェーズ開始前までに完了させることを要求  | 全AIユースケースについて事前にツールを用いた影響評価を行い、用途・期待効果・固有リスクを特定することを義務化。高リスクの場合はAccountable Officialへの報告が義務付けられている                         |   |
| 【調達時】<br>調達時のチェック項目の整理          | 調達時の確認事項(例：入出力データ等の適切な管理、有害情報の出力制御等)について整理。<br>各省が利用ケースに照らして、実際に確認する事項を判断 | 連邦政府調達における米国製AIの重要性を取り上げている。<br>各府省は要件と契約条件を明確化するために、目的・業務上の効果・システム要件・技術性能等に基づくベンダー評価を強く奨励されている               | 調達時の検討事項10点(ガバナンスやデータ保護の計画等)とフェーズ別(企画、公募、評価選定、契約・継続管理)の留意事を整理                              | 生成AIツールの調達時に、プライバシー・セキュリティ・法務の観点からリスク評価や利用規約等の確認を推奨。AIポータルでは要件を満たした「Qualified Suppliers」リストが公開され、効率的かつ信頼性の高い調達を実施 | 契約時に考慮すべき観点を5つ(AIの倫理原則、明確な説明責任、データの透明性、情報資産へのアクセス、AIのパフォーマンステスト)に整理。AI技術の急速な変化によるリスクへの対応等を求める。また、AI調達契約書の参考テンプレートを公開(任意使用) | 調達について政府内の専門機関が相談を受付。汎用的な既存行政AIサービスの優先検討を求めつつ、政府認定企業の製品やサービスからも選択可能     |
| 【運用時】<br>リスクケース・インシデントへの対応ルール※1 | リスクケース発生時の対応事項について整理。<br>提供者がインシデントの内容や行う対応等について、AI統括責任者及びアドバイザリーボードに報告。  | リスクケース・インシデントへの対応ルールの明記はなし。ただ、AIの性能が適切なレベルに達していない場合の使用の中止を求める規定が存在  | ライフサイクル全体でリスクを管理し、エスカレーションのプロセスを文書化することを推奨。リスクケース発生時の報告先や対応者の明確化、影響を受けた個人に対する救済措置の確立を求めている | 起こり得るAIのサイバーセキュリティ事象とインシデント事象について整理の上、対応計画の策定を義務化   |  |   |

【凡例】  規定あり  規定あり(限定的)  具体的な規定なし

# (参考) EU AI ActにおけるAIのリスク分類

全てのAIシステムは4つのリスクへ分類※1

許容できないリスク

人の生命や基本的人権に対して、直接的に脅威をもたらすと考えられるAIシステム

禁止



ハイリスク※2

人の健康や安全、基本的人権、または社会的／経済的な利益に影響を与える可能性があるAIシステム

ハイリスクAIの義務の遵守  
※3



限定的リスク

深刻なリスクはないが、透明性に関する特定の要件を満たす必要があるAIシステム

透明性の義務の遵守



最小リスク

リスクがごくわずか、またはリスクを伴わないAIシステム

制限なし



※1 AIシステムの特性によっては複数のリスクに該当する場合があります。

※2 欧州委員会は、AI法におけるハイリスクAIシステム規制の適用開始を、技術標準やガイドライン整備の遅れを理由に最長16カ月延期し、2027年12月までとする方針を発表しました。適用開始までの期間に、企業は必要な支援ツールを整備・確保した上で、義務を円滑に履行できる体制を整えることが可能となります。

※3 ハイリスクのAIシステムには、「リスク管理システム」「データガバナンス」「技術文書、ログ管理」「透明性」「人による監視」「サイバーセキュリティ」などの観点で定められた要求事項の遵守を求めています。

汎用目的AIモデルは2段階に分かれる

汎用目的AIモデル（GPAI）は他のAIシステムに組み込まれて使用される可能性が高いため、GPAIを提供する事業者（プロバイダー）にはさまざまな義務が課せられます。

また、GPAIの中でも一定の条件を満たすAIモデルはSystemic riskがあると判断され、より厳しい要件への対応が求められます。

Tier 1: 全ての汎用目的AIモデル

Tier 2: Systemic risk  
• 高影響が想定  
• モデルのトレーニングに使用される計算量が  $10^{25}$  FLOPs

- モデルに関する仕様書などを最新化すること
- EU著作権法を遵守すること
- 適切な情報提供を実施すること
- 規制当局に協力すること

- Systemic riskを評価し、そのリスクを低減すること
- インシデント発生時に当局へ報告すること
- サイバーセキュリティに関する保護を実施すること

など

など

2025年7月10日、「汎用AIの行動規範」が公開されています。

法的拘束力はありません。①透明性、②著作権、③安全性・セキュリティの3つの章で構成されています。

欧州委員会は2026年2月2日までに、どのようなユースケースがハイリスク/非ハイリスクに該当するかの具体例を示すガイドランスを公表予定であったが、2026年2月中旬時点では未公開となっている。

<https://artificialintelligenceact.eu/implementation-timeline/>

<https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal>

# (参考) 米国: WokeAI防止関連のガイダンス

## ■「バイアスのないAI原則によるAIへの公共信頼性の向上」

正式名称: M-26-04 Increasing Public Trust in Artificial Intelligence Through Unbiased AI Principles

発表年月日: 2025年12月11日

米国行政管理予算局がWokeAI※防止の大統領令を調達プロセスにおいて実現させるためのガイダンスである覚書 (M-26-04) を公開。米連邦政府が調達するLLMやAIが Woke AIにならないよう、AIに「真実性」と「イデオロギー中立性」を義務付け、その遵守を契約条件として強制することで、透明性と監督を強化し、公的信頼を高めることを目的としている。

※Woke AIとは: 特定の社会思想 (多様性・公平性・包摂性) に基づく価値観をAIモデル内に組み込み、その結果として事実の歪曲、人物の表象操作、人種・性別に関する偏った判断を行うようになったAI

### 文書の目的

- 政府が調達するAIが「イデオロギー的偏りなく、事実に基づく」出力を生成することを義務付ける

### バイアスのないAI原則

#### (1) 真実追求 (Truth-seeking)

- 事実の正確性・科学的検証・客観性を優先す
- 不確実性がある場合は明示する

#### (2) イデオロギー中立性 (Ideological Neutrality)

- 明示的にユーザーが求めない限り、政治・思想的に偏った判断を出力に含めてはならない

### 適用範囲

- 全ての連邦政府機関が対象
- 調達方法を問わず全てのLLMに適用
- 国家安全保障システムは形式上除外だが、可能な範囲で適用を推奨

### 調達要件

- 新規LLM契約にはこれらの原則遵守を必須項目として組み込む
- 違反したベンダーとの契約は終了することができる

### 運用開始後の監督責任

- 調達時だけでなく、導入後も継続的な監視・評価が必要

### 透明性要件

#### A) 最低限の透明性確保

- 利用規約、モデルカード・システムカード等の文書、エンドユーザー向け資料、エンドユーザーからのフィードバックの仕組み

#### B) 高リスク向けの追加の透明性要件

- 事前学習・事後学習に関する情報 (システムプロンプト、レッドチーミング結果等)、モデル評価、出力根拠の提示等

### 既存契約の更新

- 可能な範囲で、既に締結済みのAI契約も本原則に合わせて修正する

各省は2026年3月11日までに調達ポリシーを更新することが求められている

AIリスクレベルに応じて2段階の透明性要件を要求

# (参考) EU加盟国別の政府調達・利活用ルール (1/2)

## ■ドイツ「連邦行政における人工知能利用ガイドライン」

正式名称：Leitlinien für den Einsatz Künstlicher Intelligenz in der Bundesverwaltung

発表年月日：2025年3月27日

本ガイドラインは、ドイツ連邦行政における責任ある安全なAI導入を実現するための基本原則を示すものである。個別案件ごとのリスク評価を前提に、各機関の責任者が導入可否を判断し、適切なAIシステムを選定するための運用指針を提示している。

また、AI利用に関する具体的な行動指針を、政府横断の対応事項として、利用者と提供者の双方の視点から体系的に示している。EU AI Actのリスクに関する記載もされているが、明確なリスクベースアプローチの記載は確認できない。

### 1. 導入

行政におけるAI活用の意義、必要性、社会的背景

### 2. 行動の共通基盤としての価値観に基づく指導原則

- ① チャンス志向かつ責任あるAI利用
- ② 価値基盤・人間中心のAI
- ③ 信頼できる（ロバストな）AI
- ④ 透明性・説明可能性
- ⑤ 持続的で責任ある運用

### 3. AI利用の具体的な指針

- 3.1 利用者（職員）向け指針（A1～A5）
- 3.2 提供者（省庁）向け指針（B1～B9）

### 付録：大規模言語モデル（LLM）の利用

行政機関でLLMを安全・有効に使うための手順と注意点を整理

### 【利用者向け指針】

- A1. 倫理的かつ責任あるAI利用（人間としての最終的な判断責任の維持）
- A2. 個人データ開示の最小化（外部システムを利用時等の最小限の情報提供）
- A3. データ入力責任ある取り扱い（データ入力可否の事前確認）
- A4. AI生成コンテンツのチェック（誤情報や不正確な内容が混入する可能性の理解）
- A5. AI利用の透明性確保（AIを利用した旨を必要に応じて上司・同僚に明示）

### 【提供者向け指針】

- B1. 明確な意思決定（どのAIをどの目的で使うか責任を持って判断）
- B2. 利用目的・利用範囲の明確化（適用分野および目的の定義）
- B3. 目的に合った適切なAIシステムの選定（法令遵守、技術的適合性、性能の考慮）
- B4. 倫理的かつ慎重な利用（基本権保護、差別禁止、情報セキュリティの確保等）
- B5. 職員のリテラシー向上（研修の提供）
- B6. AI利用の透明性確保（どの業務でどのAIが使われているか、組織的に把握・公開）
- B7. データ開示を最小化する仕組みの整備（利用者が過剰にデータを開示しない環境の提供）
- B8. 適切なデータ入力環境の整備（ルール、手順、技術的制約等の組織的な整備）
- B9. GDPR等のデータ保護法制に準拠したAI利用（データ保護への準拠の徹底）

# (参考) EU加盟国別の政府調達・利活用ルール (2/2)

## ■フランス「AIソリューションの責任ある調達のための実践ガイド」

正式名称：Fiche pratique pour l'Achat Responsable de solutions d'Intelligence Artificielle

発表年月日：2025年12月5日

行政機関がAIを調達する際の実務的なガイドであり、調達前の事業者への確認事項や、調達文書に盛り込むべき要件が示されている。選定時の判断軸として、環境配慮の視点が重視されている。政府横断的なガバナンス構築やリスク規定等に関する記載はなく、主に調達時の確認項目の列挙となっている。

### 1. AI とは何か？

- ・AIがもたらす環境・社会への影響

### 2. AIと責任あるデジタル技術

- ・AI 導入前に確認すべきこと（AIを使う必要が本当にあるか、非AIでの解決策では不足か、想定されるリスクと便益の比較）
- ・AI 利用が正当化できる場合の考慮事項（データ/学習量の最小化等）

### 3. 調達者向け：事業者への確認事項

- ・モデルの環境負荷測定、バイアスを含まないための手法、データ管理、再生可能エネルギー利用率、水/CO2/電力の効率 等

### 4. 調達者向け：調達文書に盛り込むべき条項

- ・行政特約、技術仕様書で求めるべき条項の例（研修、報告義務や改善義務、環境影響の測定、データ管理 等）
- ・調達要領で設定する評価項目（環境・社会的配慮）

### 5. 補足資料

- ・関連規則・ガイドライン一覧、環境評価ツール一覧

### 6. 付録

- ・事業者ヒアリング質問リスト、調達文書に追加できる条項サンプル集

# (参考) シンガポール: エージェント型AIのためのモデルAIガバナンスフレームワーク

## ■「エージェントAIのためのモデルAIガバナンスフレームワーク」

正式名称：Model AI Governance Framework For Agentic AI

発表年月日：2026年1月22日

既存のModel Governance Framework for AIをエージェント型AIに拡張し、組織に対して、責任を持ってエージェント型AIを配置する方法についての指針を提供するもの。リスク軽減のための技術的および非技術的対策を推奨しつつ、最終的には人間が責任を負うことが強調されている。

### 本文書におけるエージェント型AIの定義 (参考訳)

「エージェント型AIシステムとは、AIエージェントを用いて、特定の目的を達成するために複数のステップにわたって計画を立てることができるシステムである。(中略) エージェントは通常、ユーザーが定義した目標を達成するために、複数のステップにわたって、ある程度独立して計画し、行動(例: ウェブ検索やファイル作成)を行う能力を有する。」

従来のソフトウェアの脆弱性やLLM特有のリスクに加え、エージェント特有の構成要素によってこれらのリスクが新しい形で現れたり、増幅されたりする

### エグゼクティブサマリ

#### 1章 エージェント型AIの概要

##### 1.1 エージェント型AIとは何か

- 1.1.1 エージェントの基本構成要素
- 1.1.2 マルチエージェント構成
- 1.1.3 設計が能力と限界に与える影響

##### 1.2 エージェント型AIのリスク

- 1.2.1 リスクの発生源
  - エージェント内部のコンポーネント由来のリスク
  - システムレベルで生じるリスク
- 1.2.2 リスクの種類
  - 誤った行動
  - 権限外の行動
  - バイアスのある不公平な行動
  - 情報漏えい等のデータ侵害
  - 接続システムへの障害

### 2章 エージェント型AIのためのモデルAIガバナンスフレームワーク

#### 2.1 事前にリスクを評価し、境界を設定する

- 2.1.1 適切なユースケースの選定
- 2.1.2 設計によるリスクの抑制

#### 2.2 人間が意味のある形で責任を持つ

- 2.2.1 人間側での組織内外の明確な責任分担・役割決め
- 2.2.2 意味のある人間の監督設計

#### 2.3 技術的コントロール・プロセスの実装

- 2.3.1 設計・開発時の技術的コントロール
- 2.3.2 デプロイ前のテスト
- 2.3.3 デプロイ後の継続的なモニタリングとテスト

#### 2.4 エンドユーザーの責任ある利用を可能にする

- 2.4.1 ユーザーのタイプ別の対応
- 2.4.2 エージェントと対話するユーザー
- 2.4.3 業務に組み込むユーザー

### 付録

Annex A：参考資料 Annex B：フィードバック・事例募集

# デジタル庁

Digital Agency