先進的ＡＩ利活用アドバイザリーボード
第３回ワーキング・グループ実施報告

2026/9/29　　先進的ＡＩ利活用アドバイザリーボード事務局

第３回ワーキング・グループの実施概要

【日時及び参加者】

実施日時　：2026年9月8日（火）（オンライン開催）

実施対象　：ガバメントAI源内における「エージェントチャット」「源内工房」機能の実装及びガバナンス・セキュリティ・セーフティに関する基準の策定

参加省庁　：デジタル庁

参加構成員：湯淺構成員、北村構成員、柴山構成員

【相談の背景】

■ デジタル庁では、本年度（令和8年度）中に、全府省庁等約18万人がガバメントAI源内を通じた生成AIの利用が可能となるよう、各府省庁等への源内環境の展開（大規模実証事業）を進めている（8月下旬時点：約16万人が利用可能）

■ 源内を通じた生成AIの機能をより一層発展させるため、今後（9月末～10月以降順次）、各府省庁等の源内環境において、**「AIエージェントチャット」及び「源内工房」の利用が可能な環境を試験提供予定**。同環境では、AIが自律的な判断・行動をするため、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とから、**新たに必要なルール（＝源内基準）を定め、源内の利用者、管理者及び開発者ごとに、ルール遵守を徹底する必要**

■ これらエージェントの自律性等の特性を踏まえ、**エージェントチャット及び源内工房が提供されることによる統制上の課題やセキュリティ上の安全対策等について過不足がないか各府省庁への正式実装前に改めて検討、確認**しておくことが必要。このような観点に基づき、本WGでは主に以下の内容で議論を実施した。

【本WGの議論内容】
①AIエージェントの実装に向けた安全性の確保について
②業務での利活用・展開の拡大について
③エージェント全般の利用におけるリスク管理・人の介入・関与について

①源内のエージェントチャット （AIエージェント利用環境、プロトタイプ画面）

- 通常のチャット（対話型AI）の発展形として位置づけ
- AIエージェント利用環境として自然言語による指示が可能（専門的知識は不要）
- 複数のタスクをまとめて自動実行できる
- AIエージェントのスキルを作成可能
- 処理結果は、ファイルやドキュメントとして出力できる

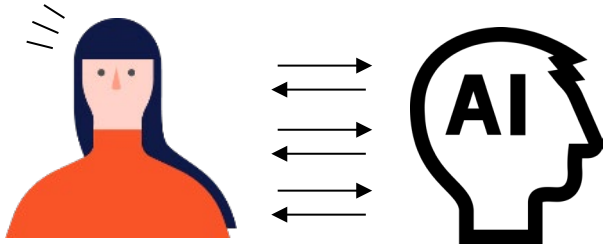
②源内工房（個人源内）のイメージ （Claude Code相当のAIエージェント利用環境、プロンプト画面）

- 自然言語による操作が可能（専門的知識は不要）
- 任意サイトへの直接アクセスが可能。複数のWebサイトを直接読み取りながら情報を収集・比較・統合するDeep Researchができる
- AIエージェントがファイル操作、コマンド実行、Web取得等を数十から数百ステップ連鎖させて処理できる
- 実際に動作するプロトタイプの開発ができる
- AIエージェントのスキルを作成可能
- 処理結果は、ファイルやドキュメントとして出力できる
- 他人が作成したスキルをカスタマイズ可能

AIエージェントの利用により政府全体の能力を強化するイメージ

マーケットプレイスを通じてスキルの共有・改良ができるようにする。
誰でも他人のAIエージェントを再利用できる。

政府職員



①バイブコーディング環境

業務プロセスの「暗黙知」

職員一人ひとりが
長年にわたり培ってきた
業務経験・専門知識

スキルを作成することで、これまで属人的に行われてきた部分のワークフローの可視化・形式知化が実現する。

「形式知」

編集・保存・共有できる
スキルとして表現

②AIエージェント利用環境

同僚・後任も実行可能

③ガバメントAIマーケットプレイス

「集合知」

マーケットプレイスを通じてスキルを共有。
他省庁職員はダウンロードしてスキルを再利用

②AIエージェント利用環境

他省庁職員も実行や改良が可能

議論概要

論点①　AIエージェントの実装に向けた安全性の確保について

セキュリティ対策

■ 外部ファイルをアップロード等する際のマルウェア対策は、クラウド基盤が提供するセキュリティチェックを利用しているのか。AIエージェントの権限に関わるMCPやAPI連携についての統制はどうなっているのか。

▸源内工房でもマルウェア対策を導入予定。PaaS上の設定で独自にマルウェア対策を施しており、アップロードファイルの自動スキャン機能も導入予定である。また、情報漏洩についても意識しており、MCPやAPI連携も外部への送信規制として制限され、無断で行うことはできない。危険性が大きいと思われるレポジトリ等については、DLできないという制御も行っている

ガバメントＡＩマーケットプレイスにおける成果物の共有

■ 源内工房でユーザが開発したスキルやコードを実環境にインストールするのは注意が必要であり、作成した成果物を他人に提供する際には、事務局による審査が重要と考える。

▸職員間でスキルを共有することを可能とする「ガバメントＡＩマーケットプレイス」についても年内にリリースする予定であり、その共有過程において内容を審査する仕組みを検討中である
▸スキル以外で、業務環境へのインストールが必要となるものの場合はその業務環境におけるアプリケーションの導入審査手続で、内製開発されるシステムの一部として使用されるものの場合は標準ガイドライン等の開発に係る規定に基づき確認される

基盤モデルのバージョンアップへの対応

■ 基盤モデルがバージョンアップされると不具合が起きる恐れがある。現在は過渡期であるため、手動でのモデルアップデートは重要であると考える。また、リリースノートやインターフェースを継続的に整備していくべき

▸基盤モデルの導入は手動で設定しており、自動更新で不具合が起きることはない、リリースノート・インターフェースの整備も継続して実施する

議論概要

論点②　業務での利活用・展開の拡大について

適用業務の拡大及び利便性向上

■ 民間向けに検討中のエージェント利用のガイドラインでも、①データの制限（環境変数の待避や利用データの確認等）、②権限の設定（ＡＩから実行承認が来た際にその内容を確認のうえで承認する運用）、③連携ツール（連携ツール利用時にリアルタイムで動作を確認する）の３つの軸を定めており、源内基準の方向性と概ね一致している。

■ 個人のプロファイリングをしてはいけないということだが、行政でもクリアランスに関する業務は増加しており、ＡＩの導入ニーズが出てくると思う。どのような目的でプロファイリングを実施しているか把握ができれば導入可としても良いのではないか。

▶プロファイリングが正当な業務であり、その目的が適切に把握できる前提であれば、担当する事務の遂行に必要な場合に限り行うことはあり得るのではないか。他方で、源内が提供するサービスを使うだけではなく、ＯＳＳに基づき独自にシステムを構築する例なども考えられるところ、少なくともこの規定（源内基準）がブロッカーになることはないと考える。

■ 操作画面での注意喚起・プロンプト集・テーブルトップ演習といった利用支援が有効であり、トークンを使い切る課題の解消にもつながると考える。

▶（ユーザリテラシーの向上は勿論のこと）ログやモデルの挙動を記録するなど、インテリジェント基盤の開発について来年度にかけて実施可否を検討中。今後一層必要になるのはトークンマネジメントの観点。アプリの利用状況の可視化に加え、モデル選択の最適化、セキュリティインシデントの可視化も検討。継続的にリファクタリングできればユーザにもメリットがあり、相互に費用を安価にするとともに、最短距離で業務の実績を上げることができると考える。

議論概要

論点③　エージェント全般の利用におけるリスク管理・人の介入・関与について

リスク判定と推論プロセス評価

- ■ 出力結果の適切さの判断では、推論プロセスも参考にはなるが、最終的には出力結果そのものの判断で一定の担保が可能であると考える。
- ■ AIエージェントは、他のAIエージェントを含め、どこまで接続するか予見できず、かつ、権限を逸脱するリスクがある。意図した通りの動作をしているかどうかをチェックするためには、中間動作と結果の両方を確認するなどのプロセスアプローチが必要ではないか。
- ■ 法的なリスクと技術的なリスクは観点が異なる。前者については行政裁量の大きさにも関わるため、個々の業務特性に応じて個別に考えるべき
- ■ AIエージェントのリスクが高いことを理解している人や研修を受けている人のみを利用可にするなど、確認方法を簡素化すべき。

▶高リスク判定シートのリスク軸について、エージェント利用時は最終結果を人が判断すれば高リスクの可能性が低いと考えてよいか、あるいは複数の自律的な動作プロセスまで踏まえて判断すべきか。源内は基本的には行政の内部事務など、国民等への処分性がない業務での利用を前提としている。最大18万人が使うツールであるため、運用の簡素化にも対応したい。

人の関与のあり方

- ■ 処分性のある行為については人間による判断の介在(ヒューマンインザループ(HITL))は必要。内部的行為・事実的行為は人が介在しないといったことはあり得るのではないか。フィジカルAIの利用では、個々の動作に人による判断を介在させることは難しい。今後ＨＩＴＬを必須としない、ヒューマンオンザループの理念に基づいたエージェント利用も議論いただきたい
- ■ 行政の処分性のある行為は私人の権利変動を伴うため、双方の合意に基づく民間事例と観点が異なることには注意が必要だが、AI利活用のバリアとしないよう、今後議論が必要ではないか

▶現在の行政手続は最終的に行政処分に至る業務が多く、職員のリテラシー向上の観点からも、当面は原則として人の確認を必須とし、将来的な課題とさせていただきたい。

今後の方向性

- ➢ 源内工房では、MCP/API連携の制限は既に導入済、マルウェア対策は展開までに実装。
- ➢ スキルの共有については、年内展開予定のガバメントAIマーケットプレイスにおいて審査の仕組みを導入。
- ➢ トークンマネジメントやモデル選択の最適化を可能とするインテリジェント基盤の実装を検討。

参考資料

参考資料

ガバメントAI 源内 とは

【背景】

- **人工知能（AI）は、生成AIを筆頭に世界各国で急速な技術進歩が進行中**
- **国際的なAI競争の激化**：米国・中国・英国が昨年相次いで国家AI戦略を発表し、政府自身でのAI積極活用を推進する方針
- **少子高齢化による公務員の担い手不足**：行政の担い手が減少する中、質の高い行政サービスを持続的に提供するにはAI活用が必須
- **AIによる行政の業務プロセス見直し、公務員の働き方改革の実現**
- **自律性の確保**：AIにおける日本の自律性を維持し、日本独自の価値創出

【目的】

1. **業務変革の実現**：政府職員の日常業務にAIを組み込み、定着させることで、業務の質向上と効率化を実現。創造的・戦略的な業務へのシフトを可能にする。
2. **政策領域へのAI展開の先導**：職員自身がAIへの理解と活用経験を深めることで、担当する政策分野や所管業界におけるAI導入を牽引できるようにする。
3. **信頼性の高いAIの価値提示**：政府が率先して信頼できるAIを活用し、その有用性と意義を国民に示す。
4. **国内AI産業の成長促進**：政府の取組を通じて国産AIの育成を図るとともに、AIへの投資意欲を喚起

■高市総理大臣指示（人工知能戦略本部（2025.12.19））

“信頼できるAIによる日本再起を実現するため、７点について指示します。第一に、『ガバメントAI源内』の徹底活用です。来年５月から１０万人以上の政府の職員が活用できるようにします。源内の活用により、創造的に業務を行い、国民の皆様に信頼できるAIの意義を示してください。・・・”

■人工知能基本計画（2025.12.23閣議決定）

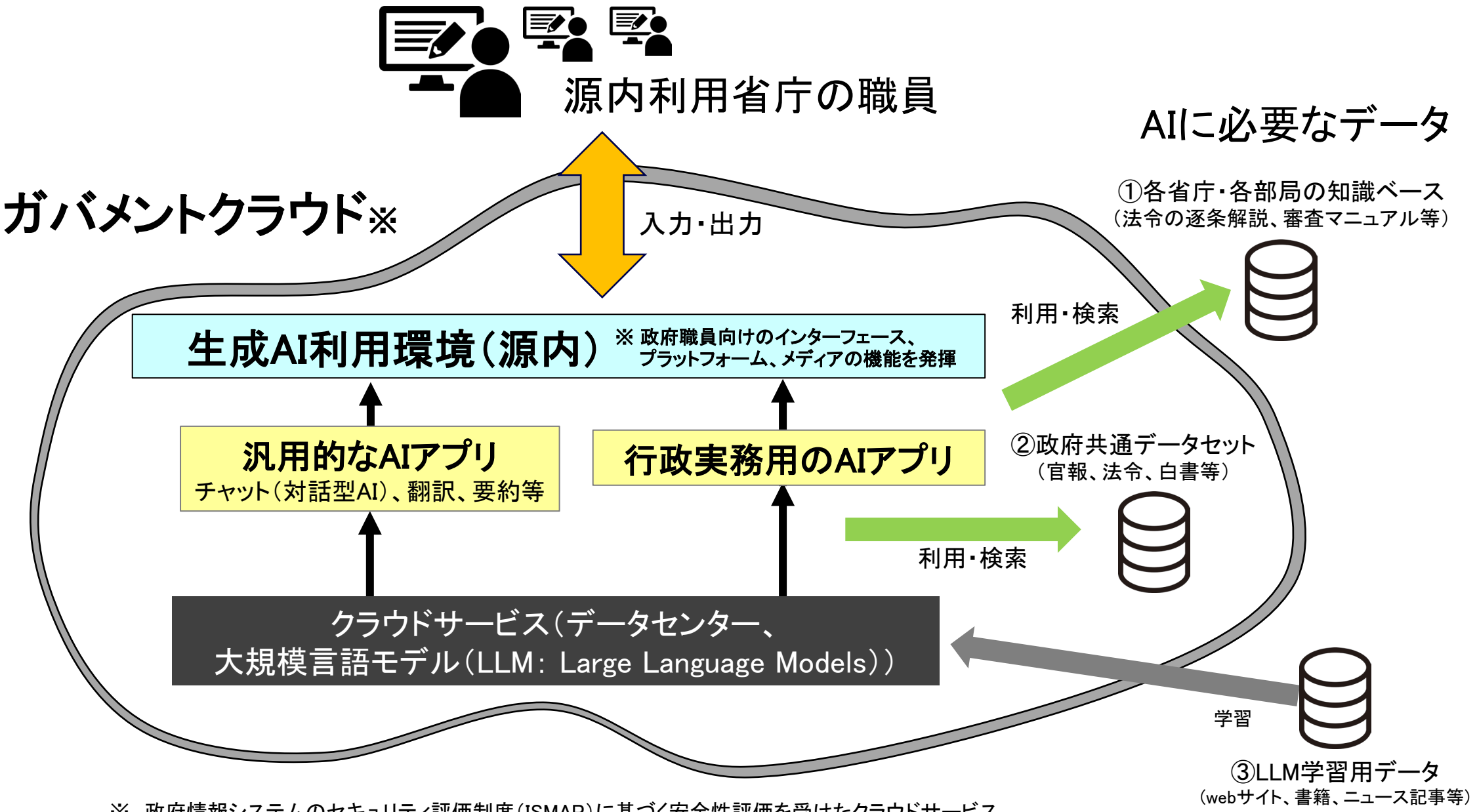
我が国でのAI利活用を促進するため、「隗より始めよ」の観点から、まずは政府自らが積極的かつ先導的に利活用する。政府職員によるAIの普段使いを浸透、定着させることにより、業務の質を向上させる。将来的には地方支分部局を含む中央省庁の全職員が業務の質の向上を実感できる環境の構築を目指すこととし、速やかに本府省庁職員が生成AIを利活用できる環境を構築する。その際、指定職・管理職による率先した利活用を促す仕組みを導入する。

ガバメントAI 源内（げんない） Generative AI

- 40種類以上のAIアプリケーションを使用できる生成AI利用環境。機密性2情報まで入力可能（各府省庁がルールを設定）
- 本年5月より、全府省庁18万人の政府職員が生成AIを利用できる環境を展開
  - 職員に対する導入支援（カスタマーサポート）及び利活用促進（カスタマーサクセス）を実施中
- 国会答弁作成支援AIを開発中。国産の大規模言語モデル（LLM）を5社選定し、源内を通じて試験的に提供予定
- 政府共通データセットを調達し、源内を通じて提供（官報79年10か月分（昭和22年5月～）等）

参考資料

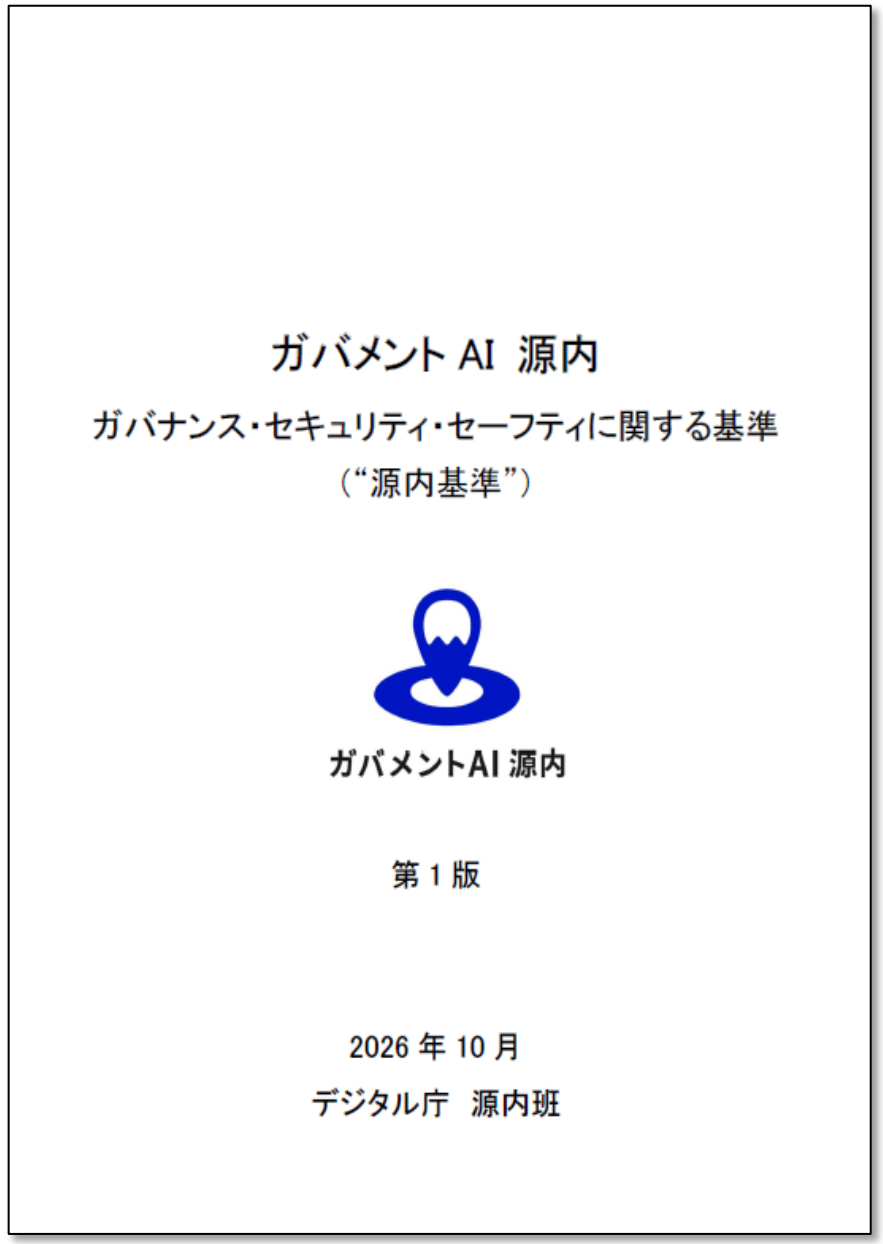
政府における生成AI利用環境（源内）の概要



※　政府情報システムのセキュリティ評価制度（ISMAP）に基づく安全性評価を受けたクラウドサービス

参考資料

“源内基準”の策定・公表について（案）


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

◆ デジタル庁では、**本年10月上旬に「ガバメントAI 源内」の利用に関するガバナンス・セキュリティ・セーフティの基準（通称「源内基準」）の策定・公表**を予定

【背景・目的等】

- デジタル庁では、本年度（令和8年度）中に、全府省庁等約18万人がガバメントAI源内を通じた生成AIの利用が可能となるよう、各府省庁等への源内環境の展開（大規模実証事業）を進めているところ（8月下旬時点：約16万人が利用可能）

- 源内を通じた生成AIの機能をより一層発展させるため、今後（9月末～10月以降順次）、各府省庁等の源内環境において、**「AIエージェント」の利用が可能な環境の試験提供を予定**。同環境では、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とから、**新たに必要なルール（＝源内基準）を定め、源内の利用者、管理者及び開発者ごとに、ルールの遵守を徹底する必要**

- 加えて、源内基準においては、「AIの積極的な活用」のほか、広島AIプロセスや事業者ガイドラインなど既存の規範等も踏まえつつ、「AIリテラシーの向上と倫理観の保持」「AI利用によって自らの専門能力を失わないこと」など、**国家公務員の源内（生成AI）への向き合い方についての方針についても記載**

（注）政府職員を対象としたAIエージェント利用環境の大規模導入は、米英など主要国においても他に類例がなく、世界的にも最先端の取組となることから、今後AIエージェントの導入を検討する自治体や民間企業が参照できるモデルケースとして、本基準を国内及び海外に向けて広く公表・周知するとともに、継続的にその内容を見直す予定

上記を踏まえ、今後、各府省庁等における源内の管理者・利用者を対象に、**別途、源内活用ガイドver1.0（2026年3月4日）の改定・周知、管理者に対するエージェント管理手順等の周知・依頼のほか、源内UI上で利用に際しての注意事項に係るポップアップ表示**等を実施予定

参考資料

「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

策定背景

- **国際的なAIガバナンスの形成**：日本はG7広島サミットを契機に「広島AIプロセス」を提唱し、生成AIの国際指針・行動規範の策定を主導。安全性・透明性・アカウンタビリティを重視する国際的枠組みの形成に貢献。
- **主要国のリスク管理の動向**：EUのAI規則やNISTのAIリスク管理フレームワーク等、各国で安全で信頼できるAI利活用を支える枠組みが整備
- **AI法と基本計画の策定**：AI法に基づき人工知能基本計画（第Ⅱ期）が閣議決定され、イノベーション促進とリスク対応の両立を原則に、政府自らがAIを積極活用する方針と具体的取組が決定。
- **重層的なAIガバナンス体制の整備**：AI法を基礎に、AI基本計画・デジタル社会重点計画・生成AI調達利活用ガイドラインが連携し、政策方針から行政実務ルールまでをカバーする重層的な枠組みが構築

ガバメントAI 源内の目的

① 政府職員の日常業務にAIを組み込み、業務の効率化及び業務プロセスの変革を促進

② 政府職員がAIの特性、能力及び限界を源内の利用を通じて深く理解し、それぞれが担う所掌事務、政策・制度及び行政分野におけるAI活用及び民間AI投資を牽引できる能力を形成。各府省庁におけるAI活用の知見、ベストプラクティス及びデータを共有し、政府全体としてのナレッジ基盤の構築と行政能力を強化

③ 政府が安全性及び信頼性に配慮したAI利用を源内において率先して実践し、その有用性・意義、課題及び適切な利用方法を社会に提示

④ 国内で開発された基盤モデル、AIサービス及び関連技術を源内において適切に評価し、改善及び高度化

⑤ 源内の知見・技術を国際展開し、世界のAI活用に貢献

■ 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ソブリンティを確保しつつ、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**

■ さらに、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るべき事項を明確化（＝“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

参考資料

「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」）　全体構造

1　総 論

1－1　源内基準の策定背景及び位置付け
背景、ガバメントAI 源内の取組、源内基準の位置付け

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、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、
源内の価値向上を実現する管理基盤、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（源内マーケットプレイス）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

2－3　エージェントチャット
利用上のリスク、利用上の注意

2－4　源内工房
利用上のリスク、利用上の注意

3　管理者の責務及び遵守事項

3－1　管理者とは

3－2　各府省庁における内部ガバナンス

3－3　各府省庁への「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　源内工房の利用許可及び安全管理措置
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、秘密情報の管理、本番環境の保護、バックアップ及び復旧 等

デジタル庁

Digital Agency