

総務省 第32回 AI ガバナンス検討会資料

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今後の検討について

2026/10

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。

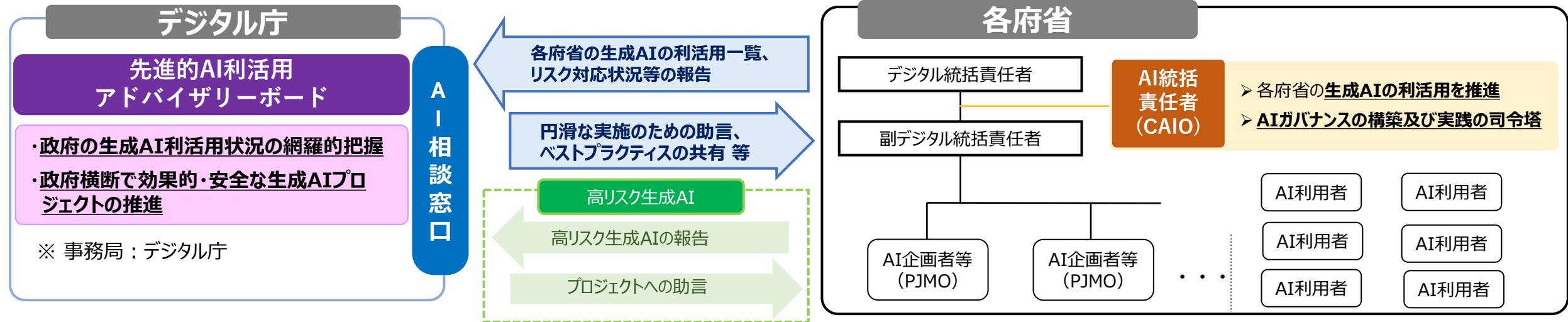
対象：原則として、**入力テキスト及び音声、出力はテキスト、画像又は音声可能な生成AIシステム** ※特定秘密や安全保障等の機微情報を扱うシステムは対象外

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

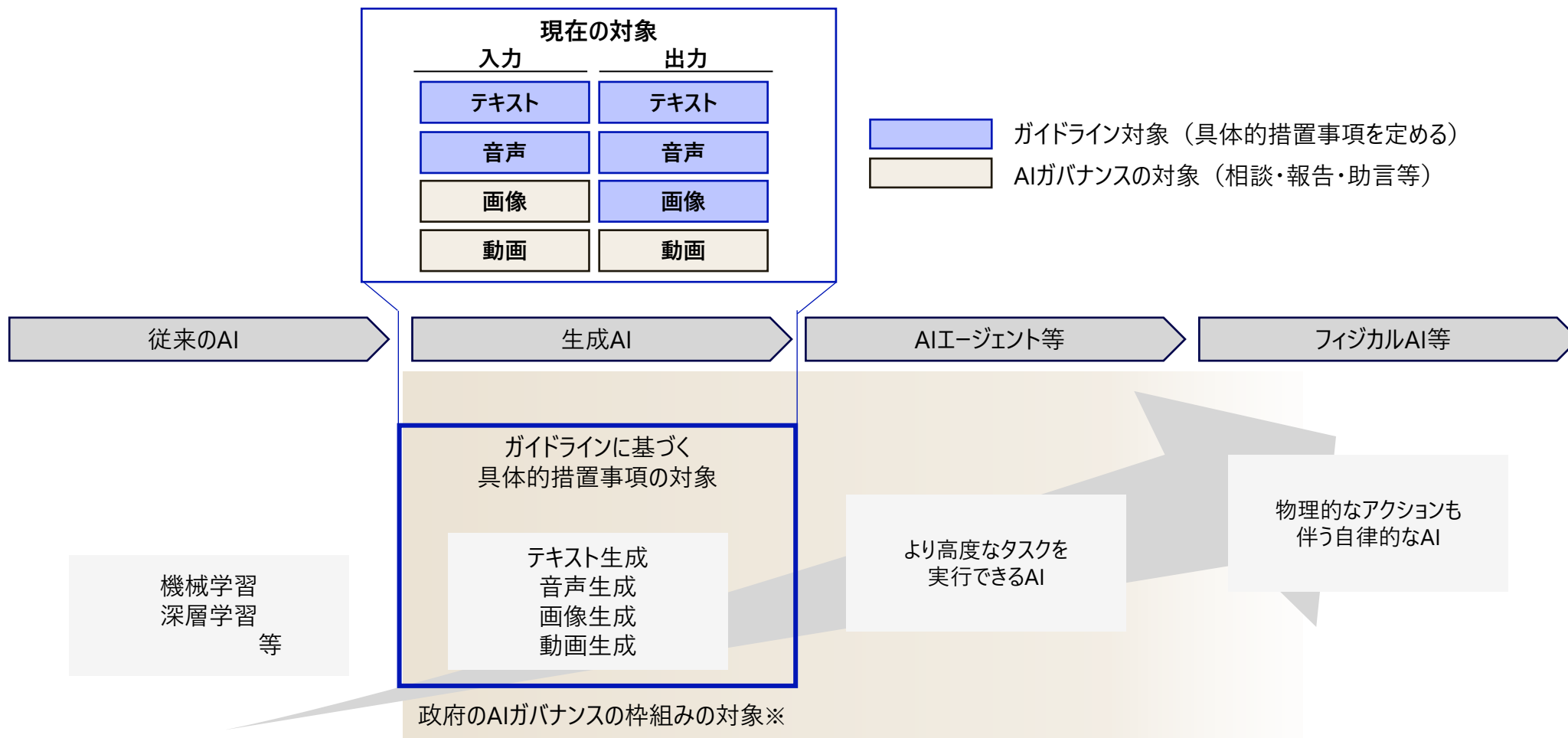
- AI統括責任者（CAIO）は、各府省の利用者（職員）に向けて**生成AIの利用ルール**を策定。
- 企画者・提供者は、本ガイドラインの「**調達チェックシート**」及び「**契約チェックシート**」を参考にして仕様書作成や事業者との契約等を行うことにより安全かつ品質の高い生成AIシステムの調達を確保。運用開始後も適切な利用や安全性や品質の確保を定期的に検証。
- 提供者及び利用者は**リスクケースが生じた場合、適切に各府省AI統括責任者（CAIO）に報告し、提供者が必要な対応を実施**。先進的AI利活用アドバイザリーボードは各ケースの報告を受け、必要に応じ再発防止策等を検討。

ガイドラインの主要改定事項

論点		課題	主な改定事項
I	政府内外の利用実態等を踏まえたガイドラインの拡充	AI技術等に着目したガイドラインの対象拡大	政府における利活用状況を踏まえ、 <u>入力テキスト又は音声、出力はテキスト、音声又は画像が可能な生成AIへ対象拡大</u> <u>生成AIシステムのうち、動画等を生成するもの、AIエージェント等は、報告・助言等のAIガバナンス体制の対象</u> （職員個人で作成可能なAIエージェントも対象）
II	生成AIのリスクに対応する記載の見直し・充実	高リスク判定基準の見直し	ユースケースの拡大の実態を踏まえ、適切にリスク判定されるよう見直し <u>要機密情報を扱うリスクについて、生成AIシステムの権限管理を適切に実施するアクセス制御に加え、生成AIシステム特有のリスク対策を追加した上で、判定基準から削除</u>
III	チェックリストのupdateやベストプラクティスの充実	調達チェックシートや契約チェックシートの記載充実	調達チェックシートは、対策例等や、ツールやベンチマーク等の具体例を追加、契約チェックシートは、成果物に関する契約上の取り決め方や定義方法等の記載を充実
IV	現行ガイドラインで記述が不足している他の制度等の記述の追加	知的財産権等対策の対策例・対策例の具体例の明示	画像や音声出力への対象拡大を踏まえ、 <u>知的財産権等対策に係る記載を入力・学習等への留意事項に追加するほか、参考資料として「知的財産権等対策参考シート」を作成</u>

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活用アドバイザーボード」への報告、デジタル庁のAI相談窓口、「先進的AI活用アドバイザーボード」の助言等の対象

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

- 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

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

- 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な規律や、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。

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

- 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いるかどうか。
- 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。

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

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いるかどうか。

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いるかどうか。
- 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いるかどうか。

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

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いるかどうか。
- 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いるかどうか。

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いるかどうか。
- 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いるかどうか。

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

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いるかどうか。
- 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。

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

III. その他全体的な利便性の向上等

各省庁等からの要望 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】

- II ①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

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

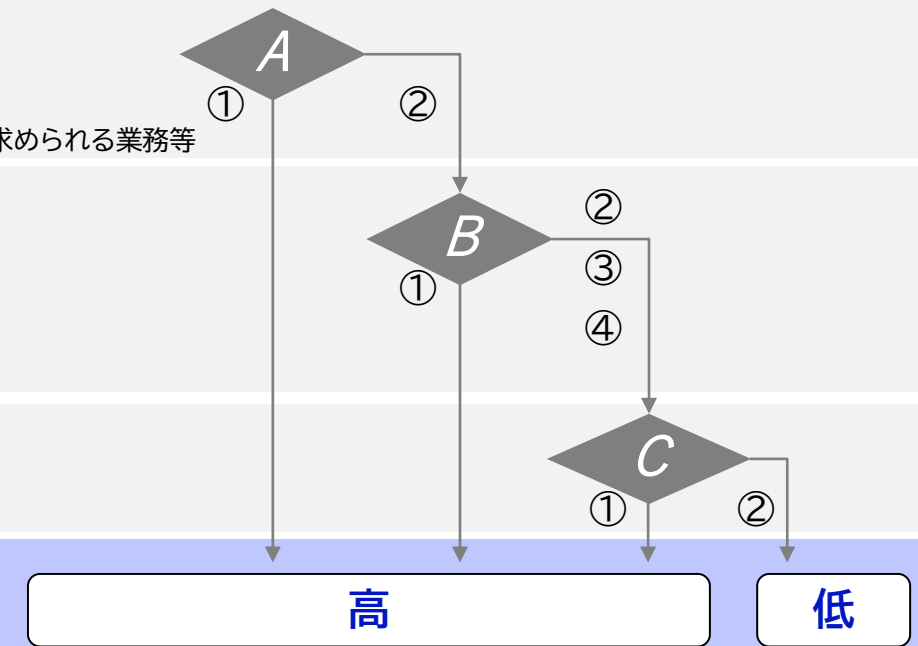
リスクの定義

- 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）



(参考) リスク判定ロジックの改定

- 各府省庁の生成AIユースケースの拡大の実態を踏まえ、リスク低減が適切に図られると考えられる場合や、ガイドライン対象の生成AI拡大によりリスクが高まる場合については、適切にリスク判定されるよう見直し
- 要機密情報又は個人情報の取り扱う場合を高リスクと判定する「C対象データ」の項目について、情報の格付に応じたアクセス制御を行うことに加え、要機密情報が入力・学習等される場合を想定した生成AIシステム特有のリスク対策を必須の実施事項と明記した上で、削除する

※リスク判定ロジックはあくまで参考ツールであり、個々の生成AIシステムのリスクレベルの最終的判断は各府省庁CAIOが行う

項目	改定の概要
C.対象データ	<p>以下の<u>必須事項を明記したうえで、要機密情報又は個人情報を取り扱う場合を高リスクと判定する項目を外す【必須の実施事項とするもの】</u></p> <ul style="list-style-type: none">・ 取り扱う情報の格付等に従い、権限を有する者のみがアクセス制御の設定等を行える機能を設ける・ 入力された要機密情報を参照等する場合に入力者以外にも漏洩するリスクへの対処をすること（特に国民等外部の者が利用する場合で、個人情報等が入力される恐れがある場合も含む）・ 要機密情報を学習、入力する際には出力結果の機密性に留意し、業務利用に問題ないか利用者が判断する
A.利用範囲	<ul style="list-style-type: none">・府省庁外の者による利用であっても、生成AIのリスクに関する知見や業務上の知見に基づき出力結果の適切さを判断できる者に限定されるのであれば、府省庁職員が利用する場合と同じリスク評価とする。
D. 出力結果の判断	<ul style="list-style-type: none">・AIエージェント等も対象とする場合、職員等が出力結果の適切さの判断を行わない場合の影響を、現時点で一般化することは困難であることから、高リスクとする。

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

- 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】

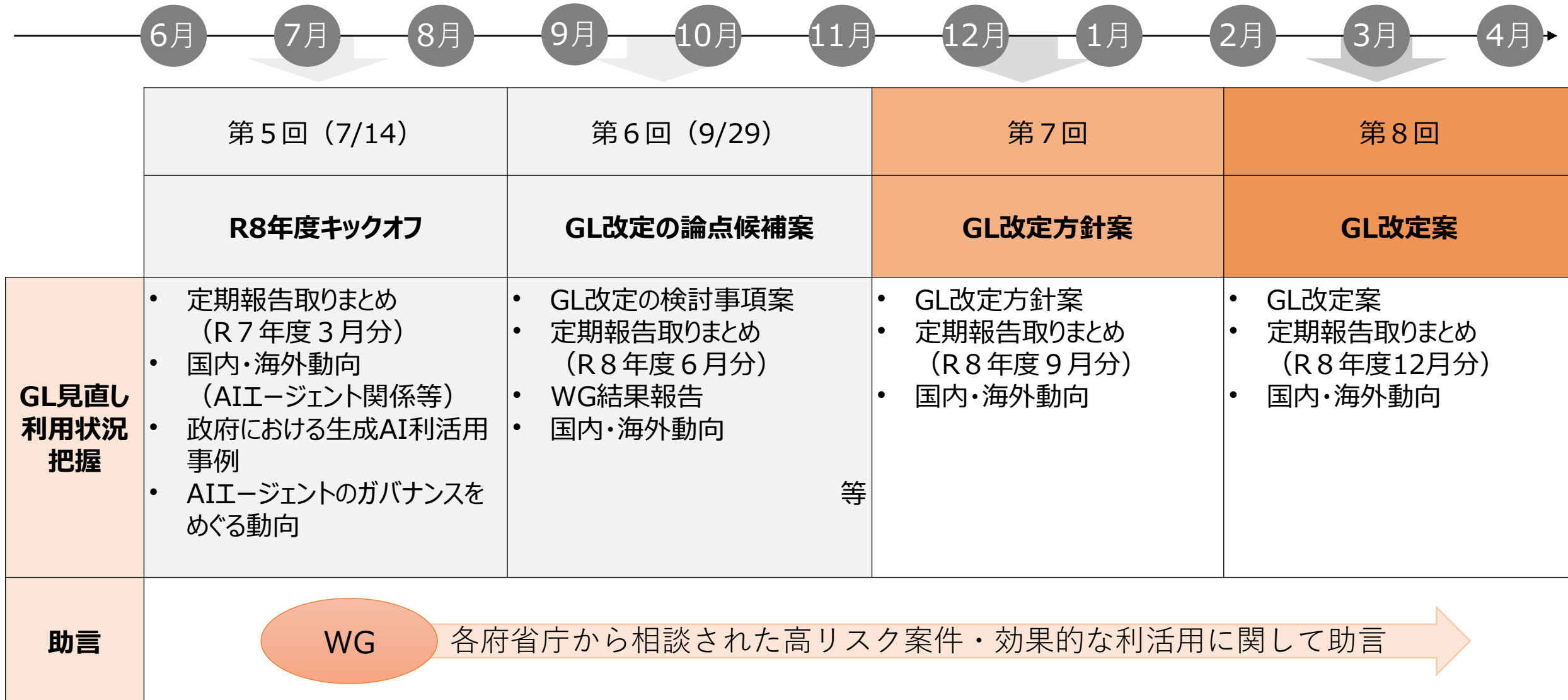
- 一般向け・民間向け及び政府向けの関連ガイドライン等において、生成AIに係るルール整備、改定がされ、今後も検討が見込まれる。

【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ー生成AIによるパブリシティ権侵害等に関する解釈指針ー」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。

アドバイザーボードR8年度開催スケジュール(案)

- 次期ガイドライン改定の論点等の検討し、年度内に改定案のとりまとめを目指す



デジタル庁

Digital Agency