



ガバメントAI 源内

ガバメントAI 源内

ガバナンス・セキュリティ・セーフティに関する基準（“源内基準”）
の策定・公表について

2026/10/20 デジタル庁 省庁業務サービスグループ ガバメントAI実装担当（源内班）

源内基準の概要

ガバメント AI 源内
ガバナンス・セキュリティ・セーフティに関する基準
（“源内基準”）



ガバメントAI 源内

第1版

2026年10月
デジタル庁 源内班

1. デジタル庁は、約18万人の政府職員が利用可能な「ガバメントAI源内」を通じて、本年10月から、「AIエージェント」^(注)の利用が可能な環境を提供し、AIによる政府業務の省力化・質の向上をより一層推進

（注） 特定の目標を達成するために、環境を検知し自律的に行動するAIシステム（「AI事業者ガイドライン第1.2版」（2026.3.31総務省・経産省）に基づく）

2. AIエージェントは、AIが自律的な判断・行動を繰り返すため、これまで以上に、利用に際しての規律や注意すべき事項が求められる。このため、源内の利用者（政府職員）、管理者及び開発者ごとに、利用・開発等におけるルールの遵守を徹底する必要

「ガバメントAI 源内 ガバナンス・セキュリティ・セーフティに関する基準」（“源内基準”）を策定・公表

- 源内基準は、政府におけるAIの安全・安心な利活用に向けた取組として、AIエージェントの導入を検討する地方公共団体や民間事業者等にとっても参考となるもの
- 源内のAIエージェント利用環境については、地方公共団体や民間事業者等においても広く活用されるよう、2027年2月を目途に、商用利用が可能なライセンスのもと、「源内OSS(オープンソースソフトウェア)Ver.2.0」として、利用・改変・再配布が可能な形で無償公開する予定

（注） 源内OSS(オープンソースソフトウェア)Ver1.0は2026年4月に公開

「ガバメントAI 源内 ガバナンス・セキュリティ・セーフティに関する基準」(源内基準) 概要

1 総論

1-1 源内基準の策定背景及び位置付け

1-2 源内の基本理念及びガバナンス原則

人間中心のAI利用、AI利用における責任主体の明確化、ガバメントクラウドによる安全なAI利用環境の確保、政府職員の入出力データの保護、政府職員による主体的な源内開発、自律性・代替性・耐遮断性に関するAIソブリンティの確保、AIソブリンティを支える疎結合構成、利用状況の可視化と透明性向上、トークンマネジメントの推進、基盤モデルの主体的な選定及び評価

1-3 源内の具体的な取組

源内が対象とするAIの範囲、政府職員向けAIアプリの内製・展開、行政実務用RAGアプリの受託開発と政府横断展開、高品質な政府共通データセットの整備、政府共通データセットの加工・品質管理及び標準化、国内外の基盤モデルの適切な選択、基盤モデルの評価・選定体制の確立、国産基盤モデルの育成・発展への寄与、源内のオープンソースソフトウェアの公開、他府省庁や地方自治体の業務用AI基盤との連携

1-4 AIエージェント利用環境の基本方針及びガバナンス

AIエージェント利用環境の導入方針、AIエージェントによる行政実務の高度化、政府が主体的に管理できるAIエージェント利用環境、エージェントチャットの位置付け及び機能、源内工房の位置付け及び機能、スキルの共有・再利用基盤(源内ライブラリ)の導入、行政手続等におけるAI利用の統制及び記録

1-5 源内における機密性情報の取扱い

1-6 源内の出力情報に係る行政文書の取扱い

2 利用者の責務及び遵守事項

2-1 利用者とは

2-2 利用者が遵守すべき基本原則

AIの積極的な活用、AIリテラシーの向上と倫理観の保持、AI活用における専門性と判断力の維持、ヒューマン・イン・ザ・ループの確保、ハルシネーションに対する注意、プロンプトインジェクションへの警戒、リスクベースアプローチの採用、高リスクな生成AI利活用の判定、シャドーAIの利用禁止、人物に関する情報等の適正な取扱い、トークンマネジメントの推進

2-3 エージェントチャット

利用者が理解すべきリスク、利用者が遵守すべき事項

2-4 源内工房

源内工房の特性と利用者が理解すべきリスク、利用者が遵守すべき事項

3 管理者の責務及び遵守事項

3-1 管理者とは

3-2 各府省庁等における内部ガバナンス

3-3 各府省庁等への「テナント管理者機能」の提供

3-4 利用状況及び利用費用の把握

3-5 エージェントチャットの利用管理及び安全管理措置

管理者が講ずべき措置、
システムの機能制限及び安全管理措置

3-6 源内工房の利用確認及び安全管理措置

管理者が講ずべき措置、
システムの機能制限及び安全管理措置

4 源内開発者の責務及び遵守事項

4-1 源内開発者の位置付け及び役割

4-2 開発及び運用に用いる環境

開発環境、ステージング環境、本番環境、共通環境

4-3 開発・運用権限の分離及び変更管理

権限及び責任の分離、本番環境の変更承認、デプロイ権限及び自動化、変更内容のレビュー、データセットの変更管理、GCAS管理権限の付与及び管理

4-4 開発・運用上の行動原則

確認すべきガイドライン群、合理的に説明できる技術選定、開発及び運用におけるAIの活用、国内外の最新動向の継続的把握、不確実性の早期低減 等

4-5 源内の開発・運用におけるセキュリティ等の統制

脆弱性への対処、セキュリティ事象への即応、変更の追跡可能性、監査証跡の不可変性、権限の最小化と分離 等

「ガバメントAI 源内 ガバナンス・セキュリティ・セーフティに関する基準」(源内基準) 背景・目的

策定背景

- 国際的なAIガバナンスの形成: G7広島サミットで「広島AIプロセス」を提唱し、安全性・透明性を重視する国際指針・行動規範の策定を主導
- 主要国のリスク管理の動向: EUのAI規則、米国のNIST AIリスク管理フレームワーク等、信頼できるAIの枠組みが各国で整備
- AI法と基本計画の策定: AI法に基づく人工知能基本計画(第Ⅱ期)を閣議決定し、イノベーションとリスク対応の両立、政府のAI積極活用を明
- 重層的なAIガバナンス体制の整備: AI法を基礎に、基本計画・重点計画・調達利活用ガイドラインが連携し、政策から実務までをカバー

ガバメントAI 源内の目的

- ① 政府業務へのAIの組み込みによる業務効率化・プロセス変革
- ② 職員のAI理解深化と政策・投資を牽引する能力形成、府省庁間のナレッジ共有による行政能力強化
- ③ 安全・信頼性に配慮したAI利用の率先実践と、その有用性・課題・適切な利用法の社会への提示
- ④ 国産基盤モデル・AIサービスの評価と改善・高度化
- ⑤ 源内の知見・技術の国際展開による世界のAI活用への貢献

- 源内の企画・開発・提供・運用に当たっては、政府職員による利活用を促進するだけでなく、自律性・代替性・耐遮断性に関するAIソブリシティを確保しつつ、安全性、正確性、信頼性、透明性、可観測性、公平性、説明可能性を最大限確保する必要
- AIの出力や処理結果を無条件に採用するのではなく、政府職員がその内容、根拠、限界及びリスクを理解した上で、必要な確認、修正及び判断を行い、最終的な意思決定とその結果に対する責任を担う必要

源内に関わる各主体が遵守すべき事項を明確化(＝源内基準)

1. 利用者 = 源内を使用する政府職員等
2. 管理者 = 各府省庁におけるAI統括責任者(CAIO)、府省内全体管理組織(PMO)、情報システム管理者等
3. 源内開発者 = デジタル庁で源内の企画・開発・提供・運用に携わる職員

1. 総論

➤ AIの積極的な活用

- AIを過度に忌避せず、業務の目的・重要度・リスクに応じて日常業務で積極的に活用
- 調査・資料作成・データ分析やエビデンスに基づく意思決定支援に活用し、新たな業務・プロセスへの適用も積極的に試行

➤ AIリテラシーの向上と倫理観の保持

- 「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」に基づき、源内を積極的に利用し、AIを業務改革に最大限活用するための知識・能力を習得
- 社会の模範として、AIの特性・限界・リスクを正しく理解する最新の知識を継続的に習得し、高い倫理観を保持

➤ AI活用における専門性と判断力の維持

- AIは能力の代替ではなく専門性を拡張する手段として利用し、過度な依存を回避
- 自ら考え・判断・検証する能力と、AIの出力を評価できる専門性(目利きの能力)を維持し、AIが使えない場合も重要業務を遂行できる能力を保持
- 若手職員は、AIを活用しつつ基礎的な知識・技能・思考力・判断力を徹底的に習得

➤ ヒューマン・イン・ザ・ループ(Human in the loop)の確保

- AIの処理を鵜呑みにせず、業務プロセスの適切な段階で確認・判断・承認・修正・監督・停止を実施
- 職員が確認していないAIの出力を、行政資料や公文書に直接転記しない

➤ ハルシネーションに対する注意

- 生成AIの出力を事実として扱わず、重要な情報は原典・法令・公表資料等の一次資料で確認
- ハルシネーションを前提に常に批判的に検証し、計算結果は別の手段で検算

➤ プロンプトインジェクションへの警戒

- 外部情報や添付ファイルを無条件に信頼せず、AIの実行内容・出力結果を確認し、意図しない動作を認識した場合は処理を停止

➤ リスクベースアプローチの採用

- 「AI調達・利活用ガイドライン」を踏まえ、AIに委ねる業務を適切に選択し、特にAIエージェントはリスクベースアプローチで対応

➤ 高リスクな生成AI利活用の判定

- ガイドラインの高リスク判定の考え方及び「高リスク判定シート」を参考に、業務の性格等を総合的に勘案してリスクレベルを判断
- 高リスクの可能性が高い場合は、目的・利用方法・想定リスク・軽減策・品質確保策を整理し、CAIOへの報告等の必要な手続を実施

➤ シャドーAIの利用禁止

- 業務では源内など組織が認めた生成AIサービスを活用
- 利便性や機能性を理由に、組織の管理・統制が及ばない「シャドーAI」を利用しない

➤ 人物に関する情報等の適正な取扱い

- AIによる個人の検索は担当事務の遂行に必要な場合に限定し、他目的に転用しない
- 本人が表明していない信条等の属性推定や将来の行動予測に使用せず、複数の情報源を統合した個人単位での蓄積も行わない

➤ トークンマネジメントの推進

- 入力・出力の双方に従量課金が発生することを認識し、コストを意識してトークンを効率的に管理・活用
- 冗長な指示や不要な文書の貼り付けを避けて簡潔なプロンプトを設計し、無用な大量出力・繰り返しの試行を回避
- 処理の難易度や求める精度に応じ、過剰に高性能でない適切な基盤モデルを選択

利用者(政府職員)が理解すべきリスクと遵守事項(抜粋)

➤ 源内では新たにAIエージェント利用環境として、「エージェントチャット」と「源内工房」を開発。源内基準では、これらの機能の利用に際して、利用者が理解すべきリスクと遵守事項を規定

	エージェントチャット	源内工房
理解すべきリスク	✓ ファイル参照・Web検索・複数段階処理が可能な一方、<u>不適切な判断・指示の影響が後続処理に連鎖するおそれ</u> ✓ 誤情報の生成、意図しない処理、機密・個人情報の不適切な参照、偏った出力等が想定 ✓ 未確認の結果を行政上の判断・通知・外部送信等に用いると、<u>国民の権利利益や行政の公平性・説明責任に影響</u> ✓ ファイル操作機能により、誤指示による<u>意図しない削除・改変・上書き等のおそれ</u> ✓ 重要ファイルは事前バックアップ、範囲限定、実行内容の確認、操作後の検証を実施	✓ 網羅的・多段階な情報収集やプロトタイプ開発を利用者自身が行えるAIコーディングエージェント環境 ✓ 調査から試作まで一つの環境で完結でき、<u>通常のAIチャットやエージェントチャットより操作の自由度が高い</u> ✓ コードの誤り、意図しない処理、ファイルの誤削除・改変、脆弱なアプリ作成、著作権・ライセンス抵触等のおそれ ✓ Webへの直接アクセスや数十～数百ステップの連続実行が可能なため、<u>サプライチェーン汚染等による意図しない情報流出のおそれ</u> ✓ <u>内容を理解しないまま実行すると、問題の影響範囲や原因を把握できないおそれ</u>
遵守すべき事項	✓ 過信せず、<u>目的・指示・参照情報・処理範囲を明確にして利用</u> ✓ 入力・参照情報は機密性・個人情報・権利等を確認し、必要な範囲に限定 ✓ 結果は根拠・数値・法令・宛先等を自ら確認し、<u>重要な判断・外部発信は確認・承認を経る</u> ✓ ファイル操作時は対象・影響範囲を事前確認し、<u>原則バックアップを確保</u> ✓ 国民の権利利益に影響する業務では、<u>リスクベースアプローチを採用</u> ✓ 利用前に注意事項をポップアップ表示し、<u>同意した場合のみ利用可能とする</u>	✓ <u>高い自由度と自律性を理解し、自らの指示とAIの処理を確認する責任を負う</u> ✓ 目的・対象業務・使用情報・外部接続・処理範囲を明確にし、実行・公開前に内容・動作・安全性を確認 ✓ 理解できない処理は実行せず、<u>異常を認識した場合は直ちに停止し管理者へ報告</u> ✓ 国民の権利利益に影響する業務での対応や利用制限は、エージェントチャットと同様 ✓ ガイドライン等に基づく対応に加え、<u>リテラシーチェック等で一定水準を満たし、管理者が確認した職員に限り利用可能とする</u>

ガバメントAI源内に係る政府方針(抜粋)

◆ 第Ⅱ期人工知能(AI)基本計画(2026年7月14日閣議決定)(抄)

第3章 AI関連技術の研究開発及び活用の推進に関し、政府が総合的かつ計画的に講ずべき施策

第1節 AI利活用の加速的推進 (前略)

社会全体でAIを利活用するに当たり、「隗より始めよ」の観点から、政府自らが積極的かつ先導的に利活用する。本府省庁を中心とする政府職員に対し、生成AI利用環境「源内」が展開された。特定の事業者や基盤モデルに過度に依存することなく、安全・安心に選択、運用できる「AI主権」を体現するものとして展開し、地方支分部局への拡大を含めて、政府職員による創造的かつ自律的なAI利活用を広く促進する。

これを起点に、政府において、AIを前提に業務や組織を再構築するAXを強力に推進する。あわせて、政府による適正な調達・利活用を先導的に行うことで、日本社会で利活用されるAIの信頼性及び透明性の確保につなげる。

【具体的な取組】

(1) 政府・自治体でのAIの徹底した利活用

- ① 政府業務における生成AIについて、エージェント型AIの導入や国内で開発されたAIモデルの積極活用も含めたガバメントAI「源内」を推進し、戦略的な利活用を進めるとともに、AIエージェントに対応するルール作りを進め、適切な利活用並びに業務の質の向上及び効率化を推進する。「源内」のオープンソース化の取組を通じて自治体のAI導入支援を進める。【◎デジタル庁、全省庁】

◆ デジタル社会の実現に向けた重点計画(2026年7月21日閣議決定)(抄)

4. 取組の方向性と重点的な取組

(1) 国のAX/DX基盤の高度化・強靱化

② 政府におけるAX/DX推進

ア「ガバメントAI(源内)」の推進

- 社会全体でのAI利活用の加速的推進に資するため、政府自らが生成AIを積極的かつ先導的に利活用する。令和7年度にデジタル庁が内製開発した生成AI利用環境「源内」について、令和8年度は全府省庁対象の実証事業を着実に推進し、行政業務におけるAI利活用による費用対効果、他の商用LLMとの性能上の比較、国産AIモデルの積極活用を通じた我が国のAI基盤の自律性確保への貢献などを検証した上で、令和9年度以降の本格利用開始を検討する。

- 政府職員において、専門的知識がなくともAIアプリを作成し、AIエージェントによる自動実行や共有が可能となる環境を令和8年度中に源内上に整備するとともに、利用に関する管理ルールを策定する。

デジタル庁

Digital Agency