第６回先進的AI利活用アドバイザリーボード

行政の進化と革新のための生成AIの調達・利活用に係るガイドライン充実に向けた検討事項案

2026/9/29　　先進的AI利活用アドバイザリーボード事務局

© Digital Agency, Government of Japan

デジタル庁
Digital Agency

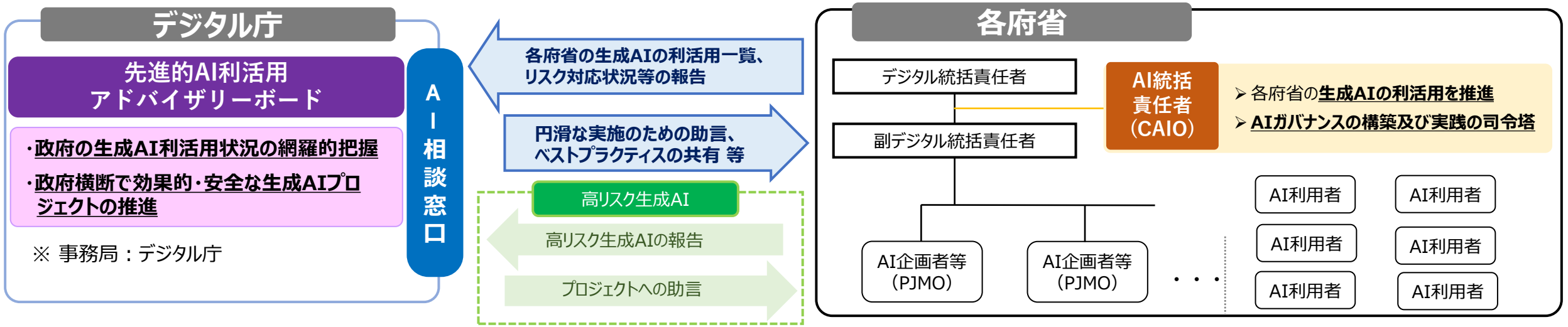
行政の進化と革新のための生成AIの調達・利活用に係るガイドラインのポイント

（１）ガイドラインの目的・枠組み等

目的：**生成AIの利活用促進とリスク管理を表裏一体で進める**ため、**政府におけるAIの推進・ガバナンス・調達・利活用のあり方**を定めるもの。
対象：**原則として、入力はテキスト及び音声、出力はテキスト、画像又は音声が可能な生成AIシステム**　※特定秘密や安全保障等の機微情報を扱うシステムは対象外
適用開始時期：**令和７年５月に運用開始。2.0版の内容は、令和８年９月１日より施行。（AIガバナンスの枠組み対象は令和８年７月１日から）**

（２）政府における生成AIの推進・ガバナンス体制の構築

- 比較的**高リスクとなる可能性がある生成AIの利用**であっても、**先進的AI利活用アドバイザリーボード**の各府省への助言や**相談窓口**等の仕組みを通じ、**安全かつ効果的AIプロジェクトとしての実施をサポート**し、先進的生成AIの利活用を促進。
  ※サプライチェーンリスクも考慮
- 各府省の**AI統括責任者（CAIO）**が、**生成AIの利活用を把握・推進し、ガバナンス、リスク管理を総括**。



（３）生成AIの調達・利活用ルール

※　各府省生成AIシステムの①**AI統括責任者（CAIO）**、②**企画者**、③**提供者**、④**利用者等**毎にルールを規定

- AI統括責任者（CAIO）は、各府省の利用者（職員）に向けて**生成AIの利用ルール**を策定。
- 企画者・提供者は、本ガイドラインの**「調達チェックシート」**及び**「契約チェックシート」**を参考にして仕様書作成や事業者との契約等を行うことにより安全かつ品質の高い生成AIシステムの調達を確保。運用開始後も適切な利用や安全性や品質の確保を定期的に検証。
- 提供者及び利用者は**リスクケースが生じた場合、適切に各府省AI統括責任者（CAIO）に報告し、提供者が必要な対応を実施**。先進的AI利活用アドバイザリーボードは各ケースの報告を受け、必要に応じ再発防止策等を検討。

（参考１） 別添各シート（高リスク判定シート、調達チェックシート、契約チェックシート）について

１．高リスクAIの判定についての参考とする判定軸（高リスク判定シート）

３つの観点を勘案し、
アドバイザリーボードに助言を求めるべきか各省において判定

A.生成AI利活用業務の性格、 B.利用範囲、C.出力結果の政府職員等による判断を経た利活用

２．調達・契約時のチェック項目について（調達チェックシート、契約チェックシート）

（１）調達チェックシート

生成AIを調達する際に**事業者への要求事項**として、**仕様書等に盛り込むべき項目を整理**

ガバナンス項目

①AI事業者ガイドライン共通の指針の遵守　②AIガバナンス（AIのメリットを最大化しつつ、リスクを統制する体制）の構築
③生成AIシステムの品質向上のため、AI業界や最新技術等の動向を把握する
④情報セキュリティインシデント・生成AIシステム特有のリスクケース発生時の対応手順の整備
⑤生成AIシステムの開発・運用に従事する者または組織についての生成AIに関する教育・リテラシー向上

開発・運用プロセス要件項目

⑥生成AIシステムへの入出力または処理されるデータの取扱いの適切な管理
⑦生成AIシステムの期待品質を満たすための取組
⑧ベンダーロックインの回避　⑨生成AIモデルのアップデートの考慮
⑩文化的・言語的考慮　⑪環境への配慮

生成AIシステムの要件項目

⑫有害情報の出力制御措置　⑬偽誤情報の出力・誘導の防止措置
⑭公平性と包摂性の確保（バイアスや差別を含む出力の抑止措置）　⑮目的外利用の防止
⑯個人情報、プライバシー、知的財産に関する適切な取扱い　⑰セキュリティ確保
⑱説明可能性の確保　⑲ロバスト性（出力の安定性）　⑳データ品質　㉑検証可能性

（２）契約チェックシート

生成AIを調達する際に**契約書で取り決めるべき項目を整理**

- 生成AIシステムに係るインプットに関する取り決め（学習の有無、データの保存方法等）
- 生成AIシステムに係るアウトプットに関する取り決め（アウトプットに関する一定の保証、アウトプットの権利帰属）
- 生成AIシステムに係る契約上の取り決め（事業者が期待品質を満たすための取組の履行又は完成させる義務）
- 生成AIシステムに係る成果物の取り決め（成果物の定義や成果物の権利帰属等）
- 生成AIを含むインシデントが発生した場合の事業者の対応義務及びその範囲に関する取り決め（被害を最小限に食い止めるため、また、原因を特定するための情報やデータの提供を含む）　等

※チェックシートの内容の適用項目や適用方法については、個別の生成AIシステムの調達形態やリスクレベルに応じて、AI相談窓口とも連携しつつ各府省において判断。

（参考２）「生成AIシステムの利活用ルール」について

１．府省毎の利活用ルール

- **各府省のAI統括責任者（CAIO）**が、適切なAIの利活用を促進するため、**「生成AIシステムの利活用ルールひな形」に基づき、各府省の利用者（職員）に向けて生成AIの利活用ルールを策定・周知。**

２．生成AIシステム毎のルール

- **企画者**が、**システム毎に、利用目的をはじめ、利活用に係るルールを策定し、各システムの利用者（職員・国民等）に周知。**

◆1. 府省毎の生成AIシステム利活用ルールの「ひな形」の構成と概要

（構成）

1．ルールの目的
2．生成AIシステムの利用に係るルール
（１）利用前のルール
（２）利用中のルール
① 入力データ又はプロンプトにおけるルール
② AI生成物利用におけるルール
3．問い合わせ先

（概要）

- 生成AIの**リスク**についての理解
- 生成AIシステムへの**入力結果や出力結果**が**必要に応じシステム側に提供**されることの理解
- **約款型クラウドサービス**は**原則として要機密情報を扱えない**ことの理解
- **国外サーバを利用**する生成AIの場合に現地政府による**データの検閲や接収を受ける可能性**があることの理解

- **要機密情報や個人情報の入力の可否**を含む利用方法の遵守
- 生成AIシステムの**利用目的の範囲内での利用**
- 正確かつ最新のデータの入力
- **知的財産権等の侵害等リスクを低減**させるよう対策を取ること

- 生成AIの**出力に基づいて行われた判断**についての**説明責任**についての理解
- 出力結果に**含まれうるバイアスを踏まえて**業務に出力結果を活用すること
- 出力結果の**正確性や根拠、事実関係等を必要に応じ確認**すること
- （第三者の著作権等の侵害の有無を含め）**安全性・公平性、客観性、中立性**に問題がある出力でないかを確認し、問題点は必ず加除修正の上で利用すること
- リスクケースやその兆候を検知した場合の**迅速な各府省報告窓口への報告**
- **著作物の著作権、その他知的財産権等を侵害等するものでないこと**等を確認すること

ガイドラインにおける拡充・見直しに関する記述

- 行政の進化と革新のための生成 AI の調達・利活用に係るガイドライン（第2.0版）においては、主に政府における生成AI施策の動向や利用状況、リスクの顕在化状況等を踏まえて随時見直しを実施していくこととしている

2.2.2 本ガイドラインが対象とする生成 AI

これらを構成要素とする生成 AI システムであって、画像や動画等を入力するもの、動画等を生成するもの、より高度なタスクを実行できるもの（AI エージェント14等）等については、政府における利活用状況に鑑み具体的対応事項は定めないが、AI ガバナンスの枠組みの対象とすることとし、（中略）**これらを含むその他の AI については、政府等における今後の利活用状況、国内外のルールの整備状況等を踏まえ、必要に応じ、本ガイドラインの拡充を検討する**こととする。

4.1.1 先進的 AI 利活用アドバイザリーボードの開催・AI 相談窓口の運用等

先進的 AI 利活用アドバイザリーボードは、**政府全体の生成 AI 施策の動向やガイドラインの運用状況を踏まえ**、関係府省庁等と連携を行いつつ**ガイドラインの見直しを行う**。

7 今後の進め方

日々技術進歩する生成 AI の政府調達・利活用においては、**今後想定されていなかったリスクが顕在化する可能性もあること等から、政府による生成 AI 調達・利活用ルールについては、随時見直していく**こととする。

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（1/10）

- ガイドラインの充実の必要性を踏まえ、検討事項候補を４つのテーマで整理している。

充実の必要性

- 政府AI調達・利活用ガイドラインについては、生成AIの技術面とユースケースの発展が著しく、今後想定されていなかったリスクが顕在化する可能性もあることから、随時見直しをすることとされている。
- ①政府におけるユースケースの広がり、②リスク緩和措置の必要性等の観点から、本年度のアドバイザリーボードにおいて論点とすべき事項や、その際に留意すべき事項について、ご意見を頂きたい。

検討事項候補を以下４テーマに整理

I. 利用実態等を踏まえた対象の拡充

II. AIエージェント特有の対応事項の導入

III. その他全体的な利便性の向上等

IV. 他ガイドライン等を踏まえた必要事項の拡充

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（2/10）

- ガイドライン上でも政府等のAI利活用状況等も踏まえた適用範囲等の拡充の検討求められている。AIエージェントの具体的規律や、画像入力及びその他のケースも具体的対応事項の対象とすることを検討してはどうか。

I.　利用実態等を踏まえた対象の拡充

AIエージェント

【背景・課題】

- 政府内でのAIエージェントの利活用の拡大が見込まれる中で、AIエージェントの一定の自律性を持ってタスクを実行する性質から、**予期しない動作をし、重大な影響を及ぼすリスクがあるため、具体的なガバナンスやルールを定めておく必要**がある。

【検討事項案】

- **利活用時の一般的なルール、リスク管理のあり方、調達上の要求事項等、具体的な規律を整備**すべきではないか（個別具体の論点候補はIIで提示）。

画像入力及びその他のケース

【背景・課題】

- 画像入力や動画出力等の利活用が一般化していく場合には、知的財産権等に応じた対策等が必要となり得る。
- 令和８年度第１回定期報告では、画像入力が増加、動画入力や動画出力は、増加しているがまだ少ない。

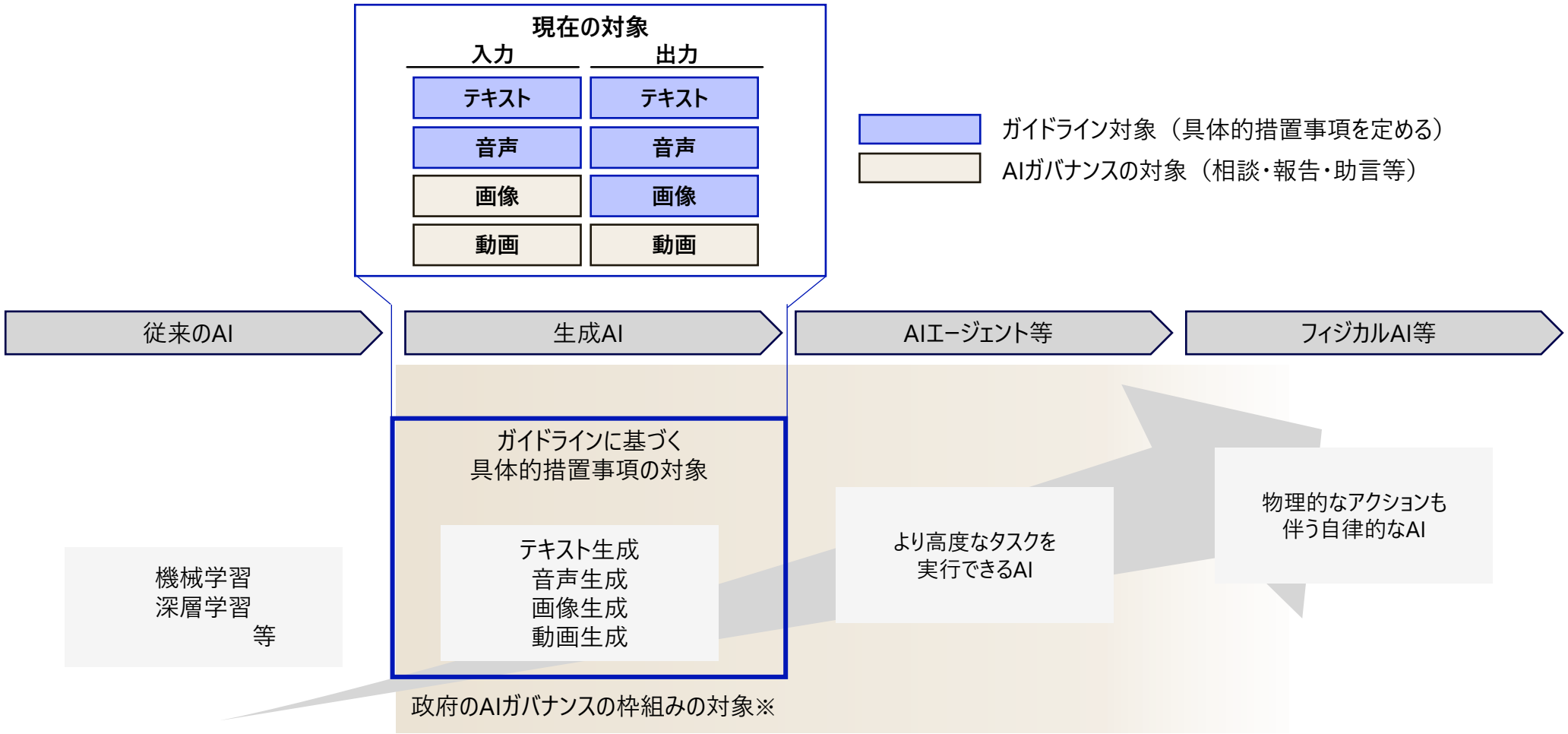
【検討事項案】

- **画像入力を新たに対象にするとともに対策事項として追加**すべきものがあるか検討すべきではないか。
- **動画等その他について対象とするかは**政府内の利用状況をもとに**引き続き検討**してはどうか。

© Digital Agency, Government of Japan

現在のガイドラインの対象

- 対象とする生成AIについて、政府のユースケース実態を踏まえて、音声入力と、また音声・画像出力へ拡充。また、その他の生成AIや生成AIを活用したAIエージェント等については、ガイドラインに基づく相談・報告・助言等の対象としている



※各府省庁のCAIOや3.1の「先進的AI利活用アドバイザリーボード」への報告、デジタル庁のAI相談窓口、「先進的AI利活用アドバイザリーボード」の助言等の対象

© Digital Agency, Government of Japan

現行の政府AI調達・利活用ガイドラインにおけるAIエージェントの定義等の記載

- 現行のガイドラインでは、AIエージェントの定義が既に脚注に示されており、当該定義は「AI事業者ガイドライン（第1.2版）」より抜粋している
- 今後の改定においても、AI事業者ガイドラインとの整合を確保しつつ、AIエージェント・エージェンティックAIの定義や位置づけを整理する必要がある

観点	行政の進化と革新のための生成AIの調達･利活用に係るガイドライン（第2.0版）	AI事業者ガイドライン（第1.2版）
AIエージェント、エージェンティックAIの定義の記載抜粋	**①AIエージェント** AI事業者ガイドラインにおいては、AIエージェントは次のように定義されている。「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。」（出典：「AI事業者ガイドライン」P.11）また、「自律」の語について脚注において、「高度な自律状態だけを指しているのではなく、ある程度の自律性を持つものも含む。」と記載されている。 （脚注14より抜粋）	**①AIエージェント** 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。 （参考としてISO/IEC 22989:2022 では、以下のように定義されている） 自動化された主体であり、環境を感知して応答し、自らの目標を達成するために行動を取るもの （「第１部AIとは　関連する用語」より抜粋） **②エージェンティックAI** AIエージェントよりも包括的かつ進化的な概念としてエージェンティックAIがある。これは、複数のAIエージェントにより自律的に意思決定を下しアクションを起こす目標主導型のAIシステムである。 （脚注18より抜粋）
その他、AIエージェント等に係る記載の抜粋	利用者が高度なタスクを実行できるAIエージェント等作成できる生成AIシステムについて、出力結果の適切さを判断せずにタスクを実行するものを対象に、作成の制限又は提供者・CAIOへの報告を求めることとしている	定義以外にAIエージェント・エージェンティックAIに関する記載はなし

© Digital Agency, Government of Japan

各国文書におけるAgentic AIの定義等の記載

- 基本的にはAgentic AIを主眼においた文書となっているが、主要なコンポーネントの考え方に大きな差分はなく、各構成要素を整理して記載がされている。
- シンガポールにおいては、セキュリティリスクの観点と対応させて整理している。

観点	国名：文書名				
	カナダ：Guide on the Use of Agentic AI	オーストラリア：Agentic AI addendum to the AI technical standard for Australian Government	シンガポール：Model AI Governance Framework for Agentic AI	シンガポール：Securing Agentic AI Addendum	Five Eyes：Careful adoption of agentic AI services
AI Agent、Agentic AIの定義記載概要	**①AI Agent** 特定の目標を達成し、変化する入力や状況に応じて行動を適応させるために、多くの場合自律的に、周囲の環境を認識し、それに基づいて行動できるシステムである。 **②Agentic AI** タスクを細分化し、連携し、外部ツールを活用し、限られた人間の監視下で長期間にわたり目標を追求できる、1つまたは複数の協調型AIエージェントで構成されるシステムを指します。	**①AI Agent** **・AI agent (component)** 環境からの入力を認識し、内部状態を維持または更新、APIまたはその他のツールを使用してアクションを選択して実行し、割り当てられた権限と制約内で定義された目標を達成するソフトウェアベースのシステム **②Agentic AI** **・Agentic AI (system pattern)** 環境を認識・推論し、内部状態を維持・更新し、推論によって目標達成方法を決定し、定義された権限と制約内で事前定義された目的を達成するために一連のアクションを独立して実行する、多段階のPRAサイクル*1を含む、協調的なAIエージェントまたはAIシステム	**②Agentic AI** 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の独立した計画、意思決定、およびアクション実行（たとえば、Webを検索したり、ファイルを作成したり）を備えています。1 エージェント型AIシステムは、個別にまたは協調して動作する1つまたは複数のAIエージェントで構成されるソフトウェアシステム	**②Agentic AI** 環境と相互作用し、データを収集し、指定された目標を達成するために自己決定タスクを実行する	**②Agentic AI** 1つ以上のエージェントで構成され、これらのエージェントは、LLMなどのAIモデルに根本的に依存して、世界の状況を解釈・推論し、意思決定を行い、行動を起こす
主要なコンポーネント	当該文書内では、詳細記載は確認できない	エージェント、モデル、メモリ、データ、ツール等図式化して記載	モデル、インストラクション、メモリ、データ、ツール等図式化して記載	エージェント、インストラクション、メモリ、データ、ツール等図式化して記載	インプット、LLMエージェント、ツール、外部データ、メモリ、アウトプット等図式化して記載
コンポーネントに対応するリスク	当該文書内では、詳細記載は確認できない	当該文書内では、詳細記載は確認できない	当該文書内では、詳細記載は確認できない	セキュリティの観点で、エージェント型AIシステムに対するOWASPの脅威分類を用いて対応関係を記載	当該文書内では、詳細記載は確認できない

公開情報に基づきデジタル庁にて調査・作成

*1PRAサイクルとは、Perceive（知覚）・Reason（推論）・Act（実行）のサイクルを指す

© Digital Agency, Government of Japan

各国文書におけるコンポーネントの図（1/3）

- オーストラリア、シンガポール（Model AI Governance Framework for Agentic AI）においては、エージェント、モデル、メモリ、データ、ツール等のAgentic AIの構成要素及び、関連するコントロール等を含めて図式化している

オーストラリア

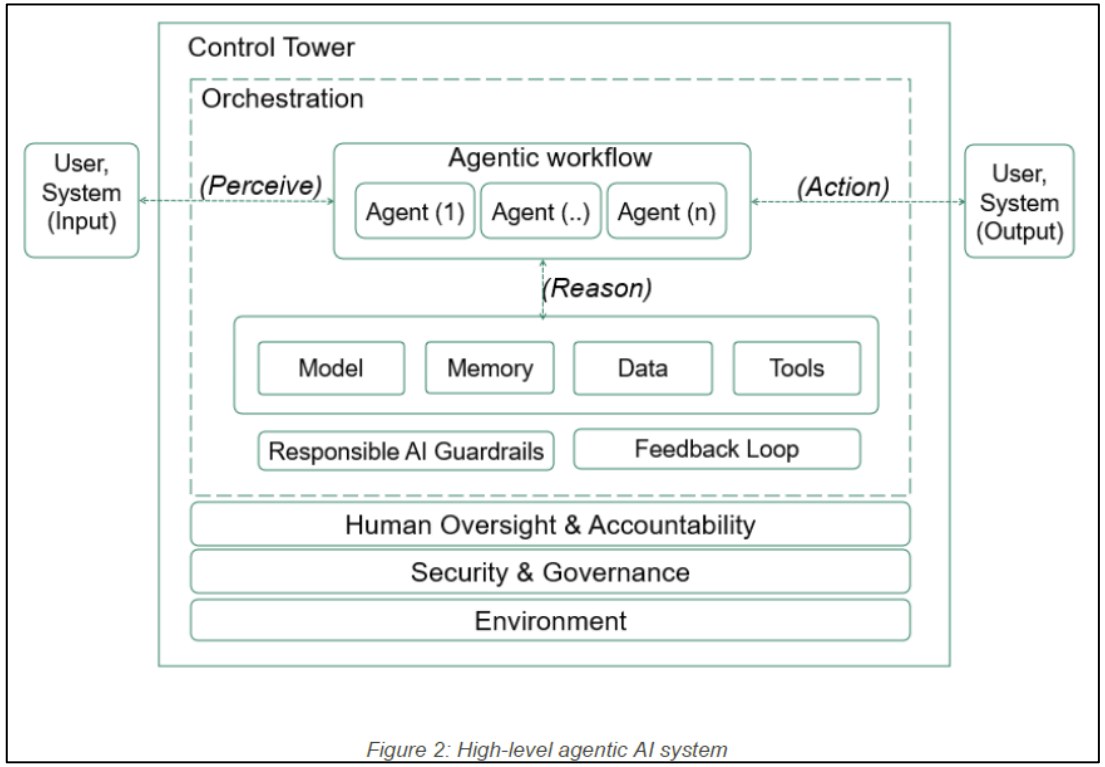
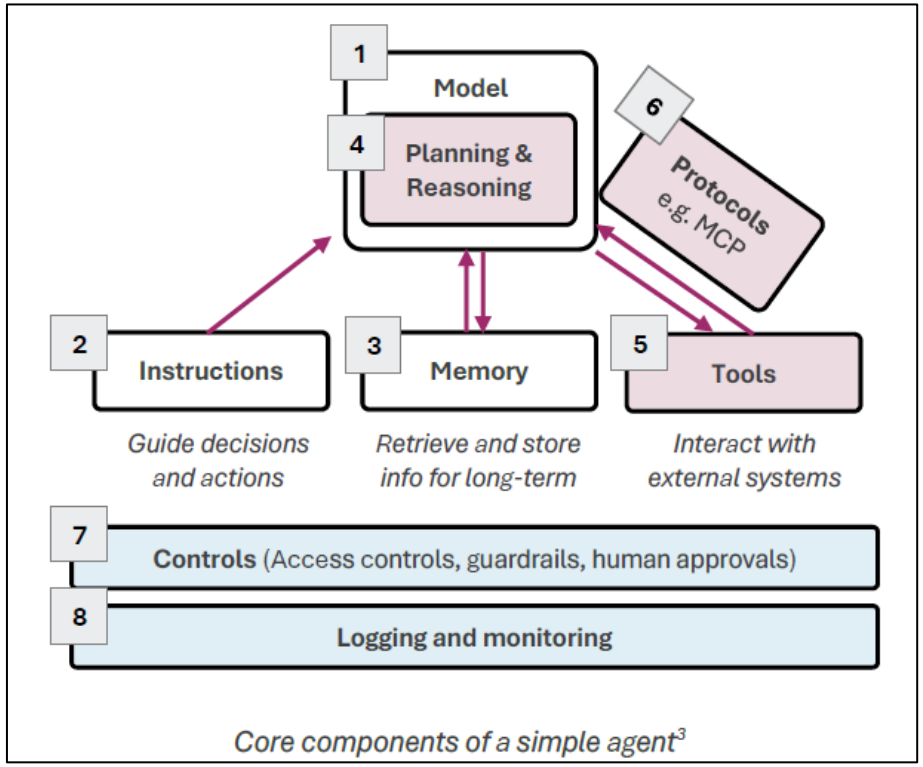


Figure 2: High-level agentic AI system

Agentic AI addendum to the AI technical standard for Australian Government、P.12より引用

シンガポール



Core components of a simple agent[3]

Model AI Governance Framework for Agentic AI、P.6より引用

© Digital Agency, Government of Japan

各国文書におけるコンポーネントの図（2/3）

- FIVE EYESにおいては、インプット、LLMエージェント、ツール、外部データ、メモリ、アウトプットの粒度感で整理されている

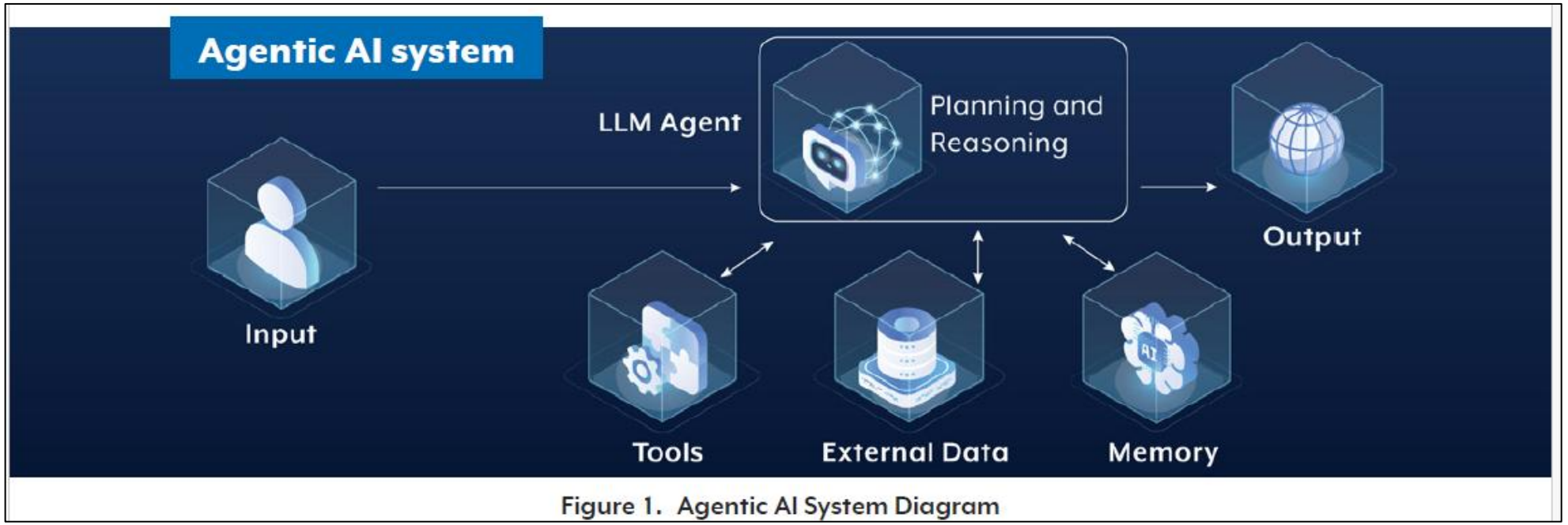


Figure 1. Agentic AI System Diagram

Careful adoption of agentic AI services、P.5より引用

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（3/10）

- AIエージェントの自律性やアクセス権限の拡大に伴う固有のリスクを踏まえ、ガバナンス、実行状況の監視、権限管理、サイバーセキュリティ等に関する対応事項を追加することについて検討を行うこととしてはどうか。

II. AIエージェントの対象拡充に伴う特有の対応事項の強化

AIエージェントにおけるリスクの考え方

【背景・課題】

- 一定の自律性を持ったタスクを実行するAIエージェントにおいては、**職員等による判断を必須とすると本来のメリットを損なうおそれがある**
- 他方で、職員等による判断を介在させずに実行すると、意図しない動作により重大な損害が生じる可能性もある。

【検討事項案】

- タスクの性質に応じて、ＡＩエージェントに付与する権限やアクセス可能な情報の範囲と、職員等の関与の在り方等を考慮した上で、AIエージェントにおける**リスクと便益のバランスの在り方を具体化**することが必要ではないか。

①自律行動に関するリスク

【背景・課題】

- AIエージェントが、決済や情報送信等を自律的に実行可能な場合、**指示の誤解や誤作動、予期せぬ動作等により、回復困難な業務損失が生ずる可能性**。

【検討事項案】

- **付与する権限は、業務遂行に必要な範囲に限定することを基本原則とし、最小権限となるよう設計・管理することが必要**ではないか。
- 送金処理、外部への情報送信等の**不可逆性の高いタスクの実行は、職員等による判断（HITL）や実行状況を監視して必要に応じ停止する仕組み**を組み込むことなどを求めてはどうか。
- また、**許認可等の行政行為など業務の性質によっては、職員等による判断を必須とすることが必要**ではないか。
- **利用目的、組織方針、権限範囲及び法令等に沿って動作することを確認する**とともに、**アライメントを確保するための対策に関する、最新動向の注視**が必要ではないか。

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（4 /10）

II.　AIエージェントの対象拡充に伴う特有の対応事項の強化

②要機密情報の漏洩リスク

【背景・課題】

- AIエージェントが**要機密情報へのアクセス権限**を持ち、**外部等への情報伝達を自律的に実行可能**な場合、誤操作等によって、**要機密情報の漏えい等が生じる**可能性がある。

【検討事項案】

- **要機密情報へのアクセスは、必要最小限とする**ことが必要ではないか。
- **要機密情報へのアクセス権限と、当該情報へのアクセス権限の範囲を超えた情報伝達手段の両方を持つ場合、職員等による判断(HITL)や実行状況を監視して必要に応じ停止する仕組みを組み込む**ことを求めてはどうか。

③AIエージェントのサイバーセキュリティ

【背景・課題】

- 従来型の生成AI においても間接プロンプトインジェクション等のリスクは存在するが、その影響は主としてテキストや画像等の出力に限定。
- 他方で、AIエージェントは、文書・画像等を生成するのみならず、**自律的にタスクを実行することから間接プロンプトインジェクション等による影響がより拡大し、攻撃者の実行手段として悪用**される可能性がある。
- AIエージェントが外部のツールやMCPサーバ等も利用する場合、従来のコードのチェックのみでは、メモリ等の汚染による不適切な動作のリスクがある。

【検討事項案】

- **AIエージェントに固有の脅威やリスク等については、リスクに応じた対策事項を求める**ことが必要ではないか。
- AIエージェントで、**外部由来の情報を取扱う場合には、タスクの実行権限のリスクに応じて、サンドボックス環境での実行など**を求めてはどうか。
- **AIエージェントに特有のサプライチェーン上のリスク**として、ツール等の出所を検証し**来歴の信頼性を確保**することや、メタデータやMCPマニフェスト等を監査対象にするなど**コード以外も含めたセキュリティ対策を求めることが必要**ではないか。
- 特に**不可逆なタスクの実行権限と要機密情報へのアクセス権限があり、かつ、外部由来の情報を取扱うときには高リスクと扱い**、情報の取得、処理、実行、出力の各段階における多層的な対策など**追加的なリスク軽減措置**を求めてはどうか。

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（5/10）

II. AIエージェントの対象拡充に伴う特有の対応事項の強化

AIエージェント間の相互作用への対応

【背景・課題】

- AIエージェントでは、**複数のエージェント間の相互作用を通じて、誤った情報、偏った判断又は不適切な指示が伝播・増幅**され、システム全体として**利用者の意図しない行動や判断に繋がるおそれ**がある。

【検討事項案】

- AIエージェント間でやり取りされる情報や指示が連鎖的な誤動作、権限外操作に繋がらないよう、最小権限の設定を求めるほか、**AIエージェントの動作ログを確認できるようにすることが必要**ではないか。

事後検証及び見直し

【背景・課題】

- AIエージェントは、処理過程や他のAIエージェントも含む**外部システムとの連携が複雑であり、問題が発生しても原因を特定できないおそれ**があるほか、**性能や環境の変化による予期せぬ動作等が生ずるおそれ**がある。
- AIエージェントは、過去の対話履歴等を短期・長期的な記憶として保存・参照した上での利用が想定されるが、**外部からの情報等の影響により、偏りが生じる可能性**がある。

【検討事項案】

- リスクに応じ**AIエージェントが問題のある動作を行った場合に検証できるようログを保存する**とともに、**検証結果に基づき、権限、接続先、エスカレーションの仕組み等を見直す**ことを求めてはどうか。
- AIエージェントが保存した記憶により、**特定の視点や情報への偏りが継続的に生じないよう、メモリの内容を確認し、必要に応じて修正できる仕組み**を求めてはどうか。

人による確認の実効性確保と効率性

【背景・課題】

- **過度な職員等への確認要求は、AIエージェント導入の目的である効率性を損なう**ほか、内容を十分確認せず機械的に承認する**「承認疲れ」が生じ、統制が形骸化するおそれ**がある。
- **利用者がAIエージェントの判断や提案を過度に信頼**し、その妥当性を十分に確認しないまま受け入れるおそれがある。

【検討事項案】

- **職員等への確認プロセスの導入では、効率性と承認疲れを考慮し、過度の確認要求とならないよう**配慮することを求めてはどうか。
- 送金処理、外部への情報送信等の不可逆性の高いタスクの実行の際の職員等による判断（HITL）においては、**職員の承認時間や却下率を確認するなど、自動化バイアスへの対応と、AIエージェントの限界に関する教育実施**を求めてはどうか。

© Digital Agency, Government of Japan

各国文書におけるAgentic AIのリスクの記載

- 各国文書において明示的にAgentic AIのリスクとして触れられている観点を整理

観点	国名：文書名		
	カナダ：Guide on the Use of Agentic AI	シンガポール：Model AI Governance Framework for Agentic AI	Five Eyes：Careful adoption of agentic AI services
Agentic AIのリスク（直接的にリスクに関連する記載より抜粋・編集）	**3. 機会とリスク** • Agentic AIの主なリスク ◦ 出力品質(バイアス、エラー、有害なコンテンツ等) ◦ 権限外の不正な行為 ◦ 不明確な権限設定 ◦ 説明責任 ◦ 追跡可能性 • より自律的なシステム ◦ 信頼できないコンテンツへの過剰な介入 ◦ 無限ループ ◦ 意図しない動作 • その他全般リスク ◦ 適切な設定なしに政府のシステムや機密情報にアクセスし、取得・開示、意図しない行動を取る ◦ 行政上の決定など影響力の高い場面では、公平性への影響、透明性の低下、成果物の生産・使用方法の説明の困難さ等などから、法的リスクの増加、法の不遵守、非倫理行為へ繋がる ◦ 重大なプライバシー侵害 ◦ 職員がAIの推奨結果等をそのまま判断として採用する ◦ 知識やスキルの衰退、人間の判断力の低下を招く	**1.2.2 リスクのタイプ** • 誤った行動（不備のあるコード生成等） • 許可された範囲や権限を越えた行動 • 偏った行動や不公平な行動 • 機密のデータの漏洩や改ざん • 接続されたシステムへの障害 **1.2.3 システミック及びマルチエージェントのリスク** Agentic Systemのスピードと複雑さは上記のリスクを高める可能性がある。相互作用するエージェントの数が増加することにより、これらのリスクを悪化させる • 速度と処理量 • 連鎖的または複合的な影響（ミスが後続の段階へと波及・増幅し被害を拡大する等） マルチエージェントが質的にどの程度異なるリスクをもたらすかは研究中であるが、以下は特に重要なリスク • エージェントの増加：中央管理がないことによりAIエージェントが制御不能に増加する可能性 • 協働の失敗：エージェントの利用にあたってのコミュニケーションや調整ミスによる意図しない失敗の可能性 • 予測不可能性及びその他の創発的行動：複数のエージェントが協力すると、結果の数は指数関数的に増加することになる	**Agentic AIのセキュリティリスク** **権限リスク：**権限の管理に伴うリスク。 • 権限の侵害と範囲の拡大（スコープクリープ） • アイデンティティやエージェントのなりすまし **設計及び構成リスク：**設計・設定不備により過剰な権限や不適切なアクセス制御が生じると、権限悪用や環境間への侵害拡大のリスクが高まる。 **行動リスク：**AIエージェントの誤作動・悪用リスク。 • 目標のずれと意図しない行動 • 欺瞞的な行動（虚偽の情報提供、意図の隠蔽等） • 創発的能力と予測不可能な行動 • 悪意のある行動 **構造的リスク：**連携構造は、高機能化と引き換えに攻撃リスクを増大させる。以下は構造の構成要素例。 • オーケストレーションとリソース • ツール • サードパーティのコンポーネント • データ • 危険のあるエージェント • 通信 **説明責任リスク：**エージェントの高度化により、以下のような要素の原因究明や責任追跡が困難になる。 • 行動とプロセス • 精度 • 可視性

公開情報に基づきデジタル庁にて調査・作成

© Digital Agency, Government of Japan

各国文書におけるリスク観点の記載比較

- 各国文書を参考する前提となる、当該文書において考慮されているAgentic AIのリスクの観点を整理

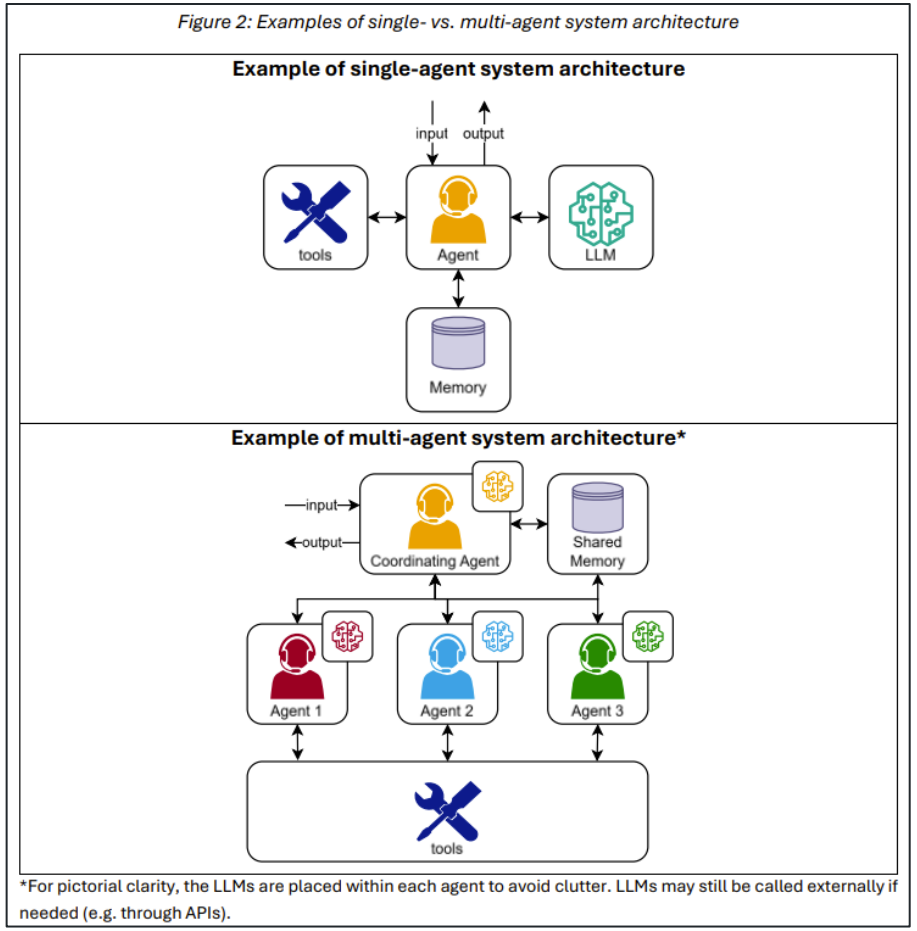
分類	観点	国名：文書名					後続スライドの「観点」との対応関係
		カナダ：Guide on the Use of Agentic AI	オーストラリア：Agentic AI addendum to the AI technical standard for Australian Government	シンガポール：Model AI Governance Framework for Agentic AI	シンガポール：Securing Agentic AI Addendum	Five Eyes：Careful adoption of agentic AI services	
①自律行動の制御に関するリスク	誤った自律行為・判断、意図を超えた挙動	○	○	○	○	○	ログ管理・人間の承認の設計
	偏った・不公平な自律行為	○	○	○	△	△	ログ管理・人間の承認の設計
	意図しない自律行動の伝播、連鎖	○	○	○	○	○	ログ管理・人間の承認の設計
②権限逸脱、情報漏洩リスク	無許可・権限外の行為	○	○	○	○	○	アクセス権限管理
	データ侵害・機微情報の漏えい／改変	○	○	○	○	○	アクセス権限管理
	過大な特権・アクセス権限及びIDなりすまし	○	○	△	○	○	アクセス権限管理
③サイバーセキュリティ、悪用リスク	プロンプト攻撃・悪意ある操作（外部ツール由来等含む）	○	○	○	○	○	セキュリティ対策
AIエージェント等に対応したガバナンス	責任所在・追跡可能性・体制の不備	○	○	△	○	○	ガバナンス体制の構築
	自動化バイアス・技能低下	○	△	○	○	△	ガバナンス体制の構築
	法令・倫理・プライバシーへの不適合	○	○	△	△	△	ガバナンス体制の構築
	停止・廃止管理の不足	△	○	△	△	△	ライフサイクルに応じた管理
	監視の不足・検知や対応の遅延	△	○	△	△	△	ID管理

公開情報に基づきデジタル庁にて調査・作成
○：当該観点をリスクとして明示的に記載　△：関連内容を含むが、記載は限定的　×：明示的な記載を確認できず

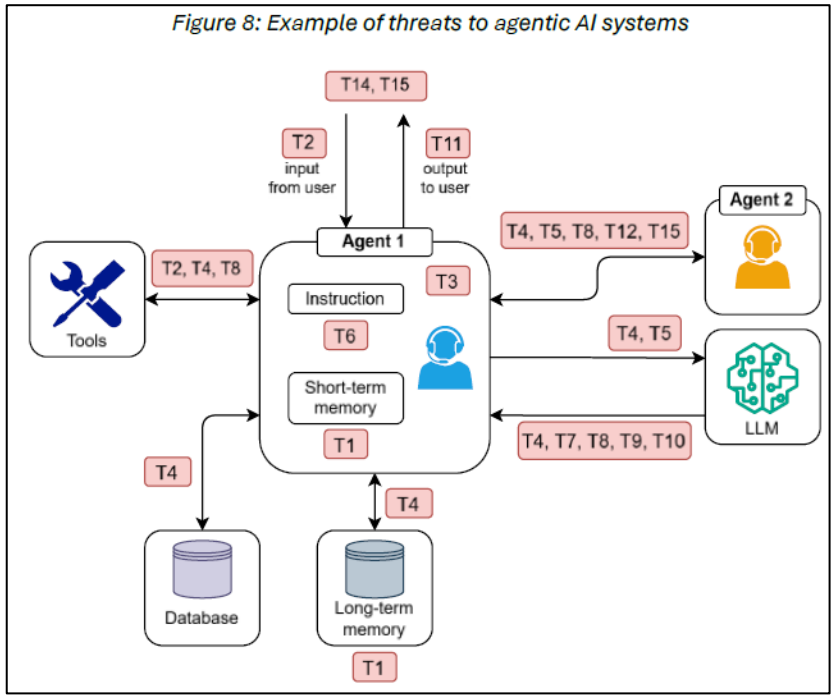
© Digital Agency, Government of Japan

各国文書におけるコンポーネントの図（3/3）

- シンガポール（Securing Agentic AI Addendum）においてはシングルエージェントシステムとマルチエージェントシステムの考え方を整理の上、Agentic AIシステムにおけるセキュリティ脅威例を紹介



Securing Agentic AI Addendum、P.15より引用



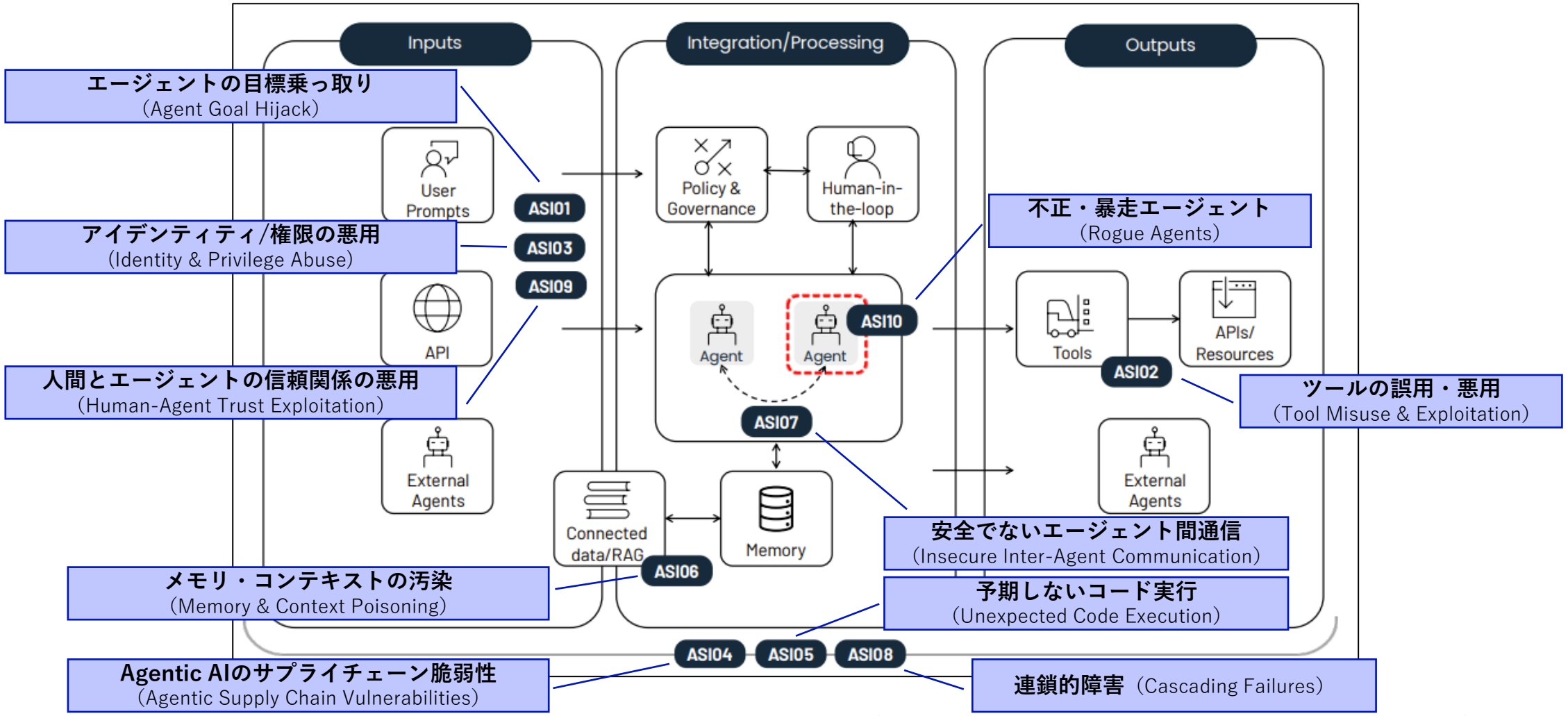
T1 – メモリ汚染
T2 – ツールの不正使用
T3 – 権限侵害
T4 – リソース過負荷
T5 – 連鎖的ハルシネーション攻撃
T6 – 意図の破壊と目標の操作
T7 – 不整合で欺瞞的な振る舞い
T8 – 否認と追跡不可能性
T9 – IDのなりすまし・エージェントIDの侵害
T10 –ヒューマン・イン・ザ・ループの過負荷
T11 – 予期しないリモートコード実行とコード攻撃
T12 – エージェント間通信の汚染
T13 – マルチエージェントシステムにおける不正エージェント
T14 –人間によるマルチエージェントシステムへの攻撃
T15 – 人間の操作

Securing Agentic AI Addendum、P.27より引用。日本語訳は参考訳

© Digital Agency, Government of Japan

OWASPにおけるAgentic AIの脅威の整理（1/2）

- OWASPは、設計・運用上で特に留意すべき代表的な脅威を10個に整理している。それぞれの脅威は、入力、処理、出力の各段階やエージェント間の連携にまたがって存在し、相互に関連しながら影響を拡大する可能性がある



OWASP Top 10 for Agentic Applications for 2026(https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/ ）より抜粋して作成
日本語訳は参考情報としての訳である

© Digital Agency, Government of Japan

OWASPにおけるAgentic AIの脅威の整理（2/2）

- 設計・運用上で特に留意すべき代表的な脅威の概要を整理

ASI ID	脅威名	概要
ASI01	**エージェントの目標乗っ取り** （Agent Goal Hijack）	プロンプトや外部データ等を通じて目標・指示・判断経路が操作され、エージェントが本来と異なる行動を実行する。
ASI02	**ツールの誤用・悪用** （Tool Misuse & Exploitation）	正規のツールが不適切に選択・連携され、権限内での情報流出、データ削除、過剰な処理等につながる。
ASI03	**アイデンティティ・権限の悪用** （Identity & Privilege Abuse）	認証情報、権限継承、委任関係等が悪用され、権限昇格や不正アクセス、無許可の操作が行われる。
ASI04	**Agentic AIのサプライチェーン脆弱性** （Agentic Supply Chain Vulnerabilities）	外部のモデル、ツール、MCP、データ、エージェント等の侵害・改ざんにより、悪意ある処理が実行環境に取り込まれる。
ASI05	**予期しないコード実行** （Unexpected Code Execution）	生成コードやツール連携、危険な関数等を介して、意図しないコードやコマンドが実行され、環境が侵害される。
ASI06	**メモリ・コンテキストの汚染** （Memory & Context Poisoning）	保存・検索される記憶やRAG等に不正確・悪意ある情報が混入し、将来の判断や計画、ツール利用が継続的に歪められる。
ASI07	**安全でないエージェント間通信** （Insecure Inter-Agent Communication）	認証、完全性、機密性等が不十分な通信を通じて、メッセージの盗聴、改ざん、なりすまし、再送が行われる。
ASI08	**連鎖的障害** （Cascading Failures）	誤情報、侵害、ツール障害等の単一の問題が、複数のエージェントや業務フローへ伝播・増幅し、全体的な障害につながる。
ASI09	**人間とエージェントの信頼関係の悪用** （Human-Agent Trust Exploitation）	AIへの過信や権威バイアスが利用され、利用者が誤った提案を承認したり、機密情報を提供したりする。
ASI10	**不正・暴走エージェント** （Rogue Agents）	エージェントが意図した機能や許可範囲から逸脱し、欺瞞、情報流出、業務妨害、自己複製等の有害な行動を継続する。

※ASIとは、“Agentic Security Initiative”を指す

OWASP Top 10 for Agentic Applications for 2026（https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/ ）より抜粋して作成.
日本語訳は参考情報としての訳である

© Digital Agency, Government of Japan

各国文書における各観点毎の記載事項概要

- ①自律行動の制御に関するリスク、②権限逸脱、情報漏えいリスクによるリスク、③サイバーセキュリティ、悪用リスクの観点について、各国文書において各観点に応じた記載がされている

分類	観点	国名：文書名				
		カナダ：Guide on the Use of Agentic AI	オーストラリア：Agentic AI addendum to the AI technical standard for Australian Government	シンガポール：Model AI Governance Framework for Agentic AI	シンガポール：Securing Agentic AI Addendum	Five Eyes：Careful adoption of agentic AI services
①自律行動の制御に関するリスク	ログ管理	エージェントが改変できない基盤にすべての行動を記録し、ツール、行動、承認、入出力、時刻、結果、パラメータ種別を保持。	意思決定を監査可能な記録で説明し、各ライフサイクル局面で実行時のメタデータを収集する。説明・監査・解釈の可能性を高める。	改ざんできない完全な監査証跡を保存し、計画・推論やMCP層の全相互作用を記録する。障害時の実行履歴を削除できないようにする。	行動、入出力、内部状態、時刻等を記録し、要求ID／DIDで分散トレーシングを実施する。改ざん検知可能な、監査ログを求める。	成果物とエージェント間の統合監査ログを要求する。長い推論によるログの肥大化、再現性と有用なシグナル抽出の困難さを考慮。
	人間の承認の設計	**送信・公開・支出・記録更新等の変更**は、低影響かつ容易に取消可能な場合を除き、**人間が確認**する。**不可逆・例外行為は事前確認**を行い、一時停止やアラートを設ける。	**支払・給付等の機微な行為、不可逆・高リスク行為**に人間の承認を必須化。意思決定を行う毎に、評価、エスカレーション、フェイルセーフを設け、履歴文書を作成。	**影響が大きい行為、不可逆な行為、通常と異なる行為**を重要チェックポイントとして人間の承認が必要な対象とする。**人間の承認基盤が機能しない場合は既定で拒否**する。	**データ変更、高リスク取引、不可逆行為に人間の確認・承認を要求**する。**異常検知時は**サーキットブレーカーで**伝播を止め、解除に人間の認可**を要する。	影響の大きい行為を事前承認なしに実行させず、承認要否は設計者・運用者が決定する。**高コスト行為やログ削除等にレビュー・隔離**を設ける。
②権限逸脱、情報漏洩リスク	アクセス権限管理	データ、ツール、行為、スコープを限定し文書化する。デフォルト設定は読取専用とし、必要な工程だけ編集・実行権限を付与する。	各エージェントに必要な情報・ツールだけを認可し、自己への権限付与や過大権限を防止する。ツールにも明確な実行制限と最小アクセスを適用する。	スコープの限定、最小権限、譲渡不可、時間／セッション限定を原則化する。付与者本人の権限を超える委任を禁止し、委任関係を記録することを求める。	人間とエージェント用のポリシーを分離し、変更・未承認の権限委任を禁止する。API、データ、ログ、ツール、ネットワークを最小権限・ホワイトリストで制御する。	権限をタスクに必要な最小限へ細粒度化し、高影響行為にはJIT資格情報を使用する。中央ポリシー決定点で実行時に継続検証し、自己変更・無期限委任を禁止する。
③サイバーセキュリティ、悪用リスク	セキュリティ対策	外部入力を「プロンプト」でなく「データ」として処理し、承認済みソース、入力無害化、外部共有の既定無効化を採用する。敵対的テスト、多層防御、キルスイッチを組み合わせる。	入力・ツール・出力にガードレールを配置し、人間による監督と継続的な評価を重ねる。承認済みインタフェース、堅牢性テスト、ロールバック、キルスイッチを必須としている。	構造的なルールベース、モデル／プロンプト、実行時の多層統制を採用する。信頼済みMCP、サンドボックス、サプライチェーン統制、反復的なシステムテストを求める。	サプライチェーン、IDアクセス、分離、外部ツール、監視等の基盤統制に能力別統制を実装する。セグメンテーション、サンドボックス、サーキットブレーカー（挙動停止機構）、敵対的評価を求める。	多層防御、ゼロトラスト、隔離、セグメンテーション、信頼済みレジストリ／SBOMを重視する。本番前のサンドボックス、レッドチーム、マルチエージェント試験を求める。

公開情報に基づきデジタル庁にて調査・作成

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（6/10）

II. AIエージェントの対象拡充に伴う特有の対応事項の強化

AIエージェント等に対応したガバナンス

【背景・課題】
- 所管部門、最終責任者、承認権限、停止判断、インシデント報告等が曖昧だと、**問題発生時に統制できず、説明責任も果たせなくなる**おそれ。
- 例えば、業務環境でタスクを実行するAIエージェント等の利用状況を把握できなければ、**脆弱性等が判明した場合に、適切な対応が困難**となり、予期しない損害が生じるおそれがある。

【検討事項案】
- AIエージェント等に**一意のIDを割り振り、使用ツール、MCPサーバ等を一覧化**したうえで、登録、承認、配布、公開、停止、廃止といった管理を行えるようポリシーを作成することを対策として求めてはどうか。
- 利用状況や管理責任者を定期的に確認し、**未承認、不要となったAIエージェント等の残存を防止**する運用を求めることを対策事項として追加してはどうか。
- 脆弱性等が判明した場合に、影響を受けるAIエージェント等を**迅速に特定し、停止又は改修できる体制の整備**を対策事項として追加してはどうか。

リスクに応じた対応事項の在り方

【背景・課題】
- AIエージェントの機能や利用方法にかかわらず**一律に対応事項を求めると、**対策が過剰となり、**利活用の便益や業務効率を損なうおそれ**がある。

【検討事項案】
- AIエージェントに関する対応事項を、**①共通事項、②リスク条件に応じた必要事項、③ベストプラクティスに分ける**こととし、整理してはどうか。

AIエージェント等作成者向けガイダンス

【背景・課題】
- AIエージェント等は、個々の利用者による作成を可能とできるが、AIエージェント固有のリスクについて作成段階で留意できるようにする必要がある。

【検討事項案】
- **利用者向けに、AIエージェント等の作成等に関するガイダンス等**を作成してはどうか。

© Digital Agency, Government of Japan

各国文書におけるAIエージェント等に対応したガバナンスの記載事項概要

- AIエージェント等に対応したガバナンスについて、各国文書において各観点に応じた記載がされている

分類	観点	国名：文書名				
		カナダ：Guide on the Use of Agentic AI	オーストラリア：Agentic AI addendum to the AI technical standard for Australian Government	シンガポール：Model AI Governance Framework for Agentic AI	シンガポール：Securing Agentic AI Addendum	Five Eyes：Careful adoption of agentic AI services
AIエージェント等に対応したガバナンス	ガバナンス体制の構築	エージェントごとに人間の所有者・説明責任者を明示し、エスカレーション経路と代理者の文書化が必要。異動・退職時は責任を移管し、不明確な場合は停止・無効化する。	機関に対し、各エージェントの責務と、その意思決定に対する人間の説明責任の割当て・文書化を必須にする。自律的な多段階処理でも人間の監督と責任を維持する。	主要意思決定者、プロダクト責任者、サイバーセキュリティ責任者、利用者にライフサイクル上の責任を配分する。外部事業者を含むバリューチェーンの義務・責任分担も明確化する。	意思決定者、AI実務者、サイバーセキュリティ実務者を統制ごとの責任主体として明示している。エージェント側もオーケストレータ等の役割・権限を定義する必要があると記載。	自律エージェントを管理するガバナンス方針を策定し、法的説明責任とリスクオーナーシップを求める。設計・開発・展開・運用ごとに主な責任主体を示している。
	ライフサイクルに応じた管理	利用前・利用中・利用後で実務事項を整理。展開後の定期監査、変更時のリスク再評価、キルスイッチ、アクセス削除、記録保存、教訓整理を求める。	設計、データ、訓練、評価、統合、監視、廃止をそれぞれコントロールする。コントロールタワー（ガバナンスの取組の１つ）、変更・版管理、ロールバック、安全な廃止を組み込んでいる。	リスク評価、人間の説明責任、技術統制、利用者責任の4つの観点を反復的に適用する。展開前テスト、展開後監視、変更管理、誤作動時の停止・フォールバックを求める。	ライフサイクルアプローチで、エージェントの構成要素と能力別のコントロールを実施している。資産・版管理、展開前テスト、継続監視、ロールバックについては詳細な明示がされている。廃止手続の明示については記載が限定的。	設計・開発・展開・運用の局面別に安全対策を整理。継続評価、権限の段階的な縮小、バージョニング・ロールバックについては詳細な明示がされている。廃止手続の明示については記載は限定的。
	ID管理	個々のエージェントとその行為を一意に識別・管理・追跡できる、明確なエージェントIDの確立を要求する。	各エージェントに固有のアイデンティティを割り当て、役割・責務・アクセス権限と対応付ける。コントロールタワーで稼働中のエージェント等を把握することが求められる。	IDは一意性、帰属先、行為資格の区別を要件とし、中央目録管理で管理すべきとする。IDと権限を中央発行・追跡し、エージェントの無秩序な増殖を防止する。	エージェントを人間ではないIDとして扱い、信頼のできる管理を実施されたレジストリと、検証可能な資格情報を要求している。DID（Decentralized Identifier、分散型識別子）、要求ごとの役割検証、スコープ付きトークンの活用を例示。	各エージェントを独自の鍵・証明書を持つ独立プリンシパルとして構成することを求める。信頼のできる管理を実施されたレジストリと稼働実体を照合し、未登録ID・鍵のアクセスを拒否することを求める。

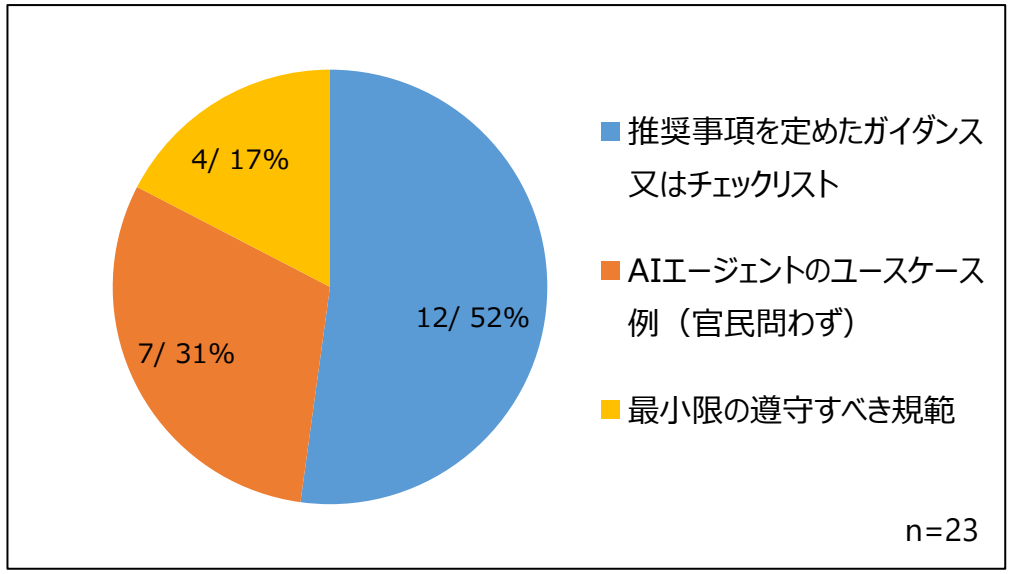
公開情報に基づきデジタル庁にて調査・作成

© Digital Agency, Government of Japan

＃12　高度なタスクを実行できるAIエージェント等の導入に当たっての要望

- AIエージェントの導入に当たっては、**約半数がガイダンス・チェックリスト**、約３割がユースケース例を要望している
- 各府省庁独自のガバナンスとしては、ノーコードでエージェントを作成・運用できるUIの提供や、最新動向に関する勉強会の開催等が実施されている

CAIO_Q20:高度なタスクを実行できるAIエージェント等の導入に当たってはどのようなものが必要か、最も適切なものをお選びください。



CAIO_Q21: Q20に関連して、その他考慮すべき事項や、府省庁独自の具体的なガバナンス体制や管理基盤がございましたらご教示ください。

- 組織内でAIエージェントの利用ルールと具体仕様を現在検討している
- AIエージェントや受託業者による生成AI利用について、ガイドラインへの追記と仕様書・契約書のひな形等を整備してほしい
- 職員がノーコードでAIエージェントを作成・運用可能な環境とあわせて、作成・公開範囲の管理やログ取得等のガバナンス統制機能が必要
- 各府省庁の具体的な活用事例、導入効果、ガバナンス運用事例を共有できる形で整理してほしい

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（7/10）

- ガイドラインの実効性向上や利用者の利便性向上の観点から、知的財産権への対応、調達・運用時の参考情報の提供、導入判断に資する情報の充実等について検討を行うこととしてはどうか。

III. その他全体的な利便性の向上等

各省庁等からの要望①　知的財産権等関連の対策追加

【背景・課題】

- 知的財産権等の利活用時の対策は主に別紙等に一定の記載をしているが、リスクを抑えるための**実効的な対策のバリエーションがまだ少なく、利用者が実際にどのような対策を取れば良いかについてのベストプラクティスは十分ではない**（特に肖像、声の利用時の対策）。

【検討事項案】

- 法務省検討会における議論等を参照し、知的財産権等に係る侵害を防止するためのベストプラクティスの充実を図るべきではないか。
- **民間のユースケースにおける画像や声の類似性等の確認手段**を参考として別途示してはどうか。

各省庁等からの要望②
調達チェックシート及び契約チェックシートの活用方法の検討

【背景・課題】

- 現行ガイドラインにおいて、生成AIシステムの導入形態に応じて、各チェックシートの活用のレベル等、**リスクと対策のバランスの調整のために導入類型例を設けているものの、各導入類型に対する対応の違いがわかりづらい**との意見があった。
- 調達仕様書や契約書等は、**各チェックシートを参考にして作成することとしているが、実際にどのように活用して調達仕様書や契約書等を作成すればよいかがわかりづらい**との声がある。

【検討事項案】

- 各府省庁の実際の仕様書や契約書を収集・分析した上で、各府省庁が調達チェックシートと契約チェックシートを参考にしやすい手法と情報提供を検討してはどうか。

© Digital Agency, Government of Japan

知的財産権等の侵害等対策として考えられる手法（特に画像、声）

- AI生成物（やそれを編集・加工したもの）を利用する場合は、知的財産権等を侵害するものでないかどうかを必ず確認することを、別紙２利活用ルールひな形で記載している
- 知的財産権（特に画像、声）侵害を回避するための確認方法について、確認機能として十分な水準を期待することが困難な技術も多い

検討の必要性	AI生成物（やそれを編集・加工したもの）を利用する場合は、既存の著作物の著作権を侵害するものでないことやその他知的財産権等を侵害するものでないかどうかを必ず確認することを、別紙２利活用ルールひな形で記載している。 一方、各府省庁より、AI生成物の同一性・類似性確認方法を具体化してほしいという要望がある。

同一性・類似性の確認技術の例（画像、音声）

種別	確認技術の例
画像	• 大量に画像データが登録されたDBを用いて、同一判定・類似部分の検知 • Web上の画像との同一判定 • ハッシュを用いた類似度判定 等
音声	• 音声解析にて、音が完全一致の場合の検知 • 声紋での類似度判定 等

© Digital Agency, Government of Japan

利活用時の知的財産権等侵害等対策の記載（1/2）

- AI生成物の利活用における、知的財産権等侵害等対策は、利活用ルールひな形にて記載している。プロンプト入力段階と、AI生成物の利用段階に分け、対策の例を記載している。

プロンプト入力段階

- プロンプトの入力において、知的財産権等の侵害等リスクを低減させるよう対策を取ること（表１参照）。既存の著作物への依拠性がないことを説明できるよう、生成に用いたプロンプト等、AI生成物の生成過程を確認可能な状態にしておくよう努めること。

表 1　プロンプト入力段階の知的財産権等の対策の例

場面	対策例
①著作物の利用	他人の著作権と同一・類似の表現が出力されないよう、他人の特定の著作物と関連付けるようなプロンプトを入力しない、自ら創作して手描きしたラフ画等、自らの著作物を読み込ませた上で出力する 等
②登録意匠・登録商標、他人の商品等表示・商品形態の利用	特定の登録意匠・登録商標等と関連するようなプロンプトを入力しない 等
③人の肖像の利用	特定の人物と関連するようなプロンプトを入力しない、特定の人物の肖像を含むデータ自体を入力しない 等
④人の声の利用	特定の人物と関連するようなプロンプトを入力しない、特定の人物の声を含むデータ自体を入力しない 等

© Digital Agency, Government of Japan

利活用時の知的財産権等侵害等対策の記載（2/2）

AI生成物の利用段階

- AI生成物については著作物性が認められるとは限らないため、著作物として保護が必要な成果物等の作成に生成AIを利用することは慎重に検討すること。
- AI生成物（やそれを編集・加工したもの）を利用する上では、既存の著作物の著作権を侵害するものでないこと（特に、既存の著作物と類似したものとなっていないこと等）やその他知的財産権等を侵害するものでないかどうかを必ず確認すること。その上で、可能な確認措置（インターネット検索等）を行っていることを適切に説明できるようにしておくことが望ましい。AI生成物（やそれを編集・加工したもの）に知的財産権等に係るリスクがある場合には、利用（インターネットでの配信、複製物の譲渡等）を避ける、権利者から許諾を得る、類似しないように作成し直した上で利用すること

表 2　AI生成物の利用段階の知的財産権等の対策の例

場面	対策例
①著作物の利用	• 生成AIを利用しない従来のコンテンツ制作と同様に、AI生成物（やそれを編集・加工したもの）について、他人の著作物と同一・類似でないかどうかを、インターネット検索等を用いて確認する • 他人の著作物と同一・類似の場合には、利用を避ける、権利者から許諾を得る、類似しないように作成しなおした上で利用する　等
②登録意匠・登録商標、他人の商品等表示・商品形態の利用	• 生成AIを利用しない従来のコンテンツ制作と同様に、AI生成物（やそれを編集・加工したもの）について、登録意匠・登録商標等と同一・類似でないかどうかを、インターネット検索等により確認する • 登録意匠・登録商標等と同一・類似と考えられる場合には、利用を避ける、権利者から許諾を得る、類似しないように作成しなおした上で利用する　等
③人の肖像の利用	• 生成AIを利用しない従来のコンテンツ制作と同様に、AI生成物（やそれを編集・加工したもの）としての人の肖像を利用する上では、特定の人物の肖像との同一性（同定可能性）をインターネット検索等により確認した上で、肖像権侵害とならないかを検討する。また、顧客吸引力を有する人物の肖像との同一性がある場合は、パブリシティ権侵害とならないかを検討する • 肖像権・パブリシティ権侵害の可能性がある場合は、利用を避ける、権利者から許諾を得る、同定可能性がないように又は権利侵害がないように作成し直した上で利用する　等
④人の声の利用	• 生成AIを利用しない従来のコンテンツ制作と同様に、AI生成物（やそれを編集・加工したもの）としての人の声を利用する上では、特定の人物の声との同一性（同定可能性）をインターネット検索等により確認した上でパブリシティ権侵害とならないかを検討する • パブリシティ権侵害の可能性がある場合は、利用を避ける、権利者から許諾を得る、同定可能性がないように又は権利侵害がないように作成し直した上で利用する　等

© Digital Agency, Government of Japan

調達仕様書、契約書の作成と利用の流れ

- 生成AIシステムの調達仕様書、契約書等は、デジタル・ガバメント標準ガイドラインを参照の上、調達チェックシート、契約チェックシートを参考に作成を実施する。

予算要求
調達の計画
評価・選定
公告・審査
契約
検収

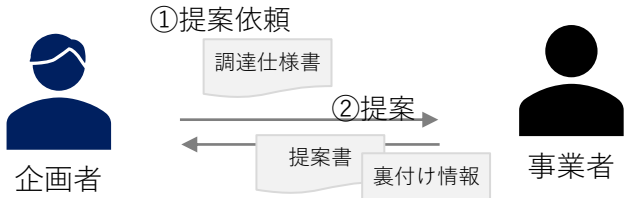
1. 調達仕様書の作成

- 生成AIシステムの調達時に事業者に求める事項は「調達チェックシート」及び「契約チェックシート」を参考に取り込む

政府情報システムの調達時に事業者に求める事項 ＋ 調達仕様書テンプレート
生成AIシステムの調達時に事業者に求める事項 ＋ 調達チェックシート
生成AIシステムを調達する際に盛り込むべき条項 ＋ 契約チェックシート
→ 調達仕様書

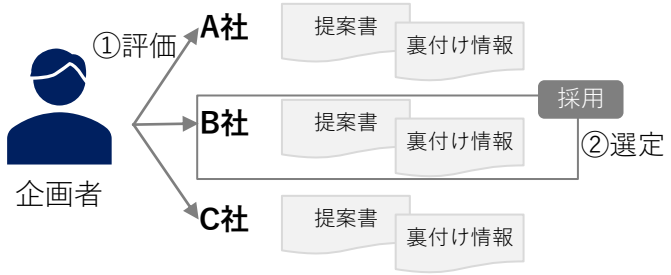
2. 事業者に提案依頼／事業者からの提案・回答

- 企画者は調達仕様書を事業者に送付し、提案依頼を行う
- 事業者は提案を行い、必要に応じて裏付け情報を提出する



3. 評価・選定

- 企画者は各社からの提案書および裏付け情報の内容から事業者を評価・選定する



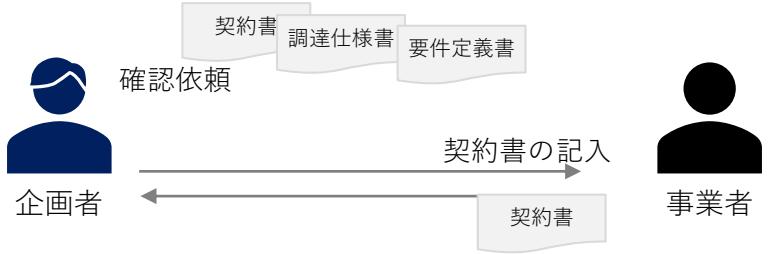
1. 契約書等の準備

- 企画者は契約書等を用意する
- 生成AIシステムの調達時に契約書等の条項として定めるべき事項は本書「契約チェックシート」を参考に取り込む

政府情報システムの調達時に事業者に求める事項 ＋ 契約書のひな形
生成AIシステムの調達時に契約に盛り込む事項 ＋ 契約チェックシート
→ 契約書等

2.事業者との契約

- 企画者は事業者と契約を行う

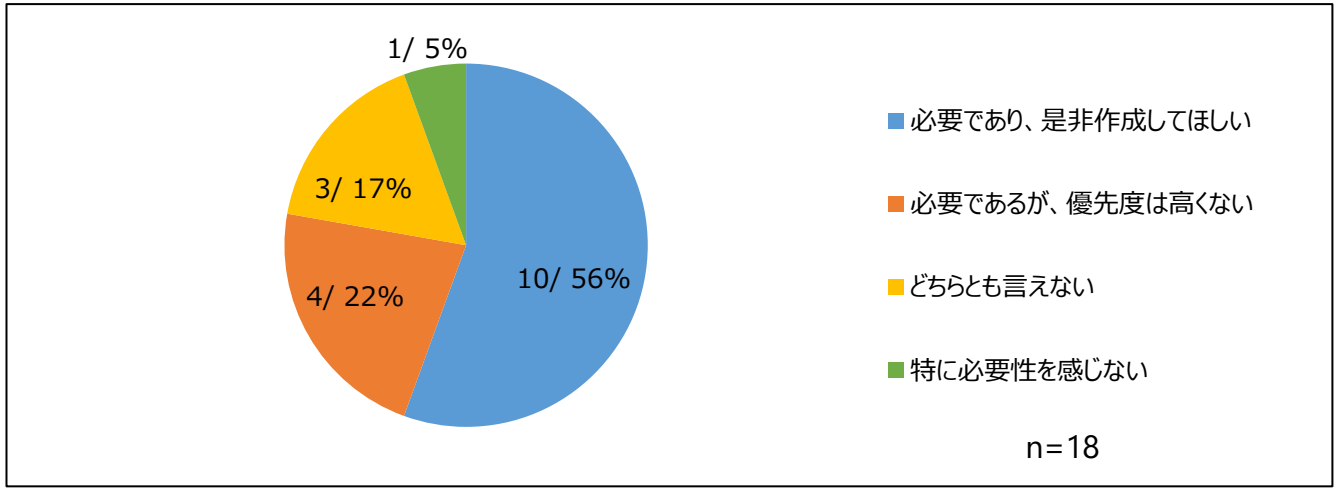


© Digital Agency, Government of Japan

#11　ガイドライン参考資料の要望

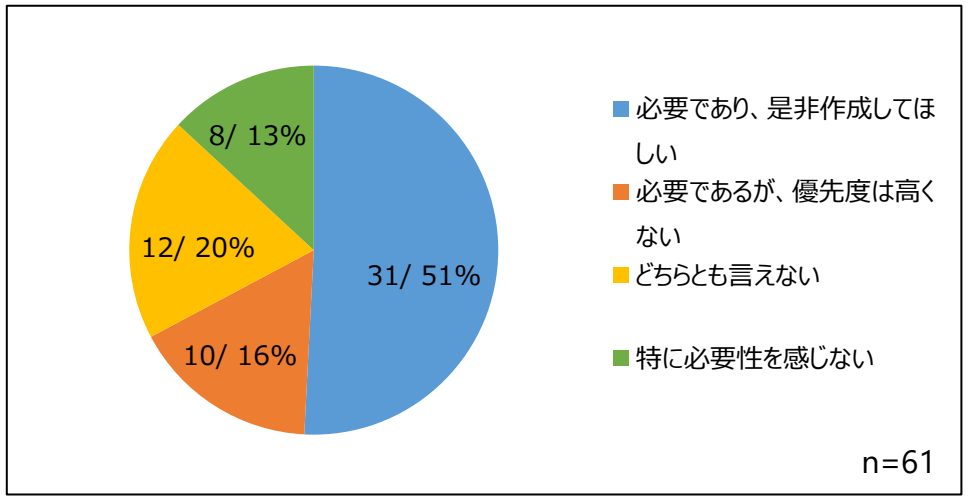
- CAIO・PJMOともに**約７割前後で調達・契約時の具体的な対応例の要望**があった
- 個別回答では、**AIエージェントや受託事業者への対応、契約フェーズ・導入類型に応じた記載例や確認事項の要望**を受けている

CAIO_Q16, 17:今後ガイドラインの参考資料として、調達仕様書や契約書のひな形を作成する場合、その必要性について、選択肢からお選びください。



- AIエージェントや受託業者による生成AI利用について、**ルールと仕様書・契約書の記載例やひな形を整備してほしい。**
- 設計・構築・運用など**契約の各段階で、記載すべき項目が分かるひな形**を作成してほしい。
- 調達仕様書・契約書のひな形は**省内ですでに準備している。**

PJMO_Q15, 16:今後ガイドラインの参考資料として、調達仕様書や契約書のひな形を作成する場合、その必要性について、選択肢からお選びください。



- ひな形があれば**導入検討のハードルが下がり、利活用が進みやすい**と考える
- **AIによる作成支援を前提とした仕様書等のひな形**を、源内の標準アプリとして公開してほしい。

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（8/10）

III. その他全体的な利便性の向上等

各省庁等からの要望③　リスク軸の選定

【背景・課題】

- 現行ガイドラインのリスク軸では、生成AIを使うに当たって一般にリスクが高いと考えられる業務例※を示しているものの、実際の案件の当てはめ時に判断しづらいとの声がある。

- （意見）リスク軸の業務の具体
  現行のリスク軸「A.生成AI利活用業務の性格」では、生成AIによる生成物の瑕疵が重大な影響を及ぼす可能性のある業務例※を挙げているが、該当業務に関する**具体的な例示はなく、判断しづらい**との意見があった。

【検討事項案】

- **Ⅱ①のリスク軸の見直しに際して、**各府省庁においてより**判断がしやすくなるようにする点も考慮すべき**ではないか。

※・国民の基本的権利や安全に大きな影響を及ぼす業務
・機微な政策分野に関する業務
・人間の生命・身体・財産に影響を及ぼす又は法人の事業に重大な影響を及ぼす業務
・資格が求められる業務
・高い説明可能性が求められる業務　等

© Digital Agency, Government of Japan

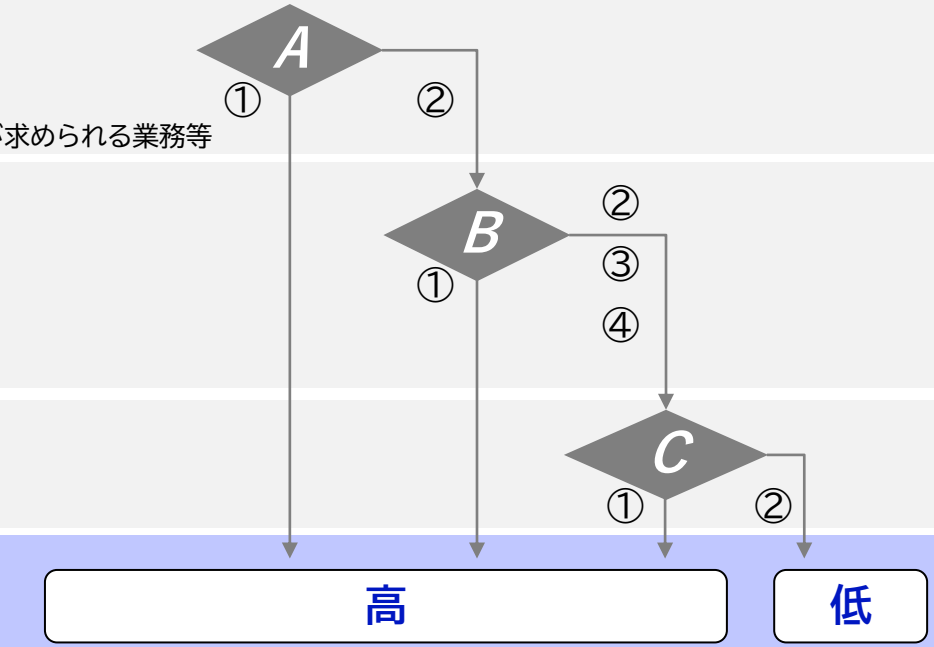
リスク判定ロジック

- ガイドライン2.0版のリスク判定ロジックにおいては、A.適用業務、B.利用範囲、C.職員等による出力結果の判断、の３つの軸を利用。また、ガイドライン本文にて３つのリスク判定の例を記載

A. 適用業務
①生成AIによる生成物の瑕疵※1が重大な影響を及ぼす可能性のある業務※2
②生成AIによる生成物の瑕疵※1が重大な影響を及ぼす可能性のある業務に適用しない
※1生成AIが提供する情報の誤り等
※2国民の基本的権利や安全に大きな影響を及ぼす業務、機微な政策分野に関する業務、
人間の生命・身体・財産に影響を及ぼす又は法人の事業に重大な影響を及ぼす業務、資格が求められる業務、高い説明可能性が求められる業務等

B. 利用範囲
①不特定外部者（一般国民等）
②特定外部者※
③政府職員による府省庁内・複数府省庁横断（共通システムでの利活用等）
④政府職員による府省庁内・単一省庁
※自身の生成AIのリスク知見又は業務上の知見等に基づき出力結果の適切さの判断が可能な者に限る

C. 職員等による出力結果の判断
①生成AIシステムの出力結果の適切さを判断せずに利活用する
②生成AIシステムの出力結果の適切さを判断して利活用する



リスクの定義
高:「高リスク」に該当する可能性が高い
低:「高リスク」に該当する可能性が低い

リスク判定の例
実際には、必ずしも表４のリスク軸を形式的に適用するに止まらず、生成 AIシステムの具体的な利用の在り方（例：C②の前提として、生成内容の出典が容易に参照可能となっている）等も合わせ、総合的にリスクレベルを判断することが重要である。
- （例１）生成 AI による生成物の誤りにより、行政手続の結果に影響を与え、法人の事業活動の停止等の大きな影響を与えるような場合（A①→高リスクの可能性あり）
- （例２）サイトに登録した個人が、自らのニーズにマッチした行政サービス等について、生成 AI を活用した検索要約サービスにより対象を絞った上で、具体的な手続ページに誘導されるもの（A②→B②→C②→低リスク）
- （例３）相手のメール内容に応じた返信メールを作成し、政府職員等による確認を経ずに、自動的に返信するもの（A②→B④→C①→高リスク）

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（9/10）

III. その他全体的な利便性の向上等

利用料及びコストの管理

【背景・課題】

- 生成AIの利用拡大に伴い、トークン数に応じた従量課金等の運用コストが増加する可能性がある。
- 特にAIエージェントでは、**処理の繰り返しや大量文書の読込み等の自律的な動作により、想定を超えるトークン利用量やコスト増加が**発生するおそれがある。

【検討事項案】

- トークン利用量等について、**利用量の把握や上限設定、アラート通知等を講じられる機能**を調達時の要求事項として追加してはどうか。
- また、想定を超える利用量の発生を早期に把握できる体制の整備など、運用上の対策事項を追加してはどうか。

ユースケースの拡大等に応じた記載の充実

【背景・課題】

- 府省庁でのユースケースの拡大に伴いより重要な業務などにも生成AIシステムが活用されるようになりつつある中、業務継続性の確保や業務の性質に応じた利用を確保する必要がある。

【検討事項案】

- 生成AIシステムについて可用性に関する記載や、海外サーバの利用に関する留意事項などに関する記載を充実させてはどうか。

既存規格等のリファレンス

【背景・課題】

- AIマネジメントシステムの国際規格であるISO/IEC42001（及び対応するJIS規格であるJIS Q 42001）等は、国内企業も含めて、認証取得が進みつつある。

【検討事項案】

- 調達時における**外部認証（主にISO/IEC42001）との対応関係を整理し、調達担当者が要求事項の充足状況を簡便に確認できるようにしてはどうか。**
- 相互運用性を確保するため、AIエージェントを作成できる環境においては、MCPやA2Aなどの広く採用が進む標準プロトコルへの準拠を求めることとしてはどうか。

© Digital Agency, Government of Japan

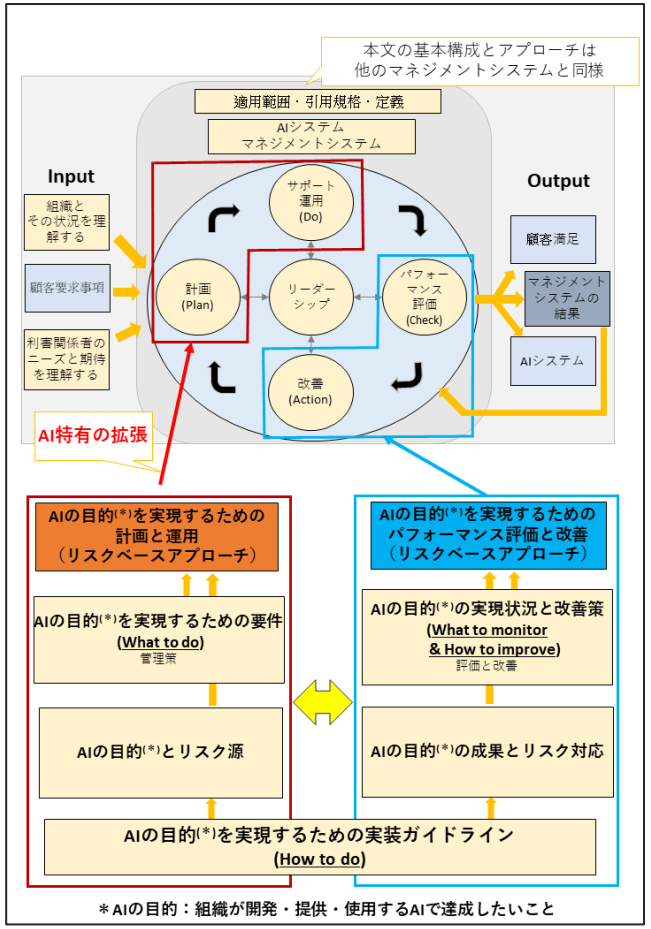
AIマネジメントシステム（ISO/IEC 42001）

- 本規格は、AIシステムを開発・提供・利用する組織向けに、AIの適切な利活用を支えるマネジメントシステムの要求事項を定めたもの。リスクベースで信頼性、透明性、説明責任、公平性、プライバシーへの配慮を求め、学習データや機械学習にも対応。ISO9001やISO/IEC27001と同様の構成を採用し、既存の管理体系と統合しやすい設計となっている

規格の概要

本規格は、AIシステムを開発、提供または使用する組織を対象とし、組織がAIシステムを適切に利活用（開発・提供・使用）するために必要なマネジメントシステムを構築する際に遵守すべき要求事項について、リスクベースアプローチによって規定したものです。信頼性や透明性、説明責任を備えたAIシステムの利活用ができるよう、そのリスクを特定し、軽減すると共に、AIの公平性や個人のプライバシーなどへの配慮についても要求しています。さらに、AIシステムに特有な学習データや機械学習について考慮するにあたっても、重要な規格になります。

マネジメントシステムの構築については、ISO9001品質マネジメントシステム（QMS）規格やISO/IEC27001情報セキュリティマネジメントシステム（ISMS）規格など既存のマネジメントシステム規格と同様のアプローチを採用し、同じ構成で要求事項を規定するなど、本規格の利用者を考慮した規格にもなっています。



経済産業省「AIマネジメントシステムの国際規格が発行されました」（https://www.meti.go.jp/press/2023/01/20240115001/20240115001.html）より抜粋

© Digital Agency, Government of Japan

AIマネジメントシステム（ISO/IEC 42001）

- この規格は、組織の状況の下でAI（人工知能）マネジメントシステムを確立し、実施し、維持し、継続的に改善するための要求事項について規定し、手引を提供するものとされている。AIシステムを活用した製品又はサービスを提供又は利用する組織によって利用されることを意図している。組織がその目的を追求する上で責任をもって、AIシステムを開発、提供又は利用し、適用される要求事項、利害関係者に関連する義務、及び利害関係者からの期待を満たすために役立てることを意図されている。

目次（抜粋）			
4	組織の状況	4.1	組織及びその状況の理解
		4.2	利害関係者のニーズ及び期待の理解
		4.3	AI マネジメントシステムの適用範囲の決定
		4.4	AI マネジメントシステム
5	リーダーシップ	5.1	リーダーシップ及びコミットメント
		5.2	AI 方針
		5.3	役割，責任及び権限
6	計画策定	6.1	リスク及び機会への取組
		6.2	AI の目的及びそれを達成するための計画策定
		6.3	変更の計画策定
		6.4	資源
7	支援	7.1	力量
		7.2	認識
		7.3	コミュニケーション
		7.4	文書化した情報

目次（抜粋）			
8	運用	8.1	運用の計画策定及び管理
		8.2	AI リスクアセスメント
		8.3	AI リスク対応
		8.4	AI システム インパクトアセスメント
9	パフォーマンス評価	9.1	監視，測定，分析及び評価
		9.2	内部監査
		9.3	マネジメントレビュー
10	改善	10.1	継続的改善
		10.2	不適合及び是正処置

その他、以下の目次項目がある
序文
1. 適用範囲
2. 引用規格
3. 用語及び定義
- 附属書 A（規定）参考となる管理目的及び管理策
- 附属書 B（規定）AI 管理策実施の手引き
- 附属書 C（参考）AI 関連の潜在的な組織の目的及びリスク源
- 附属書 D（参考）複数の領域又は分野にわたる AI マネジメントシステムの利用
- 参考文献
- 解説

日本規格協会グループ「JIS Q 42001:2025　情報技術－人工知能－マネジメントシステム　Information technology-Artificial intelligence-Management system」（https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+Q+42001%3A2025）より抜粋して作成

© Digital Agency, Government of Japan

政府AI調達・利活用ガイドラインの課題と充実に向けた検討事項候補（10/10）

- 一般向け・民間向け及び政府向けの関連ガイドライン等の整備状況を踏まえ、必要に応じて本ガイドラインの記載事項の充実を検討することとしてはどうか。

Ⅳ. 他ガイドライン等の参照による必要事項の拡充

関連ガイドライン等

【背景・課題】

●一般向け・民間向け及び政府向けの関連ガイドライン等において、生成AIに係るルール整備、改定がされ、今後も検討が見込まれる。

【検討事項案】

●知的財産権関連では、「肖像、声等の無断利用による民事責任の在り方に関する検討会取りまとめ報告書―生成AIによるパブリシティ権侵害等に関する解釈指針―」や「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」、「AIセーフティに関する評価観点ガイド」等の内容を踏まえる必要があるかを検討してはどうか。

●「AI事業者ガイドライン」などの今後の見直しがある場合、必要に応じてエージェントに関する記載の取り込みを検討してはどうか。

●他の関連ガイドライン等の策定、更新状況等を踏まえ、必要に応じて記載の取り込みを検討してはどうか。

© Digital Agency, Government of Japan

デジタル庁
Digital Agency

© Digital Agency, Government of Japan