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改定案

2026/03/18　　省庁業務サービスグループ　政府のAI調達・利活用ルール担当

デジタル庁
Digital Agency

Agenda

1. ガイドラインの充実に向けた論点と改定案内容
   (1) 政府AI調達・利活用ガイドラインの課題と充実に向けた論点候補
   (2) 論点に対するガイドライン改定案内容
   (3) 留意事項に対するガイドライン改定案内容
2. ガイドラインの改定具体的案
   (1) ガイドライン改定の全体像
   (2) ガイドラインの対象生成AI拡大の考え方
   (3) ガイドライン本紙の改定案内容
   (4) 利活用ルールひな形の改定案内容
   (5) 調達チェックシートの改定案内容
   (6) 契約チェックシートの改定案内容

1　ガイドラインの充実に向けた論点及び改定案内容

(1)　政府AI調達・利活用ガイドラインの課題と充実に向けた論点候補

(2)　論点に対するガイドライン改定方針案

(3)　留意事項に対するガイドライン改定方針

政府AI調達・利活用ガイドラインの課題と充実に向けた論点候補

ガイドライン充実の必要性について

- 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。
- 具体的には、政府において検証中あるいは今後実装予定の生成AIシステム等の動向や、関連する法令・ガイドラインの見直し状況等も踏まえ、充実させる必要。
- 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。

論点候補を以下4テーマと留意事項に整理

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

Ⅱ　生成AIのリスクに対応する記載の見直し・充実

Ⅲ　チェックリストのupdateやベストプラクティスの充実

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

Ⅴ　政府AI・調達利活用ガイドライン改定作業に当たっての留意事項

論点に対するガイドライン改定案（1/2）

論点		課題		改定案
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	①	生成AIのユースケースに応じた記載	国民向けの生成AIサービス提供に向けて、国民向け生成AIシステム利用規約作成時の留意事項を作成
		②	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、入力はテキスト又は音声、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 生成AIシステムのうち、画像や動画等を入力するもの、動画等を生成するもの、AIエージェント等は、具体的対応事項は定めないが、報告、助言等のAIガバナンス体制の対象となるよう明記
Ⅱ	生成AIのリスクに対応する記載の見直し・充実	①	高リスク判定基準の見直し	各府省庁における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、ガイドラインの対象とする生成AIの拡大によりリスクが高まる場合について、適切にリスク判定されるよう修正
		②	リスク軽減措置の具体化の検討	調達チェックシートに、対策例や対策例詳細、裏付けとなる情報の例を追加 国民向けの生成AIサービス提供に向けて、国民向け生成AIシステム利用規約作成時の留意事項を作成（再掲）
		③	リスクケースの整理	対象とする生成AIの拡大等を踏まえ、政府の生成AI利用で想定されるリスクケースの記載を追加
		④	生成AIモデルのバイアスの整理	利用目的に応じ、バイアスや出力制限が支障となるかを踏まえ、適切な生成AIを選択するべき旨を記載

論点に対するガイドライン改定案（2/2）

論点		課題		改定案
Ⅲ	チェックリストのupdateやベストプラクティスの充実	①	各役割の対応事項についての整理充実	要機密情報を扱うリスクに関して、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、要機密情報の入力・学習等への留意事項を、本文、利活用ルールひな形、調達チェックシートに追加。 デジタル社会推進標準ガイドラインの記載を踏まえ、システム監査に係る記載を追加
		②	調達チェックシートや契約チェックシートの記載の充実	調達チェックシートは、対策例や、ツールやベンチマーク等の具体例等を追加、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
		③	民間も含めたベストプラクティスの事例紹介	ベストプラクティスとなる生成AIの活用事例は、参考資料として作成
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	セキュリティ関係	ISMAP制度とクラウドサービス上で実装される生成AIの見解	ISMAPポータルサイトに公表された生成AI開発基盤と個々の生成AIモデルに関する内容を踏まえ、本文の記載を更新
			総務省AIセキュリティ分科会によるガイドラインを踏まえた対応	生成AIのセキュリティ上留意すべき点を役割ごとに追記するとともに、具体的対策の事例を調達チェックシートに追加
		知的財産関係	知的財産権等対策の対策例・対策例の具体例の明示	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
		既存のAIシステムガイドラインとのレファレンス	AI セーフティに関する評価観点ガイド(第1.10版)を踏まえた対応	評価観点ガイドの更新内容を踏まえて、主に調達チェックシートの記載を更新
			AIマネジメントシステム国際規格ISO/IEC 42001等との対応関係	ISO/IEC 42001は、活用がより進んだ段階で反映を検討

留意事項に対するガイドライン改定案

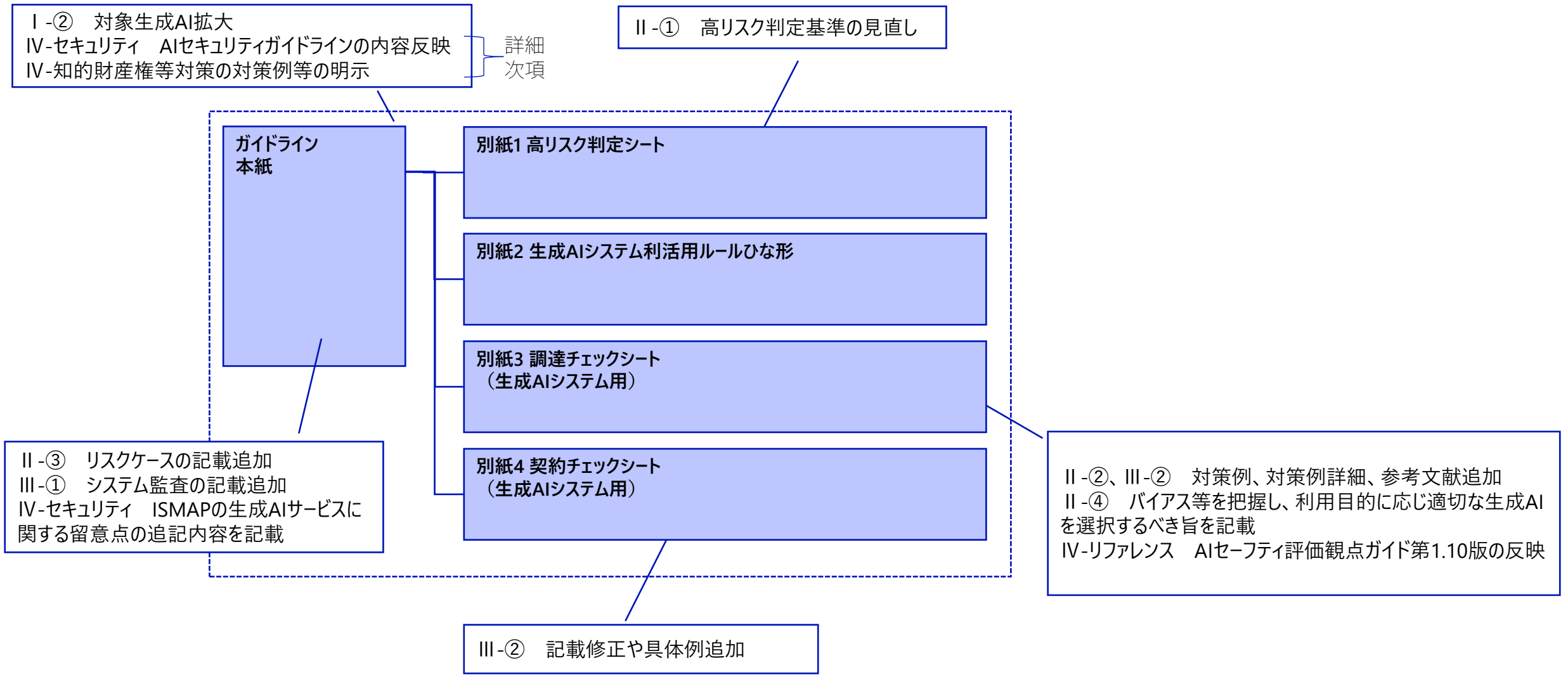
留意事項	項目		改定案
留意事項① ガイドラインの構造	(1)	本文は基本的な事項を総則的に記載	本文は基本的な事項を総則的に記載
	(2)	別紙には具体的事項を記載	別紙には技術の発展やユースケースを踏まえた具体的な事項を記載
	(3)	運用の参考になる情報のとりまとめ	ベストプラクティスとなる生成AIの活用事例は、参考資料として作成（再掲）
	(4)	タイムリーに行う情報提供手段を整備	タイムリーに情報提供を行うため、CAIOに対する注意喚起について記載
留意事項② 他ガイドライン等との関係	(1)	一般向け又は民間を想定したガイドラインは必要な範囲で取り込み	総務省AIセキュリティ分科会による「「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）」（再掲）のほか、AISI「AIインシデントレスポンス・アプローチブック」、「CAIOガイドブック、CAIO設置・AIガバナンス実務マニュアル（案）」の必要となる内容を取り込み AI事業者ガイドライン等の関連ガイドラインの新規内容や更新内容を精査して取り込み（予定）
	(2)	政府向けのシステムに関するガイドラインは生成AIの調達・利活用の観点で、特に記載すべき内容があれば調達・利活用ガイドラインに記載	「政府機関等のサイバーセキュリティ対策のための統一基準」を踏まえ、取り扱う情報の格付等に基づくアクセス制御等の必要性を追加 「DS-680.2 ウェブコンテンツガイドライン」を踏まえ、ウェブサイト等による行政情報及び機能提供を行う際に生成AIに関する部分を参考とすることを記載

2　ガイドラインの具体的改定案内容

		主に紐づく論点
(1)	ガイドライン改定の全体像	–
(2)	ガイドラインの対象生成AI拡大の考え方	Ⅰ-②、留意事項②
(3)	ガイドライン本紙の改定案内容	Ⅰ-②、Ⅱ-①、Ⅱ-③、Ⅲ-①、Ⅳ-セキュリティ、Ⅳ-知財等対策、留意事項①、留意事項②
(4)	利活用ルールひな形の改定案内容	Ⅳ-知財等対策
(5)	調達チェックシートの改定案内容	Ⅱ-②、Ⅱ-④、Ⅲ-①、Ⅲ-②、Ⅳ-セキュリティ、Ⅳ-知財等対策、Ⅳ-リファレンス、留意事項②
(6)	契約チェックシートの改定案内容	Ⅲ-②、留意事項②
(7)	知的財産権等対策参考シートの概要（政府内限り参考資料）	Ⅳ-知財等対策
(8)	国民向け生成AIシステム利用規約作成時の留意事項概要（政府内限り参考資料）	Ⅰ-①、Ⅱ-②
(9)	その他、軽微な修正（表現の修正等）	–

ガイドライン改定の全体像（１/２）

各改定方針に基づき改定・新規策定・策定予定の資料は以下である（各No.は、「論点に対するガイドライン充実方針案」のNo.）



ガイドライン改定の全体像（２/２）

政府全体でのAIに係る検討を取り込むため、セキュリティ関連及び知的財産権等対策に係る記載拡充を実施

セキュリティ関連：スコープとする生成AIの違いを踏まえつつ、政府におけるAI調達・利活用の観点で有益な情報を追加

知的財産権等：利用者の利活用時の対策は利活用ルールひな形へ、調達段階での確認事項や、開発・追加学習を行う際の事項は調達チェックシートへ、提供時の事項は本文6.5へ

総務省
「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）」

経産省
「コンテンツ制作のための生成AI利活用ガイドブック」

文化庁
「AIと著作権に関するチェックリスト＆ガイダンス」

ガイドライン本紙

別紙1
高リスク判定シート

別紙2
生成AIシステム利活用ルールひな形

別紙3
調達チェックシート（生成AIシステム用）

別紙4
契約チェックシート（生成AIシステム用）

セキュリティ関連

- LLM等に対する攻撃の概観や対策例

- 生成AIシステム全体に係る対策例
- 生成AIシステムの出力に係る対策例
- 外部データ参照に係る対策例
- 生成AIモデルに対するプロンプト入力に係る対策例

知的財産権等対策

- 提供者の対応事項

- プロンプト入力段階の対策例
- AI生成物の利用段階の対策例

- 生成AIシステムで知的財産等を取り扱う場合の要求事項
- 企画・開発において生成AIに学習を
- 行わせる場合の対策例

ガイドライン本紙の改定内容案の全体像（1~5章）

目次（抜粋）	改定内容又は主に紐づく論点・留意事項
1　はじめに	
1.1　背景	• 「人工知能基本計画」や「人工知能関連技術の研究開発及び活用の適正性確保に関する指針」の決定について追加
1.3　用語	• 表1における「生成AIシステム」の定義を修正、「生成AIモデル」の定義追加
2　本ガイドラインの目的及び適用対象	
2.1　本ガイドラインの目的	• 「人工知能関連技術の研究開発及び活用の適正性確保に関する指針」を踏まえた記載追加
2.2.2　本ガイドラインが対象とする生成 AI	• 「Ⅰ-②対象生成AI拡大」に伴う記載修正 • 1.3 用語の定義修正を踏まえ、本ガイドラインにおける「生成AIシステム」「生成AIモデル」の用語使用方法を追記
2.2.4　本ガイドラインの適用開始時期等について	• 「2.2.4 本ガイドラインの適用開始時期等について」を附則へ
3　政府における生成 AI の利活用方針	
3.2　高リスクな生成AI利活用の考え方	• 「Ⅱ-①高リスク判定基準の見直し」に伴う記載修正
4　AI の利活用促進と AI ガバナンスの強化及び推進のための体制構築	
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	• CAIOに対する注意喚起について記載（留意事項①−(４)）
5　生成 AI による便益とリスクを理解した利活用推進	
5.1　生成AIの便益	• 表5 生成AI利活用による期待される便益の例を削除（留意事項①−(１)）
5.2　生成AIによるリスク	• 「Ⅰ-②対象生成AI拡大」を踏まえ、事例追加 • 表6 AI事業者ガイドラインにおけるリスクの例を削除（留意事項①−(１)）

ガイドライン本紙の改定内容案の全体像（6章）

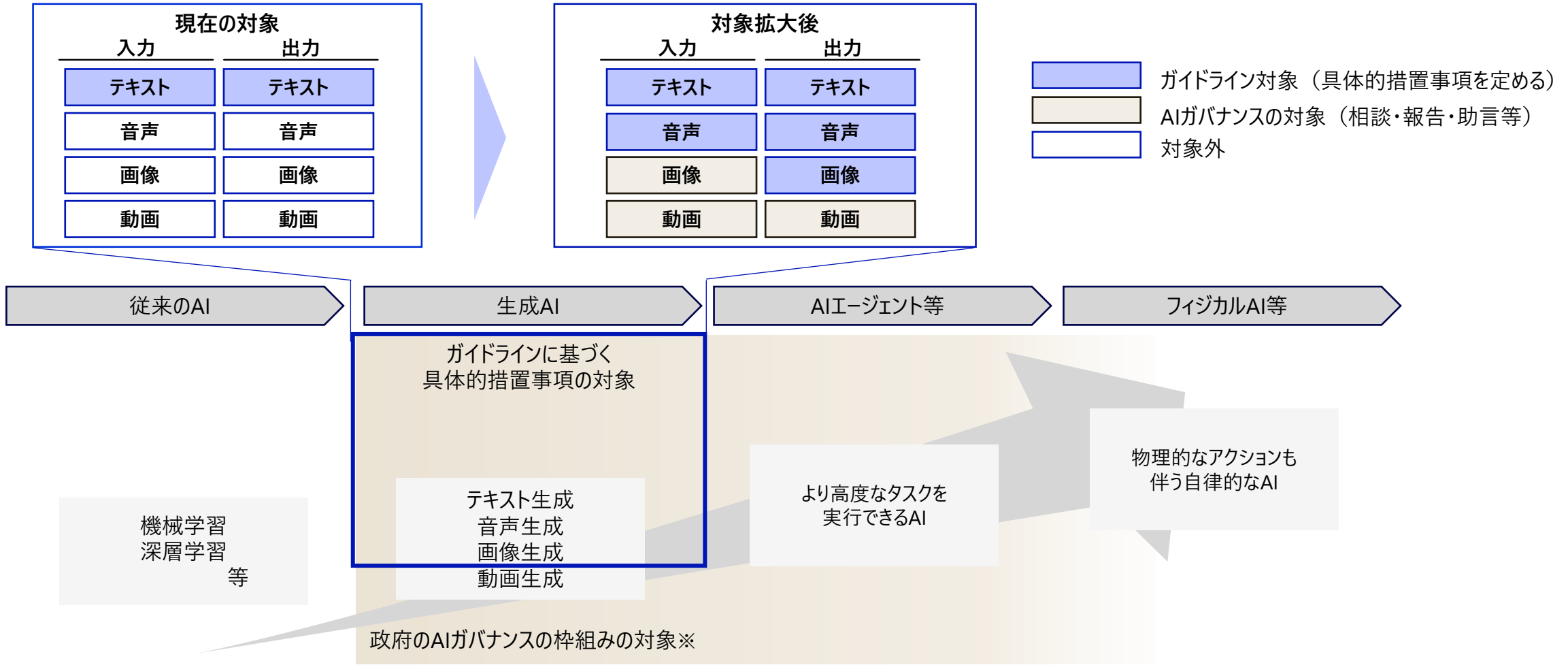
目次（抜粋）	改定内容又は主に紐づく論点・留意事項
6　政府における生成 AI の調達・利活用に係るルール	
6.1.1　各種法令・ガイドライン等を踏まえた対応事項	• 「Ⅳ-セキュリティ　ISMAPとの対応関係整理」に伴い、ISMAPポータルにて公開された記載を脚注に追加 • 「Ⅳ-知財等対策」の関係ガイドライン等に関する記載追加 • 「Ⅳ-セキュリティ」の総務省AIセキュリティ分科会による「AIのセキュリティ確保のための技術的対策に係るガイドライン（案）」を踏まえ、セキュリティ確保にあたって総則的に参考になる内容を取り込み
6.2.2　各府省庁内におけるAIガバナンスの確保	• デジタル社会推進標準ガイドラインを踏まえ、システム監査に係る記載を追加（Ⅲ-①各役割の対応事項整理）
6.3.1　生成AIシステムの企画時の対応事項	• 「Ⅱ-①　高リスク判定見直し」に伴い、要機密情報の保護のための、権限管理に関する記載、生成AIシステムの特性を考慮したログ取得に関する記載の追加 • 「DS-680.2 ウェブコンテンツガイドライン」を踏まえ、ウェブサイト等による行政情報及び機能提供を行う際に生成AIに関する部分を参考とすることを記載
6.3.2　生成AIシステム調達時の対応事項	• 調達チェックシート、契約チェックシートの詳細説明を削除又は脚注へ移動（留意事項①－（１）） • AISI「インシデントレスポンスアプローチガイドブック」を踏まえ、制御性の強化で必要な内容を追記（留意事項②－（１））
6.3.3　生成AIシステムの構築・リリース前の準備時の対応事項	• 特に不特定外部者（一般国民等）による府省庁外利用の場合は、適切な利用規約 等を整備し、個別の同意を取得する旨を追記 • 「Ⅰ-②　対象生成AI拡大」を踏まえ、利用者が高度なタスクを実行できるAIエージェントなどを作成できる生成AIシステムの場合の、適正利用の促進のための記載追加

ガイドライン本紙の改定内容案の全体像（6~7章・附則）

目次（抜粋）	改定内容又は主に紐づく論点・留意事項
6　政府における生成 AI の調達・利活用に係るルール	
6.5　政府における生成 AI システムの提供者の対応事項	• 生成AIモデルのメジャーアップデートや大規模な追加学習での考慮事項を追記 • 利用者への利用規約の提示や、知的財産権等侵害防止のための仕組みに関する情報提供等を追加 • 「Ⅱ-①　高リスク判定見直し」に伴い、アクセス制限機能の運用に関して追記 • 権利侵害等の認識後に、是正等の合理的な措置を講じる旨を追加（留意事項②−（１）） • 「Ⅰ-②　対象生成AI拡大」を踏まえ、利用者が高度なタスクを実行できるAIエージェントなどを作成できる生成AIシステムで、出力結果の適切さの判断を行わないものを作成した場合の利用者からの報告を追加
6.7　生成AIシステム特有のリスクケースへの対応	• 「Ⅱ-③　リスクケースの整理」と「Ⅰ-②対象生成AI拡大」を踏まえ、報告の対象とするリスクケース例の記載追加
7　今後の進め方	• 記載を簡素化（留意事項①−（１））
附則	• 「2.2.4 本ガイドラインの適用開始時期等について」を移動し、記載更新（留意事項①−（１））

ガイドラインの対象生成AI拡大の考え方

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エージェントなどについては、ガイドラインに基づく相談・報告・助言等の対象とする



※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

ガイドラインの対象生成AI拡大に伴う改定内容案

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エージェントなどについては、ガイドラインに基づく相談・報告・助言等の対象とする

主な改定箇所①

2.2.2　本ガイドラインが対象とする生成AI

本ガイドラインが対象とする生成AIは、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もの~~大規模言語モデル（LLM）を構成要素とするテキスト生成AI[16]~~とする~~（テキスト及び画像を生成するAI等については、テキストの生成について対象とする。）~~。

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、~~なお、画像や~~動画等を生成する~~AI~~もの、より高度なタスクを実行できる~~AI~~もの（AIエージェント等）~~、その他のAI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、具体的には、「4.1　政府全体のAIの利活用促進とAIガバナンスのための体制構築」及び「4.2　各府省庁におけるAIガバナンス体制の整備」並びに「6.7　生成AIシステム特有のリスクケースへの対応」の対象とすることとする（以降、「生成AIモデル」の語は、本ガイドラインが対象とする生成AIを構成要素とする生成AIモデルを、「生成AIシステム」の語は、本ガイドラインが対象とする生成AIモデルを構成要素とする生成AIシステムを指すものとする。）。

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~~適用範囲等の~~拡充を検討することとする[17]。

主な改定箇所②

6.3.3　生成AIシステムの構築・リリース前の準備時の対応事項
②適正利用の促進

~~利用者が高度なタスクを実行できるAIエージェントなどを作成できる生成AIシステムの場合には、出力結果の適切さの判断を行わずにタスクを実行するもの（リスク判定ロジックのC①に相当）の作成を制限するか、~~又は利用者がこれを作成したときには利用開始前に提供者又はAI統括責任者（CAIO）に報告を求める旨の利活用ルールを整備し、周知する。

主な改定箇所②

6.5　政府における生成AIシステムの提供者の対応事項

利用者が高度なタスクを実行できるAIエージェントなどを作成できる生成AIシステムにおいて、出力結果の適切さの判断を行わずにタスクを実行するもの（リスク判定ロジックのC①に相当）を作成した場合には、提供者又はAI統括責任者（CAIO）に報告させることが求められる（提供者が報告を受ける場合においては、AI統括責任者（CAIO）に報告することが求められる）。

リスク判定ロジック改定の方向性

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すこととする
※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	課題	改定の方向性
C.対象データ	•機密性２情報又は個人情報を取扱うユースケースの拡大が見込まれるが、府省庁内で適切に対策がなされれば、漏洩等のリスクは抑制されるにも関わらず、高リスクと判定される	•政府情報システムは情報セキュリティ統一基準に基づく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することとし、Cの対象データの項目は、リスク判定ロジックから外すこととする
B.適用業務	•「過失」の指す内容、主語が必ずしも明らかでなく、当てはめがしにくい •「重大な影響を及ぼす可能性のある業務」の注釈に、人間の生命・身体・財産はあるが、法人は記載がない	•「生成AIによる生成物の瑕疵（生成AIが提供する情報の誤り等）」に修正 •注釈の例示に「法人の事業に重大な影響を及ぼす業務」を追加
A.利用範囲	•B②重大な影響のない業務の場合、A②複数省庁横断と、③単一省庁で、リスク評価を異ならせる必要があるか •自治体職員や民間企業等の特定の限られた主体が利用する場合でも、 A①「国民等」として高リスクと判定される	•本ガイドラインに基づき、府省庁職員は、生成AI利活用ルールが適用され、研修等を受けることとなっているため、単一省庁と複数省庁横断で同じリスク評価とする •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.職員等による出力結果の判断	•B②重大な影響のない業務の場合に、A③単一省庁であれば、出力結果の適切さを判断せずに利用しても低リスクと判定されるが、AIエージェント等を対象とする場合の影響は未知数	•リスク判定において画像出力やAIエージェント等も対象とする場合には、B②重大な影響がないとされる業務でも、D①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、高リスクとの判定に一本化する •Aにおいて知見のある特定の府省庁外の者の場合も、同様の判定基準を適用する

リスク判定ロジック改定内容案（１/３）

「リスク判定ロジック改定の方向性」を踏まえ、「対象データ（旧C項目）」の軸をフローチャート上からは削除、「適用業務」の文言を修正、「利用範囲」の区分の見直しを実施

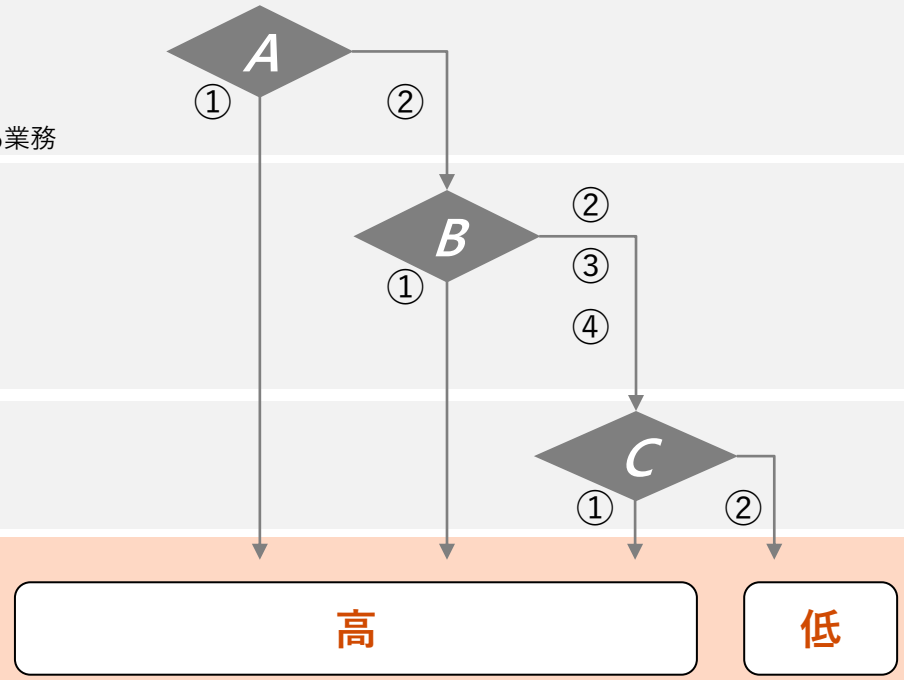
リスク判定ロジックの修正

「リスク判定ロジック改定の方向性」の改定の方向性を踏まえ、リスクロジック全体を修正（修正箇所は赤字記載）
3.2　高リスクな生成AI利活用の考え方　図３「リスク判定ロジック」

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に適用する
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

B. 利用範囲
①不特定外部者（一般国民等）による利用
②特定外部者※による利用
③政府職員等による府省庁内・複数府省庁横断での利用（共通システムでの利用等）
④政府職員等による府省庁内・単一省庁での利用
※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

方向性を踏まえたリスク判定ロジック改定内容案（２/３）

高リスク判定ロジックの見直しを踏まえ、「3.2　高リスクな生成AI利活用の考え方」の表4「リスク軸とその考え方」記載を修正している。

主な改定箇所①

3.2　高リスクな生成AI利活用の考え方　表４「リスク軸とその考え方」

リスク軸	説明	観点
~~BA~~.生成AI利活用業務の性格	~~過失~~生成AIによる生成物の瑕疵（生成AIが提供する情報の誤り等）が重大な影響を及ぼす可能性のある業務（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）において生成AIを利活用する場合、リスクが高くなると考えられる。	① ~~過失~~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において利活用する ② ~~過失~~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において利活用しない
~~AB~~.利用者の範囲・種別	調達・~~利活用する~~生成AIシステムの利用者の範囲によってリスクの大きさや影響を与える範囲の大きさが異なる。特に不特定外部者（一般国民等）が利活用する~~国民等が用いるような形で府省庁外で利活用する~~サービスは、政府職員等による府省庁内での利活用に比べリスクが高いと考えられる。また、政府職員や、外部利用者でも特定の者（地方公共団体や民間企業等）が	①不特定外部者（一般国民等）による利用国民等による府省庁外利用 ②特定外部者※による利用 ~~②~~③ 政府職員等による府省庁内・複数府省庁横断での利活用（府省庁共通システムでの利活用等）

リスク軸	説明	観点
	~~利用される場合などは~~、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と考えられる~~ことから~~場合については、~~必ずしも府省庁内利用であっても、複数府省庁横断で利活用する場合などはリスクが顕在化した際の影響範囲が大きくなるため、この点を考慮することが必要となるリスクが高いものではないと考えられる。~~	~~③~~④ 政府職員等による府省庁内・単一府省庁での利活用 ※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.要機密情報や個人情報の学習等の有無~~		
~~D~~C.出力結果の政府職員等による判断を経た利活用	生成AIの出力結果は必ずしも正しいものとは限らない。そのため、生成AIシステムの出力結果に対して政府職員等が判断せずそのまま利活用するような業務設計をする場合、リスクが高くなると考えられる。	① 生成AIシステムの出力結果の適切さを政府職員等が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政府職員等が判断して利活用する

方向性を踏まえたリスク判定ロジック改定内容案（３/３）

「対象データ（旧C項目）」の軸をリスク軸からは削除することを踏まえ、アクセス制御機能の適切な運用が必須であること、不正操作等がなされていないことの検証を行う場合等に必要なログの取得及び管理を行うこと、当該生成AIシステムが要機密情報又は個人情報を扱う場合、生成AI固有のリスクとして、入力が学習される設定となっている場合に、入力者以外にも漏洩するリスクへの対処をすることが必要の旨を別途記載

主な改定箇所②

6.3.1　生成AIシステムの企画時の対応事項

③ 企画者は、生成AIシステムにおいても、「政府機関等のサイバーセキュリティ対策のための統一基準」に基づき、権限管理を適切に実施する観点から、取り扱う情報の格付及び取扱制限等に従い、権限を有する者のみがアクセス制御の設定等を行うことができる機能を設けておくことが求められる[34]。

[34] 生成AI固有のリスクとして、当該生成AIシステムが要機密情報又は個人情報を扱う場合で、入力が学習される設定となっている場合に、入力者以外にも漏洩するリスクへの対処が必要となる。

④ 企画者は、「政府機関等のサイバーセキュリティ対策のための統一基準」に基づき、情報システムが正しく利用されていることの検証及び不正侵入、不正操作等がなされていないことの検証を行うために必要なログの取得及び管理を行う必要がある。生成AIシステムについては、「6.5　政府における生成AI システムの提供者の対応事項」での生成 AI システムの適正利用の確認のため、生成 AI システムへの入力や出力、アクセス履歴等のログの取得及び管理を検討することが求められる。

主な改定箇所③

6.5　政府における生成 AI システムの提供者の対応事項
①システムの運用

- → 生成 AI システム及び生成 AI システムが取り扱う情報へのアクセスを許可する主体が確実に制限されるように、アクセス制御機能を適切に運用する。

報告の対象とするリスクケース例追加改定内容案

対象生成AIの拡大を受けて、音声の入出力や画像の生成よるリスクケース例を追加（実際の修正箇所は次スライド参照）

リスクケース例	①プロンプト入力	②コンテンツ生成 （モデルによる出力生成）	③出力確認・利用判断	④二次利用・公開・配布
不適切な入力・情報漏洩	外部に接続された環境で、要機密情報や個人情報を誤って入力することで情報漏洩する	–	–	–
差別・偏見の出力	–	生成AIが人種・性別・文化等に関する偏見や差別を含む社会的に大きな問題となり得る出力を行う	–	–
不自然な表現の出力	–	–	–	[画像]生成した不自然な表現の画像を公開文書等に掲載したことで批判を受ける
攻撃的または危険なコンテンツの出力	–	生成AIが攻撃的または危険なコンテンツを生成する	–	–
誤情報の出力	[音声] 音声入力が不正確なことにより、それを元に作成した記録の意味内容が本来と異なるものとなり、重要な記録の正確性が損なわれる	–	生成AIが事実と異なる情報を出力し（ハルシネーション）、利用者がその情報を利用したことによって利用者もしくは第三者に不利益を与える	–
知的財産権等の侵害等	–	–	–	利用者が生成AIにより既存の作品に類似し、著作権の侵害等の問題が生じる可能性が高いコンテンツを意図せず生成し、利活用したことで当該作品に係る権利者等から削除等の申出を受ける [音声]音声回答機能が、著名な個人の音声に類似してしまい、本人から停止の申出を受ける

生成AIによるリスクと報告対象のリスクケース例追加の改定内容案

対象生成AIの拡大を受けて、音声の入出力や画像の生成よるリスクケース例を追加

主な改定箇所①

5.2　生成AIによるリスク

政府における生成AIの調達・利活用においても、~~上記のようなこうした~~リスクに留意することが必要であり、~~政府として、~~例えば、以下のようなリスクについても考慮が必要となる。

- 生成AIを用いて出力した画像が、既存の著作物に類似した要素を意図せず含んで出力され、その状態に気付かないまま利用してしまうリスク
- 音声回答機能が、著名な個人の声質や話し方と意図せずに類似してしまうリスク

主な改定箇所②

6.7　生成AIシステム特有のリスクケースへの対応

以下に、生成AIシステム特有のリスクケースの例を示す。

- → 音声回答機能が、著名な個人の音声に類似してしまい、本人から停止の申出があった。
- → 不自然な表現の画像を生成し不自然であると気付かずに利用し、公開文書等に掲載したことで国民から批判を受けた。
- → 音声入力が不正確なことにより、それを元に作成した記録の意味内容が本来と異なるものとなり、重要な記録の正確性が損なわれていることが指摘された。

セキュリティ関連の記載拡充の改定内容案（１/２）

ISMAPポータルサイトでの情報掲載を踏まえた補足情報を記載。

ISMAPに関する補足情報掲載

2026年1月9日に公開された「生成AIサービスに関する留意点について」を踏まえ、クラウドサービス事業者、生成AIモデル提供者向けにモデルの個別登録に関する周知事項を脚注に追記

6.1.1　各種法令・ガイドライン等を踏まえた対応事項

[27] クラウドサービス事業者が、生成 AI モデルを提供する事業者から生成 AI モデルの提供を受け、当該生成 AI モデルの扱うデータに対してセキュリティ管理機能を適用する自らの生成 AI 開発基盤において生成 AI サービスを提供する場合（PaaS に相当）に、当該生成 AI 開発基盤を言明対象範囲に含めて ISMAP に登録したときには、通常は当該生成 AI サービスが取り扱うデータのセキュリティは、ISMAP のセキュリティ要件を満たす状態とみなされる。この場合において、クラウドサービス事業者が生成 AI 開発基盤を言明対象範囲に含めて ISMAP の登録を行う場合には、提供される個々の生成 AI モデルを言明対象範囲に含める必要はない。また、提供される生成 AI モデルは必ずしも ISMAP に登録されている必要はない。

セキュリティ関連の記載拡充の改定内容案（２/２）

主要な改定内容として、ISMAPポータルサイトでの情報掲載を踏まえた補足情報を記載。また、総務省AIセキュリティ分科会の「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）」を踏まえての記載拡充を実施している

LLM等に対する攻撃や主な対策の概観

6.1.1　各種法令・ガイドライン等を踏まえた対応事項
LLM及びLLMを構成要素に含むAIシステムを対象として、「不正操作による機密情報の漏えい、AIシステムの意図せぬ変更や停止が生じないような状態」に対する脅威への技術的対策例を追記

⑦　セキュリティ確保の観点では、「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（P）」（総務省）が、LLM 及び LLM を構成要素に含む AI システムを対象として、「不正操作による機密情報の漏えい、AI システムの意図せぬ変更や停止が生じないような状態」に対する脅威への技術的対策例を示している。AI の性質上、脅威を生じさせる要因等を完全に排除することは困難であること、単独の対策実施により脅威を生じさせる要因を排除することは困難な場合があることを前提に、企画者・開発者・提供者それぞれが、同ガイドラインが示す技術的対策例も踏まえつつ、できる限り複数の対策を講じるなど適切にリスク対策を行い、リスクを低減することが必要である。同ガイドラインが示す直接プロンプトインジェクション攻撃、間接プロンプトインジェクション攻撃及び DoS 攻撃（サービス拒否攻撃）への主な対策の概観は、表８に示すとおりである。また、【別紙３】調達チェックシート（生成AI システム用）においては、同ガイドラインが示す技術的対策例も踏まえつつ、セキュリティ確保の評価観点からの対策例を記載している。

表8 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（P）「プロンプトインジェクション攻撃及びDoS攻撃（サービス拒否攻撃）への主な対策（概観）」に技術的な対策を例示

	AI 開発者における対策	AI 提供者における対策				
	安全基準等の学習による不正な指示への耐性の向上	システムプロンプトによる不正な指示への耐性の向上	ガードレール等による入出力や外部参照データの検証			オーケストレータや RAG [33] 等の権限管理
			入力プロンプトの検証	外部参照データの検証	出力の検証	
直接プロンプトインジェクション攻撃	○	○	○		○	○
間接プロンプトインジェクション攻撃	○	○	○	○	○	○
DoS 攻撃（サービス拒否攻撃）	○	○	○			

知的財産権等対策に係る記載拡充の改定内容案

主要な改定内容として、文化庁「AIと著作権に関するチェックリスト＆ガイダンス」、経産省「コンテンツ制作のための生成AI利活用ガイドブック」の内容を踏まえ、ガイドラインの記載拡充を実施している

参考となるガイドラインの記載追加

6.1.1　各種法令・ガイドライン等を踏まえた対応事項

⑧　生成AIの調達・利活用にあたっては、知的財産権等のリスクに対する対策を踏まえることが必要である。本ガイドラインは、「コンテンツ制作のための生成AI利活用ガイドブック」（令和６年７月５日経済産業省）や「AIと著作権に関するチェックリスト＆ガイダンス」（令和６年７月31日文化庁著作権課）に記載された対策例を踏まえた内容となっているが、詳細を確認する必要がある場合にはこれらを参考に、調達・利活用にあたって適切に判断することが必要である。

提供者の対応事項の記載追加

6.5　政府における生成AIシステムの提供者の対応事項
①システムの運用

- 利用者に対し、知的財産権等に係る対策の取組として、利用規約等を提示するとともに、可能な範囲でモデルにおける知的財産権等の侵害等防止のための仕組みに関する情報提供を行う。

その他、主要な改定内容案

AISI「AIインシデントレスポンス・アプローチブック」を踏まえ、観測性の向上に関する記載を拡充。また、他に、生成AIモデルのメジャーアップデートや追加的な学習を大規模に行った際の考慮事項を記載追加

観測性の向上のための記載追加

6.3.2　生成AIシステムの調達時の対応事項

③　企画者は、調達する生成 AI システムに関して、生成 AI モデル、アーキテクチャ等の生成 AI システムの特性、生成 AI システムの主要性能指標情報等の情報を取得する。また、リスク分析やリスク対応の検討のために合理的な範囲で、学習に使用されたデータや学習方法、データセットの情報も、取得に努める。

生成AIモデルの改善に関する記載追加

6.5　政府における生成 AI システムの提供者の対応事項
①システムの運用

生成 AI モデルのメジャーアップデートが実施される場合や追加的な学習を大規模に行った場合には、生成 AI システムの目的・用途及びコストとの関係も考慮しつつ、利用者への提供前に生成 AI システムの出力が期待品質を満たしていること、及び不適切な生成やバイアスが発生していないこと、セキュリティ対策、非機能要件や必要コストの変化等を確認した上で提供を開始する。

リスクロジック改定方向性を踏まえた改定内容案

高リスク判定ロジックの見直しを踏まえ、要機密情報の取扱いに関する記載を「（2）利活用中のルール」の「①入力データ又はプロンプトにおけるルール」及び「②AI生成物利活用におけるルール」に追記

要機密情報の取扱いに係る利活用中のルールの記載追加

①入力データ又はプロンプトにおけるルール

- ~~利用者側の不理解やミスにより生じるリスクがあることを踏まえて、~~生成 AI システムの利用目的の範囲内で、要機密情報や個人情報の入力の可否を含む利用方法を遵守し、当該生成 AI システムを適切に利活用すること（例：生成 AI システムの提供者から説明された利用方法や必要に応じてマニュアルと照らしつつ生成 AI システムを活用する。生成 AI システムの提供者から説明された利用目的範囲外の利活用や禁止されている場合には要機密情報の入力をしない。）。↵

②AI生成物利活用におけるルール

- 責任を持って生成 AI の出力結果の業務への利用判断を行うこと（例：入力データ又はプロンプト、要機密情報を入力、参照又は学習した場合の出力結果の機密性、出力結果に含まれるバイアス、音声文字起こしの固有名詞等の誤りなどに留意して、業務に活用して問題ないかを利用者が判断する。判断に迷う場合は利活用しないこととする。）。↵

知的財産権等対策に係る記載拡充の改定内容案

主要な改定内容として、文化庁「AIと著作権に関するチェックリスト＆ガイダンス」、経産省「コンテンツ制作のための生成AI利活用ガイドブック」の内容を踏まえ、「（2）利活用中のルール」に知的財産権等に係る記載を拡充

プロンプト入力段階の知的財産権等の対策例追加

プロンプト入力段階におけるルールと対策例の記載追加

- プロンプトの入力において、知的財産権等など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[4]

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など、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・登録商標など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

AI生成物の利用段階の知的財産権等の対策例追加

AI生成物の利活用におけるルールと対策例の記載追加

- 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。そのうえで、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２参照）。

表 ２ AI 生成物の利用段階の知的財産権等の対策の例[7]

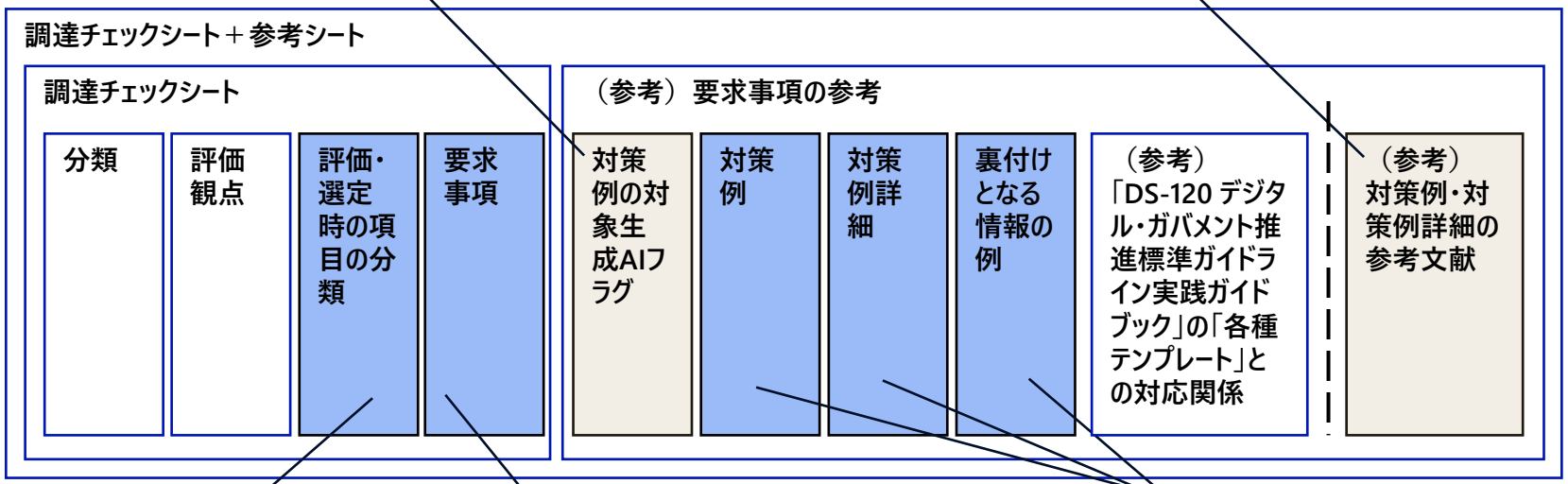
場面	対策例
①著作物の利用	・ 生成 AI を利用しない従来のコンテンツ制作と同様に、AI 生成物（やそれを編集・加工したもの）について、他人の著作物と同一・類似でないかどうかを、インターネット検索等を用いて確認する ・ 他人の著作物と同一・類似の場合には、利用を避ける、権利者から許諾を得る、類似しないように作成しなおしたうえで利用する 等

調達チェックシートの改定内容案の全体像

対象生成AIの拡大等を踏まえつつ、調達チェックシートの有用性向上のため、構成要素や記載内容を改定

【対策例の対象生成AIフラグ】
- 対象生成AIの拡大を踏まえて、対策例の対象生成AI（入力（テキスト/音声）、出力（テキスト/画像/音声））をフラグ付け

【対策例・対策例詳細の参考文献】
- 対策例・対策例詳細の記載の理解を深めるために参考になる参考文献の記載



【項目の分類見直し】
- リスク軸の見直しを踏まえ、評価・選定時の項目の分類の設定見直し

【要求事項の記載修正】
- 対象生成AIの拡大を踏まえての要求事項の記載修正
- 達成条件が明確でない記載の表現見直し
- 知的財産権等対策に係る要求事項を追加
- 制御性の向上に係る要求事項を追加

【対策例・対策例詳細・裏付けとなる情報例の記載修正、追加】
- 対象生成AIの拡大を踏まえての記載修正、対策例等の追加
- 生成AIのバイアスに係る対策例等を追加
- 生成AI品質マネジメントガイドラインを踏まえ対策例等追加
- 知的財産権等対策に係る対策例等を追加
- 制御性の向上に係る対策例等を追加

新規追加
記載の更新

要求事項の主要な改定内容案一覧

調達チェックシートの要求事項に、知的財産権等の対策（個人情報保護と一緒だったものを分離）、高い信頼性が求められる生成AIシステムの復旧対策及びベンダーロックイン回避の基本項目を新たに追加。また、達成条件が明確でない記載（「～状態としていること」等）の表現見直しや、対象生成AIの拡大等を踏まえた用語修正を実施

本改定案において新規追加・記載更新した要求事項の抜粋

【評価観点#12 ベンダーロックインの回避】

- 要求事項#12「生成AIシステムのアーキテクチャ設計と実装において、ベンダーやシステムの移行の容易性も踏まえた開発や運用が可能であること」

【評価観点#14 公平性と包摂性】

- 要求事項#20「生成AIシステムによる出力に有害なバイアス~~やを含まず~~、不当な差別の~~を含まない状態と~~入出力制限等をしていること」

【評価観点#16 個人情報、プライバシー、知的財産等】

- 要求事項#23「生成AIシステムにお~~いて取得・処理・保存するける~~個人情報~~のについて~~適切な取扱い~~が確保されるとともに、知的財産~~とプライバシー~~がの~~保護~~される状態とが図られるよう対策~~していること」
- 要求事項#24「生成段階において、知的財産権、肖像権、パブリシティ権の保護が図られるよう対策を講じていること」
- 要求事項#25「生成AIシステムの学習段階において、知的財産権等の保護が図られるよう対策を講じていること」（開発や追加学習を行う場合に適用）

【評価観点#17 セキュリティ確保】

- 要求事項#27「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を検知した後、その影響を最小限に抑えつつ、迅速な封じ込めと復旧能力を確保するための対策を講じていること」（生成AIシステムに対して高い信頼性が求められる場合に適用）

新規追加　記載を修正

ガイドラインの対象生成AI拡大に伴う改定内容案

主要な改定内容として、対象生成AIの拡大を踏まえて対策例の対象生成AIフラグの追加を実施

対策例の対象生成AIフラグ

対象生成AIの拡大を踏まえ、対策例の対象となる生成AIの識別用フラグ追加

要求事項#	要求事項	テキストを含む	音声を含む	テキストを含む	画像を含む	音声を含む	対策例	対策例詳細
1	AI事業者ガイドライン共通の指針を理解・把握・対応していることの宣言が可能であること	●	●	●	●	●	AI事業者ガイドライン「共通の指針」のチェックリストについて提出及び対応状況~~が~~を説明可能~~である~~とする	-
2	生成AIシステムの開発・運用に関して、AIガバナンス（※）が適用されていること ※AIにもたらされる正のインパクトを最大化しつつ、AIによるリスクを受容可能な水準で管理する統制の仕組み・業務	●	●	●	●	●	生成AIシステムの開発・運用にAIガバナンスのゴールを考慮~~している~~する	-
		●	●	●	●	●	生成AIシステムの開発・運用において、AIガバナンスのゴールからの乖離やリスク評価を行い、是正対応を行~~っている~~う	生成AIシステムの開発・運用にあたって、AIガバナンスのゴールからの乖離を特定~~している~~する
								生成AIシステムの開発・運用にあたって、関連するルールからの逸脱を確認したり、リスクを評価~~している~~する
								AIガバナンスのゴールからの乖離の特定やリスク評価にあたって、必要に応じて外部関係者も含め~~てい~~る

バイアスに係る記載拡充の改定案

主要な改定内容として、調達チェックシートにて「バイアスの有無が発生していないことを確認する」ように記載しているところ、政府のユースケースに応じて具体的に留意すべき点を明確化すべく、利用目的に応じ、バイアスや出力制限が支障となるかを踏まえ、適切な生成AIを選択するべき旨を追加

バイアスに係る対策例追加

評価観点#14「公平性と包括性」の要求事項「生成AIシステムによる出力に有害なバイアスや、不当な差別の入出力制限等をしていること」に、対策例、対策例詳細等の記載を追加

調達チェックシート			（参考）要求事項の参考		
評価・選定時の項目の分類	**要求事項#**	**要求事項**	**対策例**	**対策例詳細**	**裏付けとなる情報の例**
不特定外部者（一般国民等）による府省庁外利用の場合は基本項目として適用	~~19~~20	生成AIシステムによる出力に有害なバイアス~~を~~やを含まず、不当な差別の~~を含まない状態と入出力制限等を~~していること	学習データ、モデルの学習過程において、生成AIシステムに不公正なバイアスがないかを検証する~~手段を有している~~ また、生成AIモデルが、法令等に基づき特定の分野における出力制限をかけられている可能性について、整理して提供する	利用目的に応じ、バイアスや出力制限が支障となるかを踏まえ、適切な生成AIを選択する	生成AIシステムの検証プロセス・検証方法が分かる資料、生成AIモデルの提供企業の所在国におけるAIに対する規律や当該生成AIモデルの生成結果のバイアスに対する評価結果等

リスク判定ロジック改定を踏まえた改定内容案（1/2）

高リスク判定ロジックの見直しを踏まえ、生成AIシステム特有の留意点として、利用者の入力情報は当該利用者への回答の生成にのみ利用される仕組みとなるよう確保する旨を別途記載

個人情報の入力に関する記載追加

評価観点#16「個人情報、プライバシー、知的財産等」に、対策例、裏付けとなる情報の例を追加

調達チェックシート			（参考）要求事項の参考		
評価・選定時の項目の分類	**要求事項#**	**要求事項**	**対策例**	**対策例詳細**	**裏付けとなる情報の例**
個人情報を取り扱い、プライバシー~~、知的財産を取り扱う~~の保護が必要な場合は基本項目として適用	2~~3~~	生成AIシステムにお~~いて取得・処理・保存するける~~個人情報~~のについて~~適切な取扱い~~が確保されるとともに、知的財産~~とプライバシー~~がの~~保護~~される状態とが図られるよう対策~~していること	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可能性があるときには、原則として、利用者の入力情報は当該利用者への回答の生成にのみ利用される仕組みとなるよう確保する	-	個人情報等が入力される可能性がある場合、原則として、利用者の入力情報は当該利用者への回答の生成にのみ利用される仕組みの実装方法がわかる資料等

リスク判定ロジック改定方向性を踏まえた改定内容案（2/2）

高リスク判定ロジックの見直しを踏まえ、入力された要機密情報を参照又は学習する場合、権限を有する者以外の回答の生成に、当該入力された要機密情報が用いられないような仕組みとする旨を別途記載

要機密事項の取扱いに関する記載追加

評価観点#17「セキュリティ確保」の要求事項#26「生成AIシステム全体の脆弱性に対処し、不正操作による影響の防止措置を取っていること」に、対策例詳細、裏付けとなる情報の例を追加

調達チェックシート			（参考）要求事項の参考		
評価・選定時の項目の分類	**要求事項#**	**要求事項**	**対策例**	**対策例詳細**	**裏付けとなる情報の例**
基本項目	2~~3~~6	生成AIシステム全体の脆弱性に対処し、不正操作による影響~~を防いでの防止~~措置を取っていること	生成AIシステム全体に関するセキュリティ確保の対策を講じる	入力された要機密情報を参照又は学習する場合、権限を有する者以外の回答の生成に、当該入力された要機密情報が用いられないような仕組みとする	入力された要機密情報を参照又は学習する場合、権限を有する者以外の回答の生成に、当該入力された要機密情報が用いられないような仕組みとする方法がわかる資料等

セキュリティ関連の記載拡充の改定内容案（１/２）

主要な改定内容として、総務省AIセキュリティ分科会の「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）」を踏まえ、評価観点#17「セキュリティ確保」に係る要求事項＃26「生成AIシステム全体の脆弱性に対処し、不正操作による影響の防止措置を取っていること」に、対策例・対策例詳細を追加

生成AIシステム全体に係る対策例等の追加

事後学習、オーケストレータの権限等に係る対策例詳細を追加

生成AIモデルが意図しない出力を行わないよう、安全基準を事後学習させる
生成AIモデルが従うべき指示の優先度を定義し、優先度の高い指示（例：システムプロンプト）を常に優先的に処理するよう、生成AIモデルを事後学習させる
生成AIシステムの意図しない動作の防止のために、オーケストレータの権限等に不備がないことの対策を講じる。生成AIモデルや連携システムを操作するオーケストレータに係る権限を必要最小限とすることで、生成AIモデルが攻撃を受けた場合の被害拡大を抑制する

生成AIシステムの出力に係る対策例等の追加

出力の検証機能、モデル抽出攻撃等への対策例詳細を追加

生成AIシステムの意図しない動作の防止のための出力の検証機能を具備し、出力を意図しない情報が出力に含まれていないか検証し、検知した場合には応答を拒否する
単語の出現確率など、攻撃者に悪用され得る情報を必要に応じて応答から除外することで、モデル抽出攻撃への対策を行う（主にテキスト出力を念頭）

外部データ参照に係る対策例等の追加

生成AIモデルが生成AIモデルの外のデータ（RAG、Webサイトや外部のデータベース等）を参照する場合のセキュリティ確保の対策例詳細、裏付けとなる情報の例を追加

【対策例詳細】	【裏付けとなる情報の例】
RAGにて要機密情報のように参照権限が設定されている情報を利用する場合、RAG用のデータ及びデータストアへの参照権限をユーザや役割に応じて適切に設定する（主にテキスト出力を念頭）	RAG用のデータ及びデータストアへの参照権限をユーザや役割に応じて適切に設定し、不備がないかの確認方法がわかる資料等
Webサイトや外部のデータベースなど、外部データを参照する場合には、これらに意図しない出力を行わせる指示が含まれていないか検証し、そのような指示を検知した場合には、処理の拒否等の措置を講じる	Webサイトや外部のデータベースなど、外部データを参照する場合には、これらに意図しない出力を行わせる指示が含まれていないか検証し、そのような指示を検知した場合には、処理を拒否する等の仕組みの実現方法がわかる資料等
生成AIシステムに、入力プロンプトと外部参照データを明確に区分させ、外部参照データに高い注意を払わせる	生成AIシステムに、入力プロンプトと外部参照データを明確に区分させ、外部参照データに高い注意を払わせる仕組みの実現方法がわかる資料等

セキュリティ関連の記載拡充の改定内容案（２/２）

主要な改定内容として、総務省AIセキュリティ分科会の「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）」を踏まえ、評価観点#17「セキュリティ確保」に係る要求事項「生成AIシステム全体の脆弱性に対処し、不正操作による影響の防止措置を取っていること」に、対策例・対策例詳細を追加

生成AIモデルに対するプロンプト入力に係る対策例等の追加

入力プロンプトを①入力全般、②特に利用者が入力するプロンプト、③システムプロンプトの3ケースに整理し、「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）」を踏まえての内容としては、①②に係る対策例・対策例詳細を追加

（参考）要求事項の参考	
対策例	**対策例詳細**
生成AIモデルへのプロンプト入力に関するセキュリティ確保の対策を講じる（入力プロンプト全般）	プロンプト内容の有害性を生成AIモデルへの入力前に検知する等、生成AIモデルへの入力前段階での防御策を講じる 例えば、以下のような対策を講じる ・禁止リストを活用したプロンプト検知 ・有害な結果をもたらす可能性のあるコードの実行や生成を促す入力、または生成AIシステムの利用目的に照らして意図しない出力を生成させる指示がプロンプトに含まれていないか検証し、そのような入力を検知した場合にはプロンプトの一部削除による無害化や、処理の拒否等の措置を講じる ・プロンプトインジェクション等により、生成AIシステムのバックエンドシステムに意図していない操作が行われることがないよう対策する ・実装した生成AIシステムの防御策に関して、プロンプトインジェクションなどにより防御策の回避が可能とならないよう対策する
	生成AIシステムへの膨大なアクセスによる攻撃を抑制するためのレートリミットを導入する
生成AIモデルへのプロンプト入力に関するセキュリティ確保の対策を講じる（システムプロンプト）	システムプロンプトに制約事項やセキュリティ上の注意事項などを設定することで、生成AIモデルが意図しない出力を行わないようにする
	システムプロンプトに出力を意図しない機密情報（例：APIキー）等を直接記述することを避け、生成AIモデルが必要に応じて参照できるよう別個に管理する

知的財産権等対策に係る記載拡充の改定内容案（1/2）

主要な改定内容として、文化庁「AIと著作権に関するチェックリスト＆ガイダンス」、経産省「コンテンツ制作のための生成AI利活用ガイドブック」の内容を踏まえ、評価観点#16「個人情報、プライバシー、知的財産等」に係る要求事項の評価・選定時の項目の分類を整理。

評価・選定時の項目：
生成AIシステムで知的財産等を取り扱う場合の要求事項等の追加

要求事項「生成段階において、知的財産権、肖像権、パブリシティ権の保護が図られるよう対策を講じていること」を追加。併せて対策例と裏付けとなる情報の例を追加。

調達チェックシート			（参考）要求事項の参考	
評価・選定時の項目の分類	**要求事項#**	**要求事項**	**対策例**	**裏付けとなる情報の例**
生成AIシステムで知的財産等を取り扱う場合は基本項目として適用※	24	生成段階において、知的財産権、肖像権、パブリシティ権の保護が図られるよう対策を講じていること	モデルにおいて知的財産権等の侵害等防止のための仕組みを具備する	知的財産権等の保護のための措置（フィルタリング等）の実現方法がわかる資料

※①著作物の利用、②登録意匠・登録商標、他人の商品等表示・商品形態の利用、③人の肖像の利用、④人の声の利用が主に想定される。

知的財産権等対策に係る記載拡充の改定内容案（2/2）

主要な改定内容として、文化庁「AIと著作権に関するチェックリスト＆ガイダンス」、経産省「コンテンツ制作のための生成AI利活用ガイドブック」の内容を踏まえ、評価観点#16「個人情報、プライバシー、知的財産等」に係る要求事項の評価・選定時の項目の分類を整理。

評価・選定時の項目：
企画・開発において生成AIに学習を行わせる場合の要求事項等の追加

要求事項「生成AIシステムの学習において、知的財産権等の保護が図られるよう対策を講じていること」を追加。
併せて対策例、対策例詳細、裏付けとなる情報の例を追加。

調達チェックシート			（参考）要求事項の参考
評価・選定時の項目の分類	**要求事項#**	**要求事項**	**対策例**
企画・開発において生成AIに学習を行わせる場合は基本項目として適用	25	生成AIシステムの学習段階において、知的財産権等の保護が図られるよう対策を講じていること	知的財産権等の侵害等防止のため、適切なデータの学習を行う対策を講じる ※生成AIシステムの調達時に調達事業者と連携をして追加学習を行う場合や、政府職員が自ら生成AIの開発を行う場合のみ本項目の対象

その他、関係ガイドラインを踏まえた記載拡充の改定内容案（１/２）

AISI「CAIO設置・AIガバナンス実務マニュアル(案)」等を踏まえ、評価観点#8「ベンダーロックインの回避」にベンダーやシステムの移行の容易性も踏まえた開発や運用の必要性に関する要求事項を新規追加。

ベンダーやシステムの移行の容易性に関する記載追加

基本項目として、要求事項「生成AIシステムのアーキテクチャ設計と実装において、ベンダーやシステムの移行の容易性も踏まえた開発や運用が可能であること」を追加。
併せて対策例、対策例詳細、裏付けとなる情報の例を追加。

調達チェックシート			
評価・選定時の項目の分類	要求事項#	要求事項	対策例詳細
基本項目	12	生成AIシステムのアーキテクチャ設計と実装において、ベンダーやシステムの移行の容易性も踏まえた開発や運用が可能であること	標準化されたインタフェースやプロトコルを活用し、AIコンポーネントを疎結合に設計する
			モデルやベンダーの切り替えを想定した抽象化レイヤー (アダプタ) を用意し、「出口戦略」をアーキテクチャレベルで組み込む
			コンポーネントアーキテクチャを、再利用部品である生成AIモデルの置き換えが容易になるようにデザインする
			保守性の向上のため、生成AIシステムのコンポーネントをモジュール性を考慮して開発、運用する

その他、関係ガイドラインを踏まえた記載拡充の改定内容案（2/2）

AISI「AIインシデントレスポンス・アプローチブック」を踏まえ、観測性の向上と制御性の強化に関する記載を拡充

制御性の強化のための記載追加

生成AIシステムに対して高い信頼性が求められる場合、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を検知した後、その影響を最小限に抑えつつ、迅速な封じ込めと復旧能力を確保するための要求事項を追加

調達チェックシート			（参考）要求事項の参考	
評価・選定時の項目の分類	要求事項 #	要求事項	対策例	対策例詳細
生成AIシステムに対して高い信頼性が求められる場合は基本項目として適用	27	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を検知した後、その影響を最小限に抑えつつ、迅速な封じ込めと復旧能力を確保するための対策を講じていること	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のために、生成AIシステムの制御性を強化するための仕組みを具備する	改ざんされたモデルの別バージョンや代替モデルへの即時切り替えを実施する
				混入された悪意のある学習データの局所的な除去や追加学習を実施する
				異常時に影響を受けない上位権限を持ち、隔離された制御インタフェースを設置する

契約チェックシートの改定内容案の全体像

契約チェックシートの実効性向上のため、【具体例の充実化】及び【契約形態に関する記述の追記】等を実施

契約チェックシートにおける取決め事項

1. 生成AIシステムに係るインプットに関する取決め
2. 生成AIシステムに係るインプットの処理成果に関する取決め
3. 生成AIシステムに係るアウトプットに関する取決め
4. 生成AIシステムに係るアウトプットに関する処理成果の取決め
5. 生成AIシステムに係る契約上の取決め
6. 生成AIシステムに係るノウハウの取決め
7. 生成AIシステムに係る成果物の取決め
8. 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が発生した場合の事業者の対応義務、協力及びその範囲に関する取決め
9. 期待品質が満たされなくなった場合等において、そこから生じる被害を最小限に食い止めること及び、原因を特定し改善措置を講じる取決め

【インプットに係る補足説明の記載修正】
- オプトアウトの選択肢を取る場合の対応として「学習目的で使用しない」旨を明示

【インプット処理成果に係る補足説明の記載修正】
- DBだけではなくRAGをテキストデータに変換するプログラムなど、関連する成果物の権利帰属や利用方法を盛り込むべきこと等を追記

【契約に盛り込む条項内容例の記載修正・補足説明の記載修正】
- 請負契約が前提となっている状態を、生成AIシステム構築の実態を踏まえて修正（プロンプトのひな形など多様な成果物形態のイメージを契約当事者間であらかじめ共有しておくこと、生成AIシステム　構築にあたって期待品質を明確に契約条件に定めるのが困難な場合には、品質を満たすための取組履行の有無を盛り込むべき旨を追記等）
- 生成AIモデルそのものが納品物でない場合、システムコンポーネント部分の納品物を明確に特定すること等を追記

【ノウハウに係る用語の補足・契約に盛り込む条項内容例の記載修正】
- 「ノウハウ」として契約時に考慮すべき内容の明確化のため、記載内容を整理

【成果物の取決め事項追加】
- 契約形態と分けて、生成AIシステムに係る成果物の取決めを整理

契約チェックシートの改定内容案（1/3）

主要な改定内容として、AIシステム構築の実態を踏まえた記載修正を行い、それに伴い取決め事項#7「生成AIシステムに係る成果物の取決め」を追加を実施している

AIシステム構築の実態を踏まえた記載修正

取決め事項#5「生成AIシステムに係る契約上の取決め」に係る例示と補足説明の修正

契約時の項目	取決め事項	契約に盛り込む条項内容例	補足説明
基本項目	生成AIシステムに係る契約上の取決め	事業者が生成AIシステムを構築するうえで期待品質を満たすための取組の履行及び完成させる義務を定める条項	~~請負契約で事業者が生成AIシステムを完成する義務を負う場合、ユーザのサービス利用目的に照らして、どのような完成条件（完成時期、検収条件等）を定めるべきか検討して契約に盛り込むことが望ましい。~~ 期待品質に関する事項を契約に盛り込む前提として、納品物が何かを明確にする必要がある。生成AIシステムの開発の場合、生成AIモデルそのものが納品物でないときには、納品物としてシステムコンポーネント（ユーザーインターフェース・システムプロンプト・ファイルデータベース・ベクトルDB・入力フィルター・出力フィルター・外部連携コンポーネント 等）部分については、明確に特定することが望ましい。 生成AIシステムの期待品質を満たすために契約上、直接期待品質を契約条件に定める形態、又は、品質を確保するために実施すべき取組内容を定める形態があるが、納品物の性質及びその調達の目的を踏まえてどちらの形態を選択するかを検討する必要がある。 生成AIモデルに起因する性能・出力品質については、学習データや基盤モデル等の特性にも左右され、「○○以上の精度」等の成果保証を行うことが困難な場合があるため、その場合には、当該部分は成果保証ではなく、性能改善・品質向上に向けた技術的支援を受ける等の契約形態を取ることが望ましい。 但し、生成AIモデル以外のシステムコンポーネントに関しても、例えば「システムプロンプト構築を目的としたプロンプト設計・評価業務」を考えた際に、設計・評価・チューニングに関する部分は明確に成果物を定義することが困難と考えられるため、取り組みの履行有無について契約に盛り込むことが望ましく、性能要件を含めない形式としてのシステムプロンプト（指定された構造であり指定の要素が含まれている、トークン長の上限を満たす 等）、プロンプトのテストシナリオ、設定パラメータ群（回答の自由度や出力トークンの上限等）については請負契約に基づく成果物とすることが考えられる等、その調達の目的を踏まえて契約における成果物の範囲が品質担保可能な範囲内かという点を慎重に検討する必要がある。

契約チェックシートの改定内容案（2/3）

主要な改定内容として、「ノウハウ」として契約時に考慮すべき内容の明確化のため記載内容を整理し、用語の補足及び取決め事項#6「生成AIシステムに係るノウハウの取決め」に関する記載を修正している

「ノウハウ」の明確化を踏まえた記載修正

「ノウハウ」に関する用語の補足及び、取決め事項#6「生成AIシステムに係るノウハウの取決め」に係る契約に盛り込む条項内容例の修正

用語の補足
・ノウハウ：AI技術の研究・開発・利活用過程において、事業者又はユーザーが有し若しくは取得する知見、技術、情報等のうち知的財産として該当するもの。（※ノウハウについて、生データの取得や学習に適した生データ加工、学習用プログラムを用いた学習、学習済みモデルの調整、学習履歴、プロンプト履歴等が想定される。）

契約時の項目	取決め事項	契約に盛り込む条項内容例	補足説明
基本項目	生成AIシステムに係るノウハウの取決め	事業者またはユーザーのノウハウに関して、知的財産（※）として定義し、~~ノウハウの定義、~~ユーザーによる事業者の~~ノウハウの~~利用条件、事業者によるユーザーの~~ノウハウの~~利用条件、~~ノウハウの~~権利帰属に関して定めるする条項 ※発明、考案、意匠、著作物その他の人間の創造的活動により生み出されるもの（発見または解明がされた自然の法則または現象であって、産業上の利用可能性があるものを含む。）、および営業秘密その他の事業活動に有用な技術上または営業上の情報をいう。	ノウハウの定義に合致しない情報については、適用法令による制限がない限り、事業者がノウハウを自由に利用できる可能性があるため、この条項により契約による規律の対象となるノウハウの範囲やユーザーおよび事業者による利用条件（※）、権利帰属等を定めることが望ましい。 ※特に事業者の技術開発や学習目的等のサービス提供目的以外の目的で利用することを許容するか、検討して契約に盛り込むことが望ましい。

契約チェックシートの改定内容案（３/３）

主要な改定内容として、AIシステム構築の実態を踏まえた記載修正を行い、それに伴い取決め事項#7「生成AIシステムに係る成果物の取決め」を追加を実施している

生成AIシステムに係る成果物の取決め追加

取決め事項#7「生成AIシステムに係る成果物の取決め」を追加

契約時の項目	取決め事項	契約に盛り込む条項内容例	補足説明
基本項目	生成AIシステムに係る成果物の取決め	成果物について、成果物の定義、事業者がユーザーに対する成果物の完成義務又は完成条件及びその内容、成果物の権利帰属に関して定める条項	契約で規律の対象となる成果物として、成果物の定義はユーザーのサービス利用目的を十分にカバーできるよう範囲を定めること（※）、事業者が成果物を完成して提供する義務がある場合にユーザーのサービス利用目的に照らして、提供条件（提供時期、頻度、態様その他の条件）や提供する成果物の内容（性質、量、粒度その他の内容）を定めること、ユーザーが成果物に関して知的財産権等一定の権利を取得する場合の権利取得条件（権利移転の対象、対価の有無、ライセンスの有無・内容その他の条件）を定めることが望ましい。 ※システムプロンプトをコンポーネントに含む場合の納品物は、プロンプトのひな形、プロンプト文及び設定パラメータ群（回答の自由度や出力トークンの上限 等）等が考えられるが、成果物の多様な形態を踏まえて、契約書等で成果物を明確に規定し、契約当事者間で成果物のイメージをあらかじめ共有しておく必要がある。

デジタル庁

Digital Agency