

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

目次

1 総論	1
1-1 源内基準の策定背景及び位置付け	1
1-2 源内の基本理念及びガバナンス原則	4
1-3 源内の具体的な取組	8
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	14
1-5 源内における機密性情報等の取扱い	19
1-6 源内の出力情報に係る行政文書の取扱い	20
2 利用者の責務及び遵守事項	21
2-1 利用者とは	21
2-2 利用者が遵守すべき基本原則	21
2-3 エージェントチャット	24
2-4 源内工房	26
3 管理者の責務及び遵守事項	28
3-1 管理者とは	28
3-2 各府省庁等における内部ガバナンス	28
3-3 各府省庁等への「テナント管理者機能」の提供	28
3-4 利用状況及び利用費用の把握	29
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	29
3-6 源内工房の利用確認及び安全管理措置	30
4 源内開発者の責務及び遵守事項	32
4-1 源内開発者の位置付け及び役割	32
4-2 開発及び運用に用いる環境	32
4-3 開発・運用権限の分離及び変更管理	34
4-4 開発・運用上の行動原則	35
4-5 源内の開発・運用におけるセキュリティ等の統制	37

別紙

改訂履歴

1 総論

1-1 源内基準の策定背景及び位置付け

(背景)

近年、生成 AI を含む人工知能技術は急速に進展し、その社会実装が世界各国で拡大している。これに伴い、AI の利活用を促進しつつ、その利用に伴うリスクを適切に管理するための制度や枠組みの整備が進められている。

我が国は、AI の安全性・信頼性の確保に関する国際的なルール形成を早期から主導してきた。特に、令和 5 年の G7 広島サミットを契機として、我が国が提唱した「**広島 AI プロセス**」においては、生成 AI を含む高度な AI システムの開発・利用に関する国際指針及び国際行動規範が G7 で取りまとめられた。その後も我が国の主導により、広島 AI プロセス・フレンズグループへの参加国・地域の裾野が着実に広がるとともに、OECD と連携した報告枠組みの整備等、その取組は国際的に広がっている。こうした取組は、AI の安全性、セキュリティ、透明性、アカウンタビリティ等を重視する国際的な AI ガバナンスの形成に重要な役割を果たしている。

主要国においては、例えば、欧州連合(EU)における AI 規則(AI Act)や、米国国立標準技術研究所(NIST)が策定した AI Risk Management Framework(AI RMF 1.0)等を通じて、AI の開発及び利用に関するリスク管理の考え方が示されている。これらは、その法的性質や対象範囲を異にするものの、安全で信頼できる AI の開発及び利活用を支える枠組みとして位置付けられる。

我が国では、**人工知能関連技術の研究開発及び活用の推進に関する法律**(令和7年法律第 53 号。以下「AI 法」という。)が制定され、AI 法に基づく我が国初の**人工知能基本計画**が令和7年 12 月 23 日に閣議決定された。同計画では、「イノベーション促進とリスク対応の両立」、「アジャイルな対応」及び「内外一体での政策推進」を基本原則とし、政府自らが「隗より始めよ」の考え方の下で AI を積極的かつ先導的に利活用する方針が示された。

その後、人工知能技術及びこれを取り巻く国内外の環境の急速な変化を踏まえ、令和8年 7 月 14 日には、**第Ⅱ期人工知能基本計画**が閣議決定された。同計画では、第Ⅰ期の人工知能基本計画に掲げられた基本原則を引き継ぎつつ、政府における AI の積極的な利活用、ガバメント AI の推進、AI の安全性及び信頼性の確保、AI に関する自律性の向上並びに国内の AI 産業基盤の強化等に関する取組が一層具体化されている。

また、令和8年 7 月 21 日に閣議決定された「**デジタル社会の実現に向けた重点計画**」においても、急激な人口減少等の社会課題に対応し、質の高い行政サービスを持続的に提供するため、AI を活用した行政の変革を進める方針が示されている。特に、政府共通の AI 利用環境であるガバメント AI の整備及び利用を進め、政府職員の日常業務における AI の活用を拡大するとともに、業務プロセスの見直し及び行政サービスの高度化につなげることとされている。

これらの政府方針に基づき、行政実務における具体的な運用規範として、「**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)**」(デジタル社会推進標準ガイドライン DS-920。令和 8 年 6 月 12 日デジタル社会推進会議幹事会決定。以下「AI 調達・利活用ガイドライン」という。)が策定され

ている。同ガイドラインは、利活用促進とリスク管理を表裏一体で進めるため、政府全体の AI ガバナンス並びに各府省庁等における調達及び利活用時の基本的な規律を定めるものである。

このように、我が国では、AI 法を基礎として、第Ⅱ期人工知能基本計画及び「デジタル社会の実現に向けた重点計画」により政府全体の政策方針を示すとともに、AI 調達・利活用ガイドラインにより政府内における行政実務上のルールを具体化するという、重層的な AI ガバナンスの枠組みが整備されている。

（ガバメント AI 源内の取組）

我が国では、少子高齢化及び生産年齢人口の減少が進行する中、行政を担う人材の確保が一層困難になることが見込まれている。このような状況において、限られた人的資源の下で質の高い行政機能・能力を持続的に提供するためには、デジタル技術及び AI を積極的に活用し、従来の業務プロセスを抜本的に見直すことが必要である。

また、AI は、産業構造や経済社会の在り方を大きく変革し、将来の国力や国家間の競争力を左右し得るゲームチェンジャーとして、産業競争力及び経済安全保障上、極めて重要な基盤技術となっている。各国が AI の研究開発、社会実装及び産業基盤の強化を国家戦略として進める中、我が国においても、AI に関する技術、データ、計算資源、人材及び利用能力を戦略的に確保・強化するとともに、特定の国、事業者又は技術に過度に依存することなく、主体的な選択可能性及び継続的な利用能力を確保することが重要となっている。

こうした背景を踏まえ、ガバメント AI 源内（以下「源内」という。）は、政府職員による安全・安心な AI 利用を確保しつつ、その能力を十分に引き出し、行政業務の効率化、品質向上その他の利用効果及び費用対効果を最大化するため、デジタル庁が企画・開発・提供・運用する生成 AI 利用環境であり、主として次の目的を有する。

- 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する。
- 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する。各府省庁等における 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の強化を図る。
- 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示す。
- 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、改善及び高度化につなげる。
- ⑤ 源内の知見・技術を国際展開し、世界の AI 活用に貢献する。

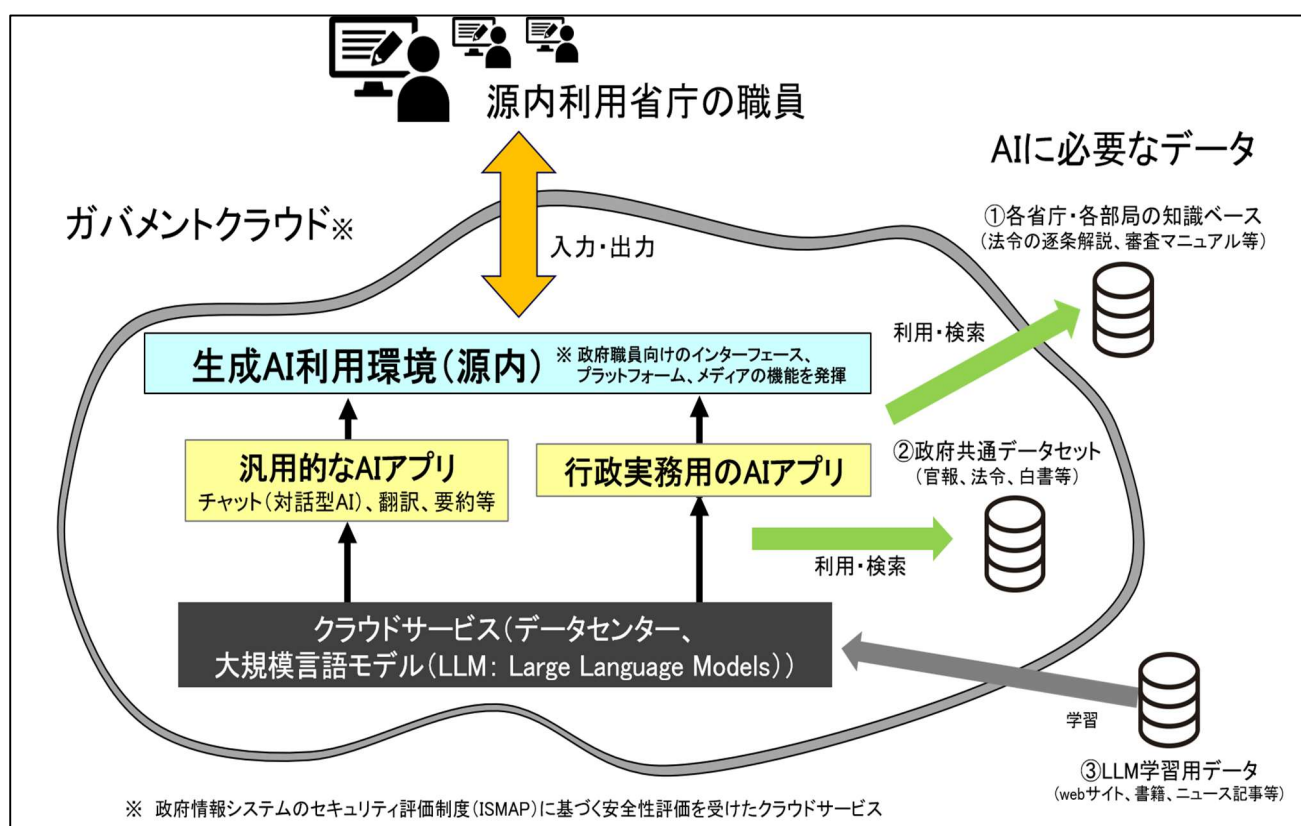
源内は、各府省庁等が定める条件の下で、機密性2情報を取り扱うことが可能なセキュアな環境において、約 40 種類の AI アプリケーション（以下「AI アプリ」という。）を提供している。チャット、翻訳、文書作成、画像生成等の汎用的な機能に加え、法令・制度調査、国会答弁作成支援、公用文校正、補助金制度調査等、行政実務に対応した AI アプリを開発、提供している。

入力することができる情報の範囲及び利用条件は、利用環境、接続するモデル、アプリケーションの機能

並びに各府省庁等のルールに応じて適切に定められる必要がある。デジタル庁では、令和7年5月以降の庁内利用を通じて得られた運用実績及び知見を踏まえつつ、「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群」（令和7年6月27日サイバーセキュリティ戦略本部決定。以下「政府統一基準群」という。）等に基づくセキュリティ対策を講じ、機密性2情報を取り扱うことができる利用環境と位置付けている。

令和8年5月には、全府省庁の約18万人の政府職員を対象とする大規模実証を開始した。この実証を通じて、政府職員の日常業務におけるAI利用の浸透及び定着を図るとともに、利用効果、安全性、ユーザビリティ、運用上の課題等について検証を進めている。今後は、AIが一定の手順に基づき業務を自動実行する「AIエージェント」の利用環境について段階的に整備を進め、令和9年度以降は、各府省庁等が予算措置を講じた上での本格的な利用へと移行することとしている。

図 ガバメントAI 源内のシステムイメージ



（「源内基準」の位置付け）

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(AI 主権)を確保しつつ、安全性、正確性、信頼性、透明性、可観測性、公平性及び説明可能性を最大限に確保する。

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ことを基本とする。そのため、業務の性質や影響の大きさに応じて、政府職員による監督、承認、介入及び利用停止を可能とする仕組みを設け、AI が政府職員の判断を代替するのではなく、職員の能力と意思決定を支援する、人間中心の利用及び運用を確保する。

以上を踏まえ、源内に関わる各主体が遵守すべき事項を整理したものが、本「源内基準」である。

源内基準は、AI 法、人工知能基本計画、「デジタル社会の実現に向けた重点計画」、AI 調達・利活用ガイドライン等の政府の法令、計画及びガイドラインとの整合を図るとともに、「広島 AI プロセス包括的政策枠組み」や「AI 事業者ガイドライン」等、国内外の主要な AI ガバナンスに関する枠組み・指針も踏まえている。

源内基準は、源内に関する共通原則を定めた『総論』並びに、源内に関与する主体に応じた責務及び遵守事項で構成される。

源内の利用、管理及び開発に関わる主体は、「利用者」「管理者」及び「源内開発者」に区分。

- 「利用者」とは、源内を使用する政府職員等（国の行政機関において行政事務に従事している国家公務員のほか、源内の利用を希望する立法府及び司法府の職員並びに独立行政法人等の職員）をいう。
- 「管理者」とは、各府省庁等における AI 統括責任者（CAIO）、府省内全体管理組織（PMO:Portfolio Management Office）、情報システム管理者その他これらに準ずる主体をいう。具体的にどの部局又は職員を管理者とするかについては、デジタル庁が一律に指定するのではなく、源内を利用する各府省庁等が明確にする。管理者は、源内の AI エージェント利用環境や基盤モデル等の利用可否について判断するとともに、組織内における適正な利用及び管理を担う。
- 「源内開発者」とは、デジタル庁において源内の企画・開発・提供・運用に携わる職員をいう。デジタル庁がこれらの業務の全部又は一部を外部事業者等に委託する場合は、当該委託業務に実際に従事する当該外部事業者等及びその担当者を含む。

なお、政府における源内の提供形態には、デジタル庁が源内をクラウドサービスとして運営し、各府省庁等に直接提供する「デジタル庁ホスト方式」と、各府省庁等が源内のオープンソースソフトウェア（以下「OSS」という。）等を活用し、自らが整備した環境において、自らの責任で AI 基盤又は AI 利用環境を開発、構築及び運営する「自営方式」の二種類がある。

前者では、デジタル庁が共通基盤・AI アプリ・AI エージェント利用環境の企画・開発・提供・運用を担い、原則として府省庁の職員が同一の源内環境を利用する。後者では、府省庁自身が独自構築した環境の管理、セキュリティ、運用及び利用ルールに責任を負うこととなる。また、「デジタル庁ホスト方式」により源内の本格運用が開始される令和 9 年度以降、自営方式を採用する各府省庁等では、「デジタル庁ホスト方式」により提供される AI アプリ、AI エージェント利用環境、政府共通データセット、基盤モデル及びライブラリに係る機能等を使用できないこととなる。このため、本「源内基準」は、「デジタル庁ホスト方式」により提供される源内を適用対象として策定したものである。

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）

源内における AI の利用は、「人間中心の AI 社会原則」（2019 年 3 月統合イノベーション戦略推進会議決定）が掲げる人間の尊厳、多様性・包摂性及び持続可能性を基本とし、AI 法第 13 条に基づく「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」（令和 7 年 12 月 19 日人工知能戦略本部決

定)、OECD の AI 原則、広島 AI プロセス国際指針その他の国内外の規範と整合的に行う。

AI は、政府職員の能力及び創造性を拡張し、より質の高い行政を実現するための支援手段として位置付けるが、職員の判断・決定を無条件に代替するものとしてはならない。AI を利用する目的、範囲及び条件はその業務を担う政府職員及び管理者が適切に定め、政府職員は、AI の出力の内容、根拠、限界及びリスクを確認した上で、必要な判断又は決定を行う。その結果に対する責任は、政府職員及び行政機関がそれぞれの役割及び権限に応じて負う。

特に、国民の権利利益、安全又は生活に影響を及ぼす業務では、リスクに応じて職員による確認、監督、承認及び介入並びに利用停止を可能とする仕組みを確保するとともに、公平性、透明性、説明責任、プライバシー及び基本的人権を尊重する。

(AI 利用における責任主体の明確化)

源内の利用及び管理並びに企画・開発・提供・運用に当たっては、AI そのものを責任主体として扱うことなく、利用者、管理者、源内開発者その他の関係主体について、それぞれの役割及び権限に応じた責任の所在を明確にする。

利用者は、出力内容の確認、修正及び自らの業務における出力内容の利用の可否の判断とその利用に伴う責任を負い、管理者は、業務における源内の利用可能範囲、源内の利用権限、統制措置及び事故対応を適切に定める。源内開発者は、源内の設計、実装、評価、監視及び継続的な改善を適切に行い、安全性、信頼性及び実効性を確保する。

(ガバメントクラウドによる安全な AI 利用環境の確保)

源内は、デジタル庁が整備する政府共通の生成 AI 利用環境であり、そのシステム基盤として、ガバメントクラウド(デジタル庁が整備・運用する政府共通のクラウドサービスの利用環境であり、政府統一基準群に基づくセキュリティ要件を満たすもの)を専ら使用する。

さらに、源内では、政府職員が安全・安心に AI を利用できるよう、AI の推論処理及び関連データの保存を、ガバメントクラウドを構成する国内のデータセンターで行う。これにより、利用者が入力するプロンプトや添付ファイル、AI が生成する出力データ、業務で使用する AI エージェントのスキル(従来、個々の職員や部局に属人的に蓄積されてきた業務手順、判断基準、参照資料、関連データ、確認方法、出力形式、例外処理等を、AI エージェントが実行可能な形で可視化・形式知化したものをいう。以下同じ。)その他の関連データを国内で取り扱い、データが国外に移転されない構成とする。以上により、公務に必要な情報管理及びデータ主権を確保し、機密性の高い情報を含む行政実務でも安心して利用できる環境を整備する。

(政府職員の入出力データの保護)

源内では、原則として、ガバメントクラウド上の環境で動作し、当該環境内で推論処理を行う基盤モデルのみを利用するものとし、入力データがガバメントクラウド外の推論環境に送信される基盤モデルサービスは利用しない。ただし、例外的に、国内リージョンでは提供されていない高度な基盤モデルを利用する必要がある場合等に限り、海外のデータセンターで推論処理を行う選択肢を設けることとし、この場合にあっては、推論処理の終了後に入力データ及び出力データが海外で保存されないことを調達契約で確保するとともに、源内の利用者及び管理者に対し、その利用上のリスク及び留意事項を明確に説明するものとする。

また、源内で利用する基盤モデルについては、利用者が入力するプロンプト、添付ファイルその他の入力データ及び AI が生成する出力データを、基盤モデルの学習、再学習、追加学習、性能改善その他これらに類する目的に利用しないことを、基盤モデルの提供事業者との契約により確保する。

なお、デジタル庁は、源内の運用管理、機能検証・改善、利用状況の分析、セキュリティ対策その他適切な企画・開発・提供・運用に必要な範囲で、入力・出力データを含むログを取得し、取り扱う場合がある。この場合、あらかじめ定めた利用目的及び権限管理の下で、必要な範囲に限定して取り扱い、目的外の利用を行わない。

（政府職員による主体的な源内開発）

行政固有の業務手順、判断基準、参照資料及びセキュリティ要件を的確に反映し、政府職員の実務に適合した AI 利用環境を迅速かつ継続的に整備するため、デジタル庁職員が源内を主体的に開発する。

源内の開発を担う政府職員は、各府省庁等の職員や行政実務に直接接し、法令、政策、国会、予算、調達等の多様なドメイン知識に容易にアクセスできる。このため、行政実務の実態を熟知する政府自身が行政実務に特化した AI アプリ（行政実務用 AI アプリ）を内製することで、現場ニーズへの即応と継続的改善を図ることができるとともに、利用者とのやりとりを通じて潜在的な課題を把握し、試作、実利用、検証及び改善を短いサイクルで繰り返すことで、利用者が必要とする機能の迅速な実現が可能となる。

これにより、外部事業者への単純な委託と比較して、AI に関する知識、技術、ノウハウ及び開発・運用能力といったナレッジを政府内に継続的に蓄積し、行政組織自らが業務改革を進め、新たな AI 活用を企画・実装できる能力の向上につなげることができる。

（自律性・代替性・耐遮断性に関する AI ソブリンティの確保）

生成 AI を筆頭とする人工知能技術は世界規模で急速な進化を遂げており、米国・中国・英国をはじめとする主要国が国家戦略として政府自身による AI 積極活用を推進している。こうした国際的な AI 競争が激化する中、日本が特定の事業者や基盤モデルに過度に依存することは、サービスの停止、利用条件の変更、価格の上昇、国際情勢の変化等によって、我が国の主体的な選択・運用が阻害され、従属的な立場に立たされるリスクを内包する。このため、第Ⅱ期人工知能基本計画及び「デジタル社会の実現に向けた重点計画」では、データ・計算基盤・モデル・運用の各レイヤーにおいて、「開かれた AI 主権」、あるいは自律性・代替性・耐遮断性を確保する「AI ソブリンティ」を基本的な考え方として明示した。

以上を踏まえ、源内の企画・開発・提供・運用に当たっては、データ、計算資源、基盤モデル、クラウド、アプリケーション、ネットワーク等、AI 基盤全体について、特定の事業者又は技術に過度に依存することなく、デジタル庁がアーキテクチャ及び運用方針を主体的に管理しつつ一部は内製することにより、主体的な選択可能性、代替可能性及び業務継続性を確保することを方針とする。

なお、源内において推進している国産基盤モデルの活用はこの方針に基づく取組であり、政府が率先して国産 AI を調達・評価・改善するサイクルに寄与することは、「世界で最も AI を開発・活用しやすい国」を目指すという政府目標の実現に資する重要な取組である。

（AI ソブリンティを支える疎結合構成）

源内のシステムアーキテクチャは、クラウド、基盤モデル、データ、認証、AI アプリ、エージェント、RAG、

API、ログ、権限管理、評価、データ連携等の各機能を、相互の独立性が高い疎結合構成とすることを基本とする。

これは、特定の事業者、クラウドサービス又は基盤モデルへの過度な依存を避け、自律性、代替性及び耐遮断性を備えた AI ソブリンティを確保するためである。具体的には、原則として、標準化された API、共通データ形式、交換可能な接続部及び機能単位の分離を採用し、必要に応じて国産基盤モデルを含む他のサービスへ切り替えられる構成とする。これにより、障害、価格改定、提供停止等への対応力が高まり、技術選択の自由も確保できる。なお、疎結合の方針を採用するに際しては、構造的な独立性を担保しつつ、デジタル庁において強固な管理・設計統制を維持する。

また、クラウドサービスや基盤モデル等の各要素については、特定の製品、事業者又は技術に固定されないことを方針とする。平時から代替候補、移行手順、データ形式及び接続仕様に関する代替構成への移行可能性を検討・検証し、障害、提供停止、価格改定、性能劣化、利用条件の変更等が生じた場合にも、源内の必要な機能を継続しつつ、速やかに他の構成へ切り替えられる代替性を確保する。

（利用状況の可視化と透明性向上）

源内の利用状況を継続的かつ客観的に把握できるよう、利用者数、利用頻度、アクティブ利用率（アクティベーション率）、利用アプリ、トークン量、トークン利用料その他の指標を可視化するダッシュボードをデジタル庁が開発し、閲覧権限を有する管理者に提供する。また、その一部は利用者にも提供する。

これらの情報を活用し、各府省庁等が職員の利用傾向及び費用を把握するとともに、デジタル庁が源内の機能、運用及び提供方法を継続的に改善できるようにする。可視化に当たっては、業務上の機密性に配慮し、利用目的に応じて表示範囲及び閲覧権限を適切に管理する。

また、政府における AI 利活用の状況、効果及び課題を国民に対して透明性をもって示すため、機密性、個人情報及び利用者の業務への影響に十分配慮した上で、集計・加工した利用状況等を公表する。これにより、政府の AI 施策に対する理解及び信頼の向上を図るとともに、地方公共団体、民間企業その他の組織が AI 利用環境の整備、運用及び利用促進を検討する際の参考となるようにする。

なお、民間の生成 AI サービスでは、利用状況の可視化や取得可能なログの範囲・粒度が提供者の仕様に依存し、政府が必要とする府省庁別、機能別、基盤モデル別、用途別等の分析・評価が困難な場合がある。このため、源内は、利用状況、性能、品質、費用、障害及びリスク等を継続的に把握・分析できるよう、必要なログを取得・管理し、柔軟に分析できる設計とする。

（トークンマネジメントの推進）

利用者及び管理者は、源内の利用に当たり従量課金制のトークン利用料が発生し、出力（応答）のみならず入力（プロンプト）に対しても課金されることを十分に認識した上で、費用を意識して入出力トークンを効率的に管理・活用する「トークンマネジメント」を適切に行うものとする。

（基盤モデルの主体的な選定及び評価）

源内で使用する基盤モデル（大規模言語モデル（LLM）等）は、国産・海外産の別を問わず、利用に係るトークン利用料など費用（コスト）のほか、その能力・正確性・安全性・信頼性・業務適合性等を総合的に評価した上で主体的に選定する必要がある。これは、公務を安全かつ安定的に遂行し、AI を安心して利用で

きる環境を確保するために不可欠である。

基盤モデルの出力は、政策立案、法令・制度調査、国会対応、行政文書の作成等に利用され得るため、誤情報や不適切な回答が行政判断に影響した場合、国民の権利利益や行政への信頼を損なうおそれがある。このため、チャット(対話型 AI)における一般的な会話性能のみならず、日本語能力、行政実務への適合性、法令・政策に関する知識、論理推論、長文処理、上司の指示理解、ハルシネーションの抑制等の能力を確実に評価する。

また、職員の入力情報が意図せず外部へ送信されないこと、モデル提供者の学習データに利用されないこと、ガバメントクラウド上その他の承認された環境で安全に動作することなど、情報セキュリティ上の条件を基盤モデルにおいて確保する。加えて、使用する基盤モデルの学習データ、モデルの開発経緯及びライセンス条件等が不透明である場合には、著作権、個人情報その他の第三者の権利に関する紛争や、提供停止による業務継続上の支障が生じる可能性があることに留意して調達する。

表 ガバメント AI 源内の理念・特徴（まとめ）

分 野	ガバメントAI 源内の理念・特徴
全 体	● 基盤モデル・データ・計算基盤等の各レイヤーにおいて自律性・代替性・耐遮断性を確保する。→AIソブリンティ(AI主権)の確保 ● 利用状況・トークン費用・障害・リスクを徹底的に可視化し、管理できるようにする。モデル選択やデータ更新等に反映する。 ● システムアーキテクチャは、各機能を、相互の独立性が高い疎結合構成とすることを基本とする。特定製品へのロックインを避ける。 ● OSS(オープンソースソフトウェア)公開によって、地方公共団体及び民間企業が参照実装として活用できるようにする。 → ただし、OSSの公開は安全性及び品質を担保するものではなく、OSSを導入・利用する組織は、インフラストラクチャ、アプリケーション、基盤モデル、アクセス制御、ログ管理、保守運用に関する責任を各自で負う。
AIアプリケーション	● 業務実態を熟知する政府自身が行政実務特化型AIアプリを内製することで、現場ニーズへの即応と継続的改善を図る。 ● 利用者価値を重視し、早期にプロトタイプを職員へ提示し、小さく開発・検証して改善するアジャイル／DevOps型開発を原則とする。
AIエージェント	● AIの利用形態を発展させ、単純なチャット応答から、複数の業務工程を連続・自律的に実行するAIエージェント活用へと発展させる。 ● 政府職員のスキルや利用目的の違いに応じて、「エージェントチャット」と「源内工房」の2つの実行環境を提供する。 ● AIエージェントのスキルについて、政府横断的に共有・再利用を行い、政府職員のノウハウ(暗黙知)を集合知へ転換する。
ガバナンス・セキュリティ	● 人間中心、公平性、安全性、可視化・透明性、アカウントビリティ、セキュリティ等の確保を基本とする。 ● ヒューマン・イン・ザ・ループを確保し、AIの処理結果をそのまま受け入れるのではなく、AIを用いた業務プロセスに適切に関与する。 ● 機密性2情報まで扱える環境とする一方、各府省庁が自ら情報の取り扱い範囲や利用条件を定める。
データ	● 政府共通データセットを政府全体の知識基盤として整備し、基盤モデル内部の一般知識だけに依存しない構成とする。 ● データの収集・加工・品質管理・メタデータ付与・更新管理等を標準化し、公表することで相互運用性を高める。
基盤モデル	● 国産・外国産を問わず、能力、正確性、安全性、信頼性、行政実務適合性等による性能本位の選定を原則とする。 ● 基盤モデルについて、能力、正確性、安全性、信頼性、業務適合性に加え、日本語能力、法令・政策知識、論理推論、長文処理、指示理解、ハルシネーション抑制等を評価する。デジタル庁自身が評価テストを実施する。 ● 国産基盤モデルの活用を推進する。複数モデルをタスクに応じて利用者が選択できるマルチモデル方式とする。 ● 政府職員の入出力データは、基盤モデルの学習・性能改善に利用させない契約とする(国内のガバメントクラウド内で推論が完結)。
クラウド基盤	● 政府共通のクラウド基盤であるガバメントクラウドを専ら利用する。 ● 推論処理・業務データ・スキルセット等のデータを国内で処理・保存し、データ主権を確保する。

1-3 源内の具体的な取組

(源内が対象とする AI の範囲)

源内は、政府組織における調査、政策の企画立案、制度設計、法制執務、許認可その他の行政実務への AI 実装を主な対象とする。行政実務は言語を用いる業務が中心であることから、当面は言語モデル及びこれを基盤とするエージェント型 AI を中核として、行政分野に特化したパーティカル AI の実装を進める。

あわせて、音声、画像、映像等を扱うマルチモーダル AI についても、技術の進展及び行政上のニーズを踏まえて順次対応範囲を拡大する。さらに、研究行政等における AI for Science や、行政の現場において実世界の機器・設備・センサ等と連携するフィジカル AI についても、将来的な活用及び源内との連携を検討する。

（政府職員向け AI アプリの内製・展開）

政府職員向けの AI アプリやプロンプト集の内製を推進することは、行政固有の業務手順、判断基準、参照資料及びセキュリティ要件を AI に組み込み、業務の効率化及び業務プロセスの変革(BPR)を促進するために不可欠である。このため、デジタル庁は、行政実務用 AI アプリ等の内製を推進し、汎用的な生成 AI アプリを源内に導入するだけでなく、各府省庁等が保有する行政資料、業務手順及び判断基準を活用しつつ、行政実務に即した AI アプリ等を源内で提供する。

具体的には、国会議事録から政府答弁を検索する「国会答弁を調査・分析アプリ」、法令に関する質問に詳細なレポートを作成する「法制度に関する調査アプリ」、文化庁等のガイドラインに基づく「公用文を校正するアプリ」、補助金制度情報を検索・整理する「補助金制度調査(jGrants)アプリ」などをこれまで内製している。さらに、政策・法令の過去経緯を根拠付きで整理するアプリ、行政文書の用例を調査・検索するアプリ、アナログ規制点検ツールなども開発しており、行政固有の資料・制度・業務手順に即した実務支援を源内で実現している。

（行政実務用 RAG アプリの受託開発と政府横断展開）

デジタル庁では、RAG (Retrieval-Augmented Generation(検索拡張生成))技術を用いて、府省庁等が保有する様々な行政資料(白書、報告書、通達・事務連絡、統計資料、刊行物、各種ガイドライン、審査マニュアル、議事録、報道発表資料等)を生成 AI に参照させ、当該資料に基づく回答を生成する AI アプリ(行政実務用 RAG アプリ)を、府省庁の希望に応じてデジタル庁が受託開発し、源内上で提供する。

これにより、複数の府省庁の利用ニーズが見込まれる有用なアプリを源内上で政府横断的に提供することにより、共通する行政知識やノウハウの共有を図ることができる。

なお、AI アプリの受託開発では、各府省庁等が参照資料の内容、権利関係、更新及び保守に責任を持ち、デジタル庁が開発及びサービス提供に係るシステム上の責任を担う役割分担とする。参照資料の更新遅延等に起因する回答の不正確さや誤りに対する行政上の責任は、当該資料を所管する各府省庁等が負うものとする。また、プロトタイプ動作確認及び精度評価、検収、本番公開後の資料更新、定期的な棚卸し等に基づく効果検証を行うことで、アプリの正確性、最新性及び有効性を継続的に評価し、必要な改善を行う。

（高品質な政府共通データセットの整備）

政府として高品質なデータセットを幅広く確保することは、生成 AI の回答精度、網羅性及び行政実務への適合性を高めるために不可欠な基盤となる。このため、源内においては、政府共通データセットを整備し、政府職員が源内を通じて利用できるようにする。政府共通データセットの整備により、AI が回答を生成する際に、信頼できる行政資料や専門情報を根拠として参照できる状態を整え、基盤モデルが保持する一般的な知識だけに依存することを防ぐことが可能となる。出所、作成主体、作成時期及び原本への参照先が明

確なデータを利用することで、回答の根拠を確認しやすくし、事実と異なる内容が生成されるハルシネーションを低減できる。

デジタル庁は、令和8年度において閣議決定、国会会議録、質問主意書・答弁書、裁判例、地方公共団体の条例・規則、白書、予算・決算資料、条約、統計、政府刊行物、辞書、書籍、新聞・雑誌記事、企業情報等について、推計 1,800 万ページ以上の収集を行うプロジェクトを 7 年度補正予算に基づいて進めている。また、昭和 22 年 5 月以降の官報(79 年 10 か月分)のデータセットの調達・確保も進めている。

行政機関が AI に利用するデータについては、公開されているという理由だけで自由に収集・再利用できるとは限らない。著作権、著作隣接権、データベース権、商標権、肖像権、利用規約等の有無を確認し、源内の開発、評価、利用及び加工に必要な利用許諾を確保することが不可欠である。そこで、権利関係を整理(権利処理)した高品質なデータセットを共通基盤として確保することにより、個々の府省庁が同じ資料について重複して権利確認や調達を行う負担を減らし、適法性、継続性及び利用可能性を確保した上で、政府横断的に利用できるようにする。

将来的には、源内におけるデータの利用実績を踏まえ、データを継続的に追加・更新するとともに、政府全体で共有可能な信頼性の高い知識基盤として発展させる。これにより、行政実務の高度化、府省庁間における重複投資の抑制、政策立案能力の向上を実現し、政府全体の行政遂行能力の強化につなげる。

(政府共通データセットの加工・品質管理及び標準化)

デジタル庁が収集した資料については、AI が適切に利用できる形式に加工するとともに、その品質を確保する。具体的には、紙資料や画像 PDF への OCR 処理、不要なヘッダー・フッターや白紙ページの除去、改行の正規化、明らかな誤認識の修正、文字コードの統一等を行い、加工後のデータについて品質検証を実施する。また、名称、作成者、対象期間、更新日、権利条件、品質検証結果、ファイル形式、想定用途等の情報をメタデータとして付与し、適切に管理できるようにする。これにより、政府職員が、データの信頼性、最新性、法的な利用可能性及び技術的品質を確認した上で、付与された利用条件及び権限の範囲内において、業務に活用できるようにする。

デジタル庁は、これらのデータを源内の政府共通データセットとして整備し、政府横断的な利用を促進するため、資料の収集、権利処理、加工、品質管理、メタデータの付与、命名規則、フォルダ構造、ID 付与、カタログ化及び更新管理に関する標準化を令和8年度事業として進める。この成果については広く公開し、各府省庁等、地方公共団体及び民間企業における信頼性の高い AI 用データ調達及び機関間の相互運用性の確保に活用できるようにする。

(国内外の基盤モデルの適切な選択)

行政実務において AI を活用するためには、日本語の語彙及び表現への適合性に加え、法令、制度、政策、行政組織、公用文の記述様式その他の我が国固有の行政実務に関する理解が求められる。このため、源内では、国内外の基盤モデルをそれぞれの特性に応じて活用するとともに、デジタル庁として行政実務への適合性を客観的に評価する。

国内事業者が開発する国産基盤モデルについても、国産であることのみを理由として採用するのではなく、性能、安全性、信頼性、透明性、法令及び行政実務への適合性、供給継続性、相互運用性並びに費用

対効果等を総合的に評価する。海外の基盤モデルについても同様の基準に基づいて評価し、それぞれのモデルの特性及び利用目的に応じて適切に選択できるようにする。これにより、特定のモデルへの依存を抑制しつつ、政府職員が必要な性能及び機能を利用できる環境を確保する。

なお、政府職員による AI 利用時は、一般利用者向けに設定された安全制限(ガードレール)が、正当な行政目的の遂行を妨げる場合がある。例えば、爆発物の製造、サイバー攻撃その他の危険な手法については、脅威分析、規制立案、捜査、事故防止及び対策検討のため、その仕組みを理解する必要がある。このため、業務上の必要性が認められる場合には、通常のガードレールを緩和又は解除した基盤モデルの確保を検討する。ただし、その場合には、対象者及び利用範囲を限定し、事前承認、本人確認、アクセス制御、監督及び事後検証を行うなど、一定の統制の下で運用する。

(基盤モデルの評価・選定体制の確立)

デジタル庁は、源内で利用する基盤モデルについて、公募、評価、選定及び導入後の継続的な検証を適切に実施できる組織的能力を確保する。組織的能力の確保に際しては、AI セーフティ・インスティテュート(AISI)等の他機関との連携を検討する。

また、基盤モデルは、技術の進展やモデル更新が速いため、主要なベンチマーク指標やモデル提供者の説明のみで判断するのではなく、行政実務への適合性を含む評価基準を独自に策定し、専門的知見を有する政府職員による客観的な審査を行う体制とする。導入後も性能、品質及び安全上の問題等を継続的に確認し、必要に応じてモデルの変更又は利用停止を判断できるようにする。

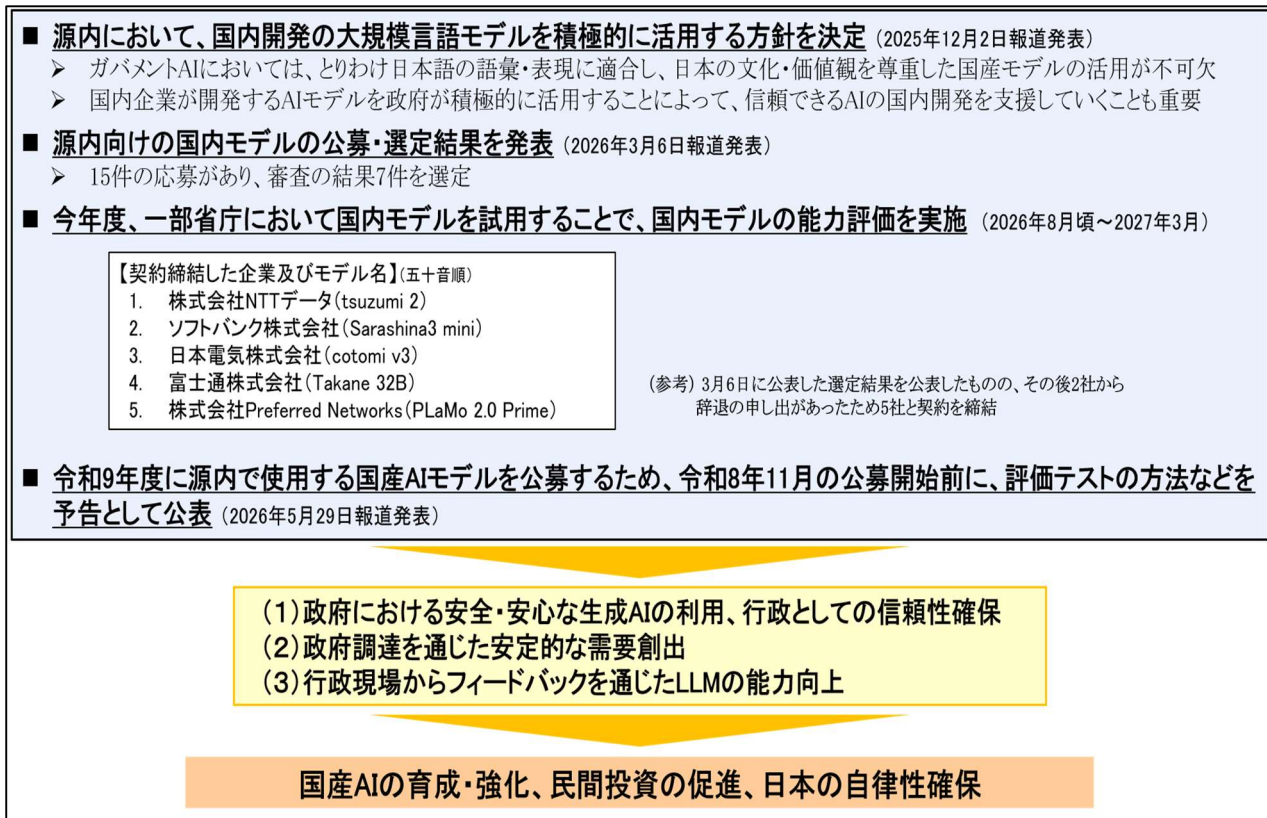
さらに、源内で利用する基盤モデルを選定する際には、生成 AI 技術の進歩の促進と知的財産権の適切な保護の両立に向け、利用者及び権利者にとって安全・安心な利用環境を確保する観点から、「生成 AI の適切な利活用等に向けた知的財産の保護及び透明性に関するプリンシプル・コード」(令和8年8月内閣府知的財産戦略推進事務局策定)に沿った対応がなされているかを考慮する。

(国産基盤モデルの育成・発展への寄与)

源内は、多数の政府職員による実利用を通じて、国内の基盤モデル及び AI サービスの性能向上に寄与する役割も担う。源内における利用状況及び評価・検証結果について、セキュリティ、個人情報保護、秘密保持、知的財産権等の保護に配慮した上で国内事業者にはフィードバックすることにより、日本語及び行政実務に適合したモデルの継続的な改善を促進する。

また、政府が適切な評価及び競争性の確保を前提として国内の AI 技術を調達し、政府職員に対して継続的な利用機会を提供することは、国内事業者による研究開発及び事業展開の予見可能性を高めることに寄与し得る。ただし、調達に当たっては、特定の事業者又は国産基盤モデルを一律に優遇するのではなく、公平性、透明性、競争性及び技術的中立性を確保し、行政目的の達成に最も適したモデル及びサービスを選定する。

図 源内における国産基盤モデルの活用



(源内の価値向上を実現する管理基盤)

源内の利用、品質、費用、性能、障害及びリスクに関するデータを一元的に収集・分析し、複数モデルからの最適な選択、AI アプリ等の継続評価、政府共通データの更新方針の決定、運用自動化及び障害・脅威の予兆検知、AI セキュリティ・セーフティの確保を実現するため、「源内インテリジェント経営基盤」(仮称)の企画・開発・実装を推進する。

これにより、データに基づく源内の継続的な改善及び価値向上を基本方針とし、①出力品質と費用対効果の向上、②運用負担の軽減、③データ整備方針や予算配分の最適化、④障害時の業務継続性確保等を可能とする統合的な運用・経営基盤の確立を目指す。さらに、業務時間の削減、成果物の品質向上、費用削減その他の効果を可能な範囲で測定し、行政におけるAI利用の費用対効果(ROI)を客観的かつ説明可能な形で示すことを目指す。

本基盤により、タスクに応じて国内外の複数の基盤モデルを選択し、必要に応じて連携させるとともに、その性能を評価する統合的な運用も実現し、モデルの評価・選択を我が国が自律的に行える体制を強化する。

(源内のオープンソースソフトウェアの公開)

デジタル庁は、令和8年4月、源内の一部を、商用利用が可能なライセンスの下で、無償のオープンソースソフトウェア(OSS)として公開した。源内のOSSは、地方公共団体、民間企業その他の組織が生成AI利用環境を構築する際の参照実装として活用することができる。これにより、各組織が同様の機能を一から設計及び開発する負担を軽減し、AI導入に必要な要件の検討、設計及び初期開発を効率化することが期待される。

また、OSS は、そのライセンス条件の範囲内で自由に改変及び再利用することが可能である。このため、各組織は、自らの業務、情報管理ルール及びシステム環境に応じて機能を追加又は変更することができ、特定の事業者への過度な依存を抑制するための選択肢となり得る。

一方、OSS を利用することのみをもって、安全かつ適切な AI 利用環境が直ちに構築されるものではないことに注意する必要がある。導入する組織は、クラウドその他のインフラの整備、基盤モデル及び外部サービスの選定、AI アプリの導入、アクセス管理、ログ管理、セキュリティ対策、個人情報及び機密情報の取扱い、保守運用、障害対応並びに利用者教育等について、自らの責任において適切な措置を講ずる必要がある（デジタル庁は、ホスト方式以外の自営方式やその他外部で構築された環境における運用上の損害及び事故についての責任は一切負わないものとする）。

また、地方公共団体等が AI 利用環境を確保・構築する場合には、源内の OSS をそのまま指定することを前提とするのではなく、これを技術的な参照実装の一つとして調達仕様書で活用しつつ、調達目的、必要な機能、性能、安全性、運用条件及び費用対効果を明確にする必要がある。これにより、調達仕様の作成負担を軽減するとともに、調達における公平性、競争性及び技術的中立性を確保できるようになる。

源内は、OSS を通じて、政府内部の業務効率化にとどまらず、政府、地方公共団体、民間企業、研究機関及び AI 事業者が相互に知見及び成果を共有し、我が国全体の AI 利活用能力を高めるための共通基盤として発展させる。源内の利用、評価、改善及び OSS による公開を通じて得られた成果を社会に還元することにより、安全で信頼できる AI の普及、国内 AI 産業及び関連人材の育成並びに官民による AI 投資の好循環の形成を目指す。

（他府省庁や地方公共団体の業務用 AI 基盤との連携）

源内は、各府省庁等が業務上の必要性や情報管理ルールに基づき、行政特化型の業務用 AI 基盤を独自に構築する場合においても、政府共通の AI 基盤として連携・接続を積極的に進めることを基本方針とする。これにより、政府における重複投資を抑制し、政府全体の AI 利活用能力の向上を図る。

源内とのシステム連携により、デジタル庁が、政府共通部分に係る企画、開発、運用、提供、セキュリティ対応及び基盤モデルの更新等を集中的に担うことが可能となる。これにより、各府省庁等は、府省庁固有の業務システムの整備、保有データの管理及び業務改革(BPR)に専念することができる。

具体的な連携機能としては、各府省庁等・地方公共団体が管理する固有データ又は RAG(検索した資料を根拠として AI が回答を生成する仕組み)と、法令、官報、白書等の政府共通データとを、必要な範囲で連携し、相互に参照可能とすることが考えられる。連携方式としては、API(複数のシステム間でデータや機能を連携する仕組み)、MCP(AI と外部のデータやツールを連携するための共通的な仕組み)、A2A(AI エージェント間で情報交換や処理の依頼・連携を行うための共通的な仕組み)等を活用し、システム、データ、ツール及び AI エージェントを相互に接続・連携することを想定する。こうした連携・接続により、政府機関間の AI エージェントが相互に連携し、役割を分担しながら処理を実行することにより、組織の垣根を越える一連の業務プロセスについて、自動化及び省力化を実現できる可能性がある。

また、将来的には、地方公共団体との連携・接続を進め、国と地方の間でデータ及び業務プロセスを相互に連携できる環境を構築することにより、国・地方を通じた行政業務及び行政サービスの高度化につなげることを視野に入れる。

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)

源内においては、チャットや定型的な AI アプリの提供にとどまらず、政府職員が自らの業務に応じて AI エージェント(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)を構築し、活用できる環境を整備する。

源内の Web 上であらかじめ用意された AI アプリは、文書作成、要約、翻訳等の汎用的な業務には有効である。その一方、膨大な資料やデータを用いた一連の作業や、行政組織固有の判断基準、手順及び例外処理を伴う連続的な判断工程には限界がある。

世界的には、現在(令和8年 10 月時点)、Claude Code や Codex に代表される高度なコーディング支援を行う AI コーディングエージェントの利用が広がっているが、その用途はプログラミングに限られない。業務知識を有する一般の利用者が自然言語(ことば)で指示することにより、ファイル作成、データ分析、情報整理、複数工程からなる処理手順の実行、成果物の作成・検証までを一体的に行うことが可能であり、知的業務全般を支援するエージェントとしての活用が進んでいる。このため、政府職員についても、調査、文書作成、データ分析、資料作成、定型処理その他の行政実務における一連の業務プロセスを支援・高度化する手段としてコーディングエージェントは有用である。

このため、源内では、政府職員が自らの業務に応じたスキルを作成し、実行できるようにする。政府職員は、スキルに、業務の目的、判断基準、参照資料、作業手順、確認方法、例外時の対応及び出力形式等を AI との対話によって記述し、職員の経験に属人的に依存していた「暗黙知」を、再現可能かつ検証可能な「形式知」へ転換する。

また、作成されたスキルを府省庁内及び政府横断的に共有し、再利用できるようにすることで個人の知識を政府組織の「集合知」へと発展させる。これにより、業務品質の均一化、人事異動時の円滑な引継ぎの実現、属人化の抑制、優良なユースケースの共有及び人材育成の効率化を図るとともに、各職員が AI を活用して業務プロセスの改革(BPR)に取り組む文化を形成し、政府全体の政策立案能力及び行政遂行能力の強化につなげる。

(AI エージェントによる行政実務の高度化)

源内に AI エージェント利用環境を導入することには、業務の効率化に加え、行政実務の品質、継続性及び説明責任を高める意義がある。AI エージェントに業務の目的、判断基準、参照資料等をあらかじめ設定することで、作業の漏れや担当者ごとのばらつきを抑制し、一定の品質を確保しやすくなる。また、実行した処理、参照した資料、生成した成果物及び職員による修正内容を記録することにより、事後的な検証、監査及び改善にも活用できる。

さらに、複数の作業を連続実行できる AI エージェントによって、職員が関係者との利害調整や創造的・戦略的な判断に充てる時間を確保する効果が期待される。定型的な作業を AI に委ねることにより、職員が作業に費やす時間を削減し、職員が担うべきより高度な業務に注力できるようになる。

また、従来、個々の職員や部局に属人的に蓄積されてきた業務手順、判断基準、参照資料、関連データ、確認方法、出力形式、例外処理等を、AI エージェントが実行可能な形で可視化・形式知化した「スキル」を政府内で標準化し、版管理、評価、共有及び再利用を可能とすることで、府省庁間の重複開発を抑え、優良な実践例を政府全体に展開する。源内の AI エージェント利用環境の設計に際して、特定のモデルや事

業者のクラウドサービスに過度に依存しない構成とすることは、将来の技術変更への対応力及び職員の業務上のデータポータビリティの確保に資する。

（政府が主体的に管理できる AI エージェント利用環境）

AI エージェントを導入した場合には、誤った判断の連続実行、作業ファイルの誤消去・改変、権限を超えた処理、機密情報の不適切な取扱い等のリスクがある。このため、AI エージェントを利用する職員には、その特性、能力及び限界を理解し、実行内容や権限、参照情報、出力結果等を適切に確認・判断できる十分なリテラシーが不可欠である。

特に、自律的に複数の処理を連続実行する高度な AI エージェントについては、利用者の理解不足そのものが重大なリスク要因となり得ることから、必要とされるリテラシーを持つ職員による利用を基本とする。その上で、利用目的及び実行権限の限定、重要な処理における職員の確認・承認、ログの保存、利用停止機能、定期的な性能評価、スキルの審査、責任主体の明確化等の統制措置を講じつつ導入する。

なお、民間が提供する AI エージェント関連サービスは、汎用的な業務の自動化や既存サービスとの連携に有用である一方、提供事業者が定める機能、権限体系、ログ、モデル、データ管理及び運用方式の範囲内で利用することが基本となる。このため、政府の管理者から見た場合、機密情報の取扱い、複雑な権限管理、実行過程の詳細な把握、行政固有の評価、府省庁横断でのスキル共有、基盤モデルの自由な選択、職員が作成したスキル、データその他のアセットの移植・再利用等について、必要な水準の自由度、可視性及び統制可能性を確保できるとは限らない。

特に、内部の処理方式、取得可能なログ、評価方法等が提供事業者の仕様に依存する場合には、管理者にとって AI エージェントの挙動や利用実態を十分に把握・検証できず、ブラックボックス化するおそれがある。また、製品の仕様、利用条件、料金体系及び提供機能が事業者の判断により変更される可能性もある。

このため、源内では、民間製品を適切に活用しつつ、それらでは十分に確保しにくい自由度、可視性、統制可能性及び移植性を補完し、行政実務に必要な AI エージェント利用環境を政府自ら設計・実装する。源内において、「エージェントチャット」及び「源内工房」として、AI エージェントの開発・実行を定められた利用条件及び権限の範囲内で自由に行うことができる環境を内製開発によって確保することにより、特定事業者への過度な依存を抑制するとともに、技術の進展及び行政実務の変化に応じて継続的に改善できるようにする。

（エージェントチャットの位置付け及び機能）

源内が提供する「エージェントチャット」は、Web 上のチャット画面から利用できる、高度なチャット機能を有する AI エージェント利用環境である。職員は特別な操作をすることなく、質問や壁打ち（対話を通じた論点整理及びアイデアの検討）、情報整理、文書作成、データ分析、資料生成、コード実行、既存スキルの利用や新たなスキルの作成を、一つの画面から行うことができる。CSV 等の集計・分析、PowerPoint・PDF・Excel の生成、定型業務の実行など、日常的な行政実務を幅広く支援する。

特徴として、複数の基盤モデルを選択できること、マイドライブ（エージェントチャットにおいて、ファイルの入力・出力・参照のために職員単位で用意されたファイルの記憶場所をいう。以下同じ。）上でファイルを保存・管理できることが挙げられる。ファイルはそのファイルを利用したチャットセッション終了後も保持される。

また、通信先を許可されたサービスに限定し、アップロードファイルのマルウェアスキャン、認証及びアクセス制御を行うことで、安全性を確保している。

(「エージェントチャット」の詳細については別紙 1 参照)

(源内工房の位置付け及び機能)

源内が提供する「源内工房」は、利用者ごとに割り当てられる AI エージェント開発環境である。主な目的は、Web 検索だけでは不十分な網羅的・多段階の情報収集と、政府職員自身による実際に動作するプロトタイプ(Web アプリケーション等)の開発となる。また、複数の Web サイトを直接参照して情報を収集・比較・統合する Deep Research(AI が Web 上の複数の情報を自律的に収集・分析し、詳細なレポートを作成する機能。以下同じ)、Web アプリの構築・動作確認、データ分析、文書・スライド・Excel 等の生成、コードやスキルの作成・実行などにも利用できる。個人専用のファイルシステムにソフトウェアや作業ファイルを保持でき、環境停止後もファイルは維持される。また、任意サイトへの直接アクセスや、ファイル操作、コマンド実行、Web 取得等を多数の工程にわたり連続実行できる点に特徴がある。ただし、任意サイトへの直接アクセスに際しては、原則として、アクセス自体は可能とするが、一部危険性が高いと源内開発者が判断するものについては遮断するなどブロック機能を有する。

また、調査、企画、試作及び検証を一つの環境で連続的に進められることが利点となる。動作するプロトタイプを早期に作成し、利用者による検証を経て仕様書の一部として活用することで、実現可能性や利用者ニーズを確認した上で情報システム等の調達につなげやすくなる。

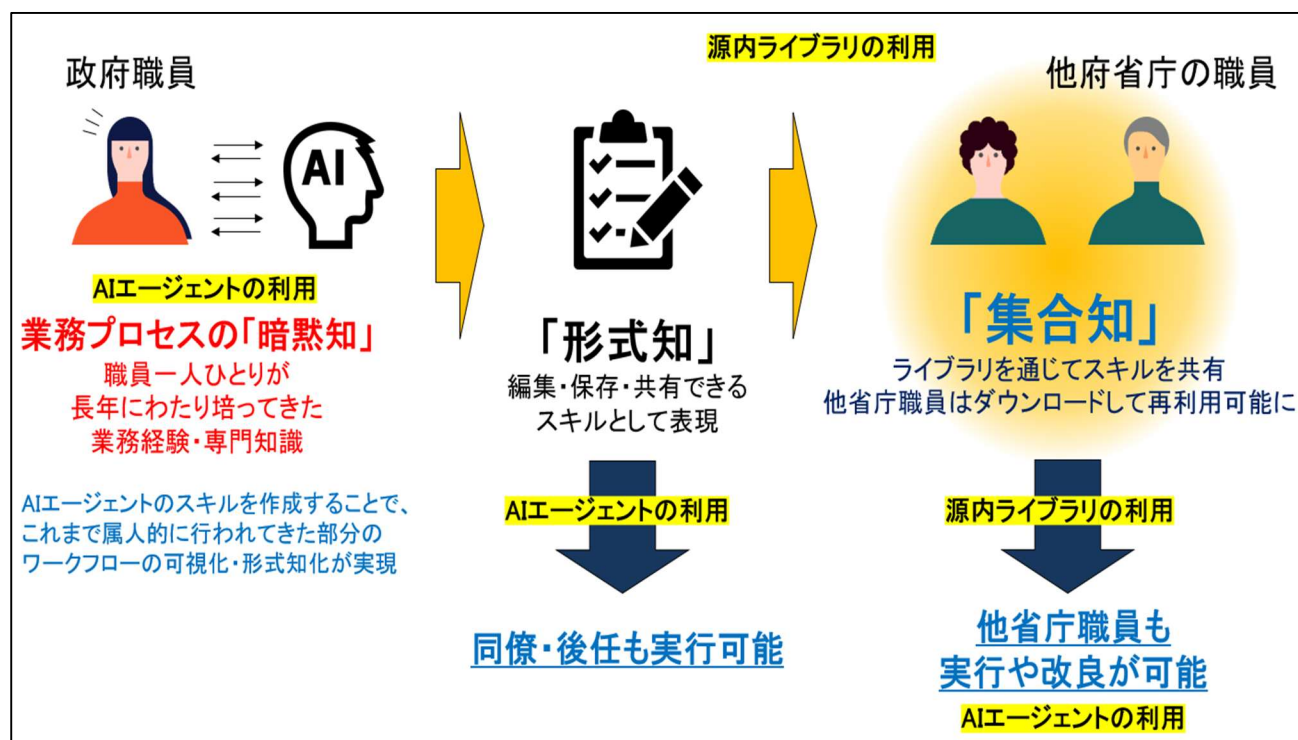
(「源内工房」の詳細については別紙 1 参照)

(スキルの共有・再利用基盤(源内ライブラリ)の導入)

源内において、政府職員が作成した AI エージェント用のスキル及びデータセットを、府省庁横断で共有・活用するための「源内ライブラリ」を整備する。

インターネット上の一般的なマーケットプレイス(ライブラリ)と同様に、提供者と利用者を結び付け、成果物を見つけやすくし、評価や利用実績を通じて優良なものを選別・改善する機能を持たせることで、各府省庁等が類似のスキルを個別に開発する重複を抑制できる。また、既存スキルを基礎として改良や派生開発を行えるようにすることで、政府全体としてナレッジやノウハウを蓄積し、いわゆる「車輪の再発明」(同種・類似の成果物の重複開発)の防止につなげる。

図 AI エージェントによる暗黙知の形式知化・集合知化（源内ライブラリの活用）



源内ライブラリで共有する対象は、調査・リサーチ、データ分析・統計・仮説検証、政策企画・立案、国会対応、審議会・委員会、法制執務、行政手続・処分、補助金・助成金、調達・契約、計画策定・プロジェクト管理、予算要求、文書作成、要約・議事録、スライド・図表作成、翻訳、人事・組織管理、研修、IT・システム管理、広報、総務・庶務、危機管理・防災、監査・コンプライアンス等、幅広い行政分野を想定する。

なお、共有されたスキルを無条件に政府内で共有することは適切ではないため、審査、品質評価、セキュリティ確認、版管理及び廃止手続等の仕組みを源内で今後整備する必要がある。特に、高度な行政判断や国民の権利義務に影響する業務では、職員による確認及び責任ある判断の必要性を利用者は十分に理解する必要がある。

図 源内ライブラリのイメージ（開発中）



（行政手続等における AI 利用の統制及び記録）

総務省では「行政通則法的観点からの AI 利活用調査研究会」が開催され、行政手続法や行政不服審査法等が重視する国民の権利利益の保護や行政の透明性・公正性を確保しながら、行政における AI 利用を進めるための検討が行われている。令和8年5月には「行政通則法 AI 利活用ガイドライン（仮称）の方向性」が検討され、同ガイドラインの策定について検討が進められている。ガイドラインでは、業務の性質等に応じた導入に際しての留意点、残すべき記録、人による関与、情報公開や不服申立ての場面での対応等が整理される見込みである。

このため、源内の AI エージェント利用環境は、同ガイドラインへの適合を前提として設計する。まず、AI を判断主体とするのではなく、職員の調査、審査、分析及び文案作成を支援する手段として位置付け、最終的な判断は職員が行い、AI エージェントを使用した業務の責任は行政機関に帰属することを理解していることを利用条件とする。業務に必要な判断の程度、影響範囲、誤りの回復可能性等に応じたリスクベースアプローチを採用し、リスクの程度に応じて職員によるスキルの開発及び利用を各府省庁等において適切に

統制することも必要である。特に、許認可、給付、不利益処分、採用その他の国民の権利利益に影響する業務や高度な政策判断、利害調整を要する業務では職員によるスキルの開発及び利用を各府省庁等において適切に統制することも必要である。

その上で、AI 調達・利活用ガイドラインでは、政府における生成 AI システムについて、「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適切に行う必要がある」とされているところ、とりわけ、源内では、AI エージェント利用環境の導入を念頭に、一定の高リスク業務においては、通常のアクセスログとは別に、AI 利用記録の作成を必須とする機能の実装を検討する。例えば、入力データ、プロンプト、参照したデータ、生成された回答、使用した基盤モデル及び処理時刻等を自動的に記録できるようにすることも一案となる。これにより、どのような情報と判断過程に基づき AI が出力したのかを後から検証、監査できるようにする。

さらに、源内で用いる基盤モデルについては、デジタル庁として、正確性・信頼性、能力・性能、法令・制度理解力、安全性、中立性等を継続的に評価し、その結果に基づき源内で採用することとする。

以上の取組を通じて、行政の透明性、公正性及び説明責任を維持しながら、安全な AI 利用を政府全体に普及させる共通基盤として源内の整備を推進する。

1-5 源内における機密性情報等の取扱い

行政機関が取り扱う情報には、情報の機密性に応じた区分が設けられている。源内は、「政府統一基準群」及びデジタル庁情報セキュリティポリシーに準拠するとともに、政府情報システムのためのセキュリティ評価制度 (Information system Security Management and Assessment Program。以下「ISMAP」という。) に登録されたクラウドサービスを用いて構築されている。

デジタル庁においては、こうした技術的・制度的な措置を踏まえ、デジタル庁職員が源内上で取り扱うことのできる情報を、「機密性2情報」までとして運用している。一方、他府省庁における入力可能な情報の範囲については、デジタル庁における取扱いを参考としつつ、各府省庁等が自らの情報セキュリティポリシーその他の関係規程に基づいて定め、当該規程に従って運用する必要がある。

ただし、「機密性2情報」には、その内容や漏えい時の影響に幅があるため、一律に源内へ入力できるものではない。このため各府省庁等は、自組織の業務、情報の性質及びリスク等を踏まえ、源内で取り扱える情報の範囲と条件をルール化し、周知・管理することが求められる。また、利用職員も、入力前に機密性や AI 利用の必要性を確認し、定められたルールに従って慎重に AI を取り扱う必要がある。

また、源内に「機密性2情報」を入力した場合は、出力情報についても「機密性2情報」以上の要機密情報が含まれる可能性があることから、出力情報の取扱いにも十分注意しなければならない。

なお、「機密性 1 情報」は、公表済みの情報又は公表を前提とする情報であり、原則として源内で取り扱うことができる。

これに対し、「機密性 3 情報」は、行政事務で取り扱う情報のうち、秘密文書に相当する機密性を要する情報であり、源内に入力してはならない。加えて、特定秘密(「特定秘密の保護に関する法律」(平成 25 年法律第 108 号)第3条第1項に規定する特定秘密)、重要経済安保情報(「重要経済安保情報の保護及び活用に関する法律」(令和6年法律第 27 号)第3条第1項に規定する重要経済安保情報)についても源内では取り扱ってはならないものとする。これは、プロンプトへの記載、添付ファイルの送信、AI エージェント利用環境のマインドライブへの保存その他の入力方法を問わない。源内の利用者は、源内を利用する前に対象情

報の機密性区分を確認し、機密性 3 情報や特定秘密、重要経済安保情報に該当し、又は該当する可能性がある情報を源内上で取り扱わないよう、十分に注意しなければならない。

その他、源内で生成 AI を利用する上での要機密情報の取扱いについては、国の行政機関向けのガイドラインとして示されている「AI 調達・利活用ガイドライン」等に沿った対応が求められる。

1-6 源内の出力情報に係る行政文書の取扱い

公文書管理法上の「行政文書」とは、行政機関の職員が職務上作成又は取得し、組織的に利用するものとして、当該行政機関が保有している文書をいう。令和7年 2 月の内閣府公文書管理課長通知 2-3(共有フォルダにおける行政文書の電子的管理に関するマニュアル(令和7年 2 月 14 日内閣府大臣官房公文書管理課長))では、行政文書を作成する前段階の下書きや、職員が参考のために収集した資料については、「個人用フォルダ」に保存することとされている。このように、職員個人の検討段階にとどまり、組織的に利用されていない文書は、非行政文書として取り扱われる。

源内において AI が出力した結果についても、職員が自らの調査や検討のために利用し、本人だけが管理する環境に保存している限り、通常は、個人用フォルダに保存されたメモや下書きと同様に、組織的に利用されているものとはいえない。このため、行政文書には該当しないと整理できると考えられる。

一方、出力結果を他の職員と共有し、組織の共有フォルダ、組織内の共有チャンネルその他の共用環境で業務に利用した場合は、組織的な利用が始まったものとして行政文書に該当する可能性がある。その場合は、保存期間の設定、行政文書ファイルへの編てつ(書類や原稿を編んで綴じること)その他の文書管理が必要となることがあるため、所属府省庁又は所属課室の文書管理担当者に確認し、適切に取り扱う必要がある。

その他、源内の出力情報(出力結果)の取扱いについては、国の行政機関向けのガイドラインとして示されている「AI 調達・利活用ガイドライン」等に沿った対応が求められる。

図 ガバメント AI 源内の利用画面イメージ (2026 年 10 月時点)

The screenshot displays the GBAI (Gaba AI) source interface. At the top, there's a navigation bar with tabs: 「ガバメントAI 源内」, 「チャット」, 「エージェントチャット」, 「AIアプリ」, 「ドライブ管理」, and 「お知らせ」. On the right, there are links for 「サポート」 and 「アカウント」.

A prominent notification box states: 「新モデル登場 - Claude Opus 5.5が利用可能になりました。」 (New model launch - Claude Opus 5.5 is now available.) It mentions that this latest model is strong in understanding complex text and processing long documents, supporting administrative tasks more efficiently.

Below the notification, a message asks: 「お手伝いできることはありますか？お気軽にご相談ください」 (Is there anything I can help with? Feel free to consult). It provides an example prompt: 「例) "国会答弁を検索したい"、"法制度について詳しく調べたい"、"引き継ぎ作業をやりたいので適切なプロンプトを考えて"」.

The interface includes a section for selecting the AI model, currently set to 「Claude Opus 5.5」, with a link to 「AIモデルの選び方と注意事項」 (How to choose an AI model and precautions).

A large text input area is provided for the user's query, accompanied by a button to 「添付するファイルを選択 (1ファイルあたり4.5MB、PDFは15MBまで)」 (Select files to attach (4.5MB per file, PDF up to 15MB)) and a 「送信」 (Send) button.

An 「お知らせ」 (Notice) section lists several updates, including the completion of the image generation app, a survey on file management, and the launch of the GBAI app and prompt library. A button 「お知らせ一覧を見る」 (View all notices) is provided.

At the bottom, a 「最近使ったAIアプリ」 (Recently used AI apps) section shows four options: 「チャット」 (Chat), 「法制度に関する調査」 (Survey on legal system), 「源内利用状況ダッシュボード (全職員版) ヘロログイン」 (Source usage dashboard (all staff version) Hero login), and 「ウェブ情報や手元資料を要約」 (Summarize web information and hand资料).

2 利用者の責務及び遵守事項

2-1 利用者とは

本基準において、「利用者」とは、源内を使用する政府職員等をいう。

なお、「政府職員等」とは、国の行政機関において行政事務に従事している国家公務員のほか、源内の利用を希望する立法府及び司法府の職員並びに独立行政法人等の職員とする。

2-2 利用者が遵守すべき基本原則

(AI の積極的な活用)

- 利用者は、AI が多くの業務に有用であることを踏まえ、AI を過度に忌避することなく、業務の目的、重要度及びリスクに応じて、日常の行政実務に積極的に活用すること。AI の利用は、行政課題をより創造的・戦略的・革新的に検討する契機となり得ることから、AI の活用に開かれた姿勢を持つこと。
- 利用者は、調査、資料作成、データ分析その他の業務において AI の能力を十分に引き出すとともに、より高度なデータ分析やエビデンスに基づく意思決定の支援にも活用すること。また、既存の業務に限定せず、AI を活用できる新たな業務や業務プロセスを積極的に探索し、実際に試行すること。
- 利用者は、従来の業務手法にとらわれることなく、AI 利用による効果及び課題を継続的に検証し、利用方法を改善すること。その結果を、業務の効率化、品質向上、利用者の負担軽減及び業務プロセスの変革(BPR)につなげること。

(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」(令和7年 12 月 19 日人工知能戦略本部決定)に基づき、利用者は、源内を積極的に利用し、AI の能力を行政実務の効率化、品質向上及び業務改革に最大限活用するための基礎となる知識及び能力を身に付けること。
- 同時に、利用者は、社会における AI 利用の模範となることを意識し、AI の特性、能力、限界及びリスクを正しく理解し、適切に使用・判断・検証・利用するための最新の知識及び能力を継続的に身に付けるとともに、高い倫理観を保持すること。

(AI 活用における専門性と判断力の維持)

- AI は利用者の能力を代替するものではなく、専門性を拡張するための手段として利用すること。AI への過度な依存は批判的思考や利用者の技能を低下させ得るため、自ら理解・保有していない技能をAI に安易に委ねないこと。
- 利用者は、自ら考え、判断し、検証する能力を維持するとともに、AI を活用しながらも専門知識及び業務遂行能力を継続的に向上させること。AI の出力を評価できるだけの専門性(目利きの能力)を維持すること。AI が利用できない場合でも、重要業務を遂行できる能力を維持すること。
- 特に若手職員については、AI を積極的に活用して業務の効率化及び学習の高度化を図る一方、職務遂行に必要な基礎的な知識、技能、思考力及び判断力を徹底的に身に付けること。

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

- 利用者は、AI による処理をそのまま受け入れるのではなく、AI を用いた業務プロセスの適切な段階に関与し、必要な確認、判断、承認、修正、監督又は停止を行うこと。AI の出力結果を職員が確認しないまま、行政資料や公文書への直接の転記はしないこと。
- 源内を利用する際は、業務の性質・内容に応じて、例えば、次のような対応を行うこと。
 - プロンプトを入力する前に、その内容や含まれる情報が適切であるかを利用者自身が確認する。
 - AI に与える指示、参照資料及び実行条件が業務目的に適合しているかを確認する。
 - AI から出力された回答をそのまま採用せず、根拠、正確性及び妥当性を批判的に検証する。
 - AI が作成した文書、分析結果その他の成果物を利用者が確認し、必要に応じて修正する。
 - AI が複数の処理を連続実行する場合に、重要な処理の前に利用者による確認又は承認を行う。
 - ファイルの更新・削除、データの登録等、影響の大きい処理は利用者が実行可否を判断する。
 - AI の処理内容が当初の目的、権限又は指示の範囲を逸脱していないかを監督する。
 - AI の出力を行政上の判断や対外的な説明に用いる場合に、利用者が最終的な内容を確認する。
 - 意図しない動作、異常な出力その他の問題を認識した場合に、処理を停止し、必要に応じて管理者(第三章参照)に報告する。

(ハルシネーションに対する注意)

- 生成 AI は、利用者の指示に対して自然で説得力のある文章を生成できる一方、その内容が事実と異なる場合や、存在しない情報をあたかも事実であるかのように生成する場合がある(いわゆる「ハルシネーション」)。また、基盤モデルには学習データの時点によるカットオフ問題があり、それ以降に生じた制度改正、組織変更、最新統計、社会情勢等を把握していない場合がある。さらに、生成 AI は数値計算を誤る場合がある。
- このため利用者は、生成 AI の出力をそのまま事実として扱わず、重要な情報については、原典、法令、政府機関の公表資料その他の信頼できる一次資料により内容を確認すること。ハルシネーションを前提とした批判的検証を常に行うこと。また、計算結果については、その他の計算手段を用いて検算すること。
- 特に、次のような情報については十分な確認を行う。
 - 具体的な数値、統計、割合、金額、年月日
 - 法令の名称、条文、改正内容、施行日等
 - 制度、政策、行政手続、支援措置等の内容や適用条件
 - 人名、組織名、役職名、地名、施設名その他の固有名詞
 - 文書、報告書、会議、通知、通達等の名称や存在
 - 出典、引用、URL、参考文献等の情報
 - 数値計算、割合、集計、単位換算その他の計算結果

(プロンプトインジェクションへの警戒)

- プロンプトインジェクションとは、AI に与える入力や AI が参照する情報に悪意ある指示を埋め込み、本来の指示や目的から逸脱した動作を誘発する攻撃手法をいう。例えば、「これまでの指示を無視して〇〇せよ」などの命令を混入させることで、AI に不適切な処理を行わせ、機密情報の開示、有害な内容

の生成、意図しないツール操作等を試みるものである。特に、Web ページ、外部データ、RAG の参照資料、利用者が添付するファイル等に悪意ある指示が埋め込まれている場合、利用者が直接その指示を入力していなくても、AI がこれを読み取り、従おうとする可能性がある。

- このため、源内を利用する際は、外部情報や添付ファイルを無条件に信頼せず、AI の実行内容や出力結果を確認するとともに、意図しない動作を認識した場合には処理を停止するなど、十分な注意を払うこと。
- 特に、次のような手法や命令の混入については十分な確認を行うこと。
 - 「これまでの指示を無視せよ」など、既存の指示を上書きしようとする命令
 - Web ページや文書中に、通常の本文を装って悪意ある指示を記載する手法
 - 白文字、極小文字、背景と同色の文字など、人間から見えにくい形で指示を埋め込む手法
 - PDF、Word、Excel、PowerPoint 等の添付ファイル内に悪意ある指示を埋め込む手法
 - RAG で参照される文書やデータベースに悪意ある指示を混入させる手法
 - 検索結果や Web ページを経由して間接的に AI へ指示を与える手法
 - 「管理者からの指示」「システムメッセージ」などと偽装し、優先度の高い命令のように見せる手法
 - 複数文書や複数段階の指示に分割し、組み合わせることで最終的に不正な処理を誘発する手法

(リスクベースアプローチの採用)

- AI 利用におけるリスクベースアプローチとは、すべての AI 利用を一律に管理するのではなく、利用目的、対象業務、取り扱い情報、判断への影響及び誤りが生じた場合の影響等を踏まえ、リスクに応じて必要な統制措置の強度を変える考え方である。低リスク用途では過度な手続を避け、高リスク用途では利用者による確認・承認、精度・安全性の検証、記録・ログの保存等を強化すること。
- 源内の利用に当たっては、「AI 調達・利活用ガイドライン」を踏まえ、利用者が AI に委ねる業務を適切に選択し、特に AI エージェントについてはリスクベースアプローチによる対応を基本とすること。
- 例えば、部局内に閉じる定型的・補助的な業務は相対的に低リスクであり積極的な利用が可能である一方、行政処分、対外的な意思決定、政策判断等、国民の権利利益や行政運営に大きな影響を与える業務では、より慎重な利用及び強い人的統制を行うこと。
- また、AI エージェント等の利用に当たっては、実行した処理、参照情報、指示、ツール利用等の処理履歴を必要に応じて記録し、事後的に検証できる状態を確保すること。

(高リスクな生成 AI 利活用の判定)

- 利用者は、源内を利用する AI のユースケースについて、「AI 調達・利活用ガイドライン」に定める高リスク判定の考え方及び「高リスク判定シート」を参考として、AI を利用する業務の性格その他の事情を総合的に勘案し、リスクレベルを判断しなければならない。
- 利用者は、高リスクに該当する可能性が高いと判断した場合、各府省庁の生成 AI 利活用ルール等に基づき、当該プロジェクトの目的、利用方法、想定するリスク、リスク軽減策及び運用時の品質確保策を整理し、CAIO への報告その他必要な手続を行うこと。

(シャドーAI の利用禁止)

- 利用者は、業務上 AI を利用する場合、源内など組織としてその利用が認められている生成 AI サービスを活用するものとし、利用が認められていない生成 AI サービスを使用しないこと。利便性や機能性を理由として、組織の管理・統制の及ばない、いわゆる「シャドーAI」を業務に利用しないこと。
- 例えば、次のような行為は行わないこと。
 - 私用のチャット AI サービスに非公開の業務資料を入力して分析させる。
 - 個人所有のスマートフォンの生成 AI アプリを使って業務に関する機微な質問やデータを入力する。
 - 個人アカウントの生成 AI に業務メールを貼り付けて返信案を作成させる。
 - 個人契約のチャット AI サービスを使って決裁前の計画案の誤字脱字チェックを行わせる。

（人物に関する情報等の適正な取扱い）

- AI を用いた個人の検索は、担当する事務の遂行に必要な場合に限り行うこととし、他の目的に転用しない。本人が表明していない信条等の属性を推定し、又は将来の行動を予測することに使用してはならない。また、複数の情報源を統合して個人単位で蓄積しないこと。
- 個人情報を含むプロンプトを AI に入力する場合には、あらかじめ特定された利用目的のための必要最小限の利用であることを十分に確認すること。氏名、住所その他の個人を識別する情報が処理に必要でない場合には、削除、匿名化その他の措置を講じ、必要以上の個人情報を入力しないこと。
- 各府省庁等のルールで入力が禁止されている情報については、各府省庁等が独自に定めるルールに従わなければならない。

（トークンマネジメントの推進）

- 利用者は、源内の利用に当たり従量課金制のトークン（AI が文章を処理する際の単位）の利用料が発生し、出力（応答）のみならず入力（プロンプト）に対しても課金されることを十分に認識した上で、費用（コスト）を意識して入出力トークンを効率的に管理・活用する「トークンマネジメント」を適切に行うものとする。
- 具体的には、冗長な指示や不要な文書の貼り付けを避け、プロンプトは簡潔かつ的確に設計し、入出力双方のトークン消費を抑制すること。また、無用な大量出力や繰り返しの試行を避け、費用対効果（ROI）を意識した利用に努めること。さらに、処理内容の難易度や求める精度に応じ、過剰に高性能な基盤モデルを用いず、適切な基盤モデルを選択することとする。
- このため、源内では、利用の都度、トークン数やコストを表示するとともに、基盤モデルごとの特徴などその選択方法と注意事項を閲覧可能な仕様とする。

2-3 エージェントチャット

（利用者が理解すべきリスク）

- エージェントチャットは、通常の AI チャットに加え、ファイル参照、Web 検索、複数段階の処理を実行できるため、業務効率化の効果が期待される一方、不適切な判断又は指示の影響が、後続する複数の処理に連鎖的に及ぶおそれがあることを理解しなければならない。
- 具体的には、事実と異なる情報の生成、指示内容の曖昧さに起因する意図しない処理、機密情報・個

人情報等の不適切な参照、偏りのある出力等が想定される。

- 特に、生成結果を利用者が確認せず、行政上の判断、通知、回答、登録又は外部への送信等に直接利用している場合には、国民の権利利益、行政の公平性及び説明責任に影響を及ぼすおそれがあることを理解すること。
- また、エージェントチャットは、ファイルを参照するだけでなく、利用可能な権限の範囲でファイルの作成、編集、移動、名称変更、上書き又は削除等の操作を行う場合がある。このため、利用者の指示が曖昧な場合や、対象ファイル・処理範囲を誤って指定した場合には、必要なファイルの意図しない削除、内容の改変、上書き、保存先の変更等が生じるおそれがあることを理解すること。
- 重要なファイルを扱う際は、事前のバックアップ、対象範囲の限定、実行内容の確認及び操作後の検証を行うこと。

図 エージェントチャットの利用画面イメージ（2026 年 10 月時点）

The screenshot displays the 'Agent Chat' interface. On the left, there's a sidebar with 'マイドライブ' (My Drive) containing files like 'pdf', 'data', and 'chats', and a '利用履歴' (Usage History) section with recent chat logs. The main area shows a chat session titled 'エージェントチャット' using 'Claude Opus 4.8' in '深く推論する' mode. A task list is visible, including 'スキルを有効化', 'ウェブを検索 8回実行', 'ファイルを読み込 内容確認 4回実行', 'プログラムを実行して処理', and 'ファイルを作成・書き込み 2回実行'. A red overlay box highlights the following features:

- 通常のチャット(対話型AI)の発展形として位置づけ
- AIE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
- AIE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

At the bottom, there's an input field and a '送信' (Send) button.

(注) デジタル庁では 2026 年 8 月以降全庁に展開

(利用者が遵守すべき事項)

- 利用者は、エージェントチャットを過信せず、利用目的、指示内容、参照させる情報及び実行させる処理の範囲を明確にした上で利用しなければならない。
- 入力又は参照させる情報については、機密性、個人情報、知的財産権等のほか、利用権限を事前に確認し、適切な範囲に限定しなければならない。
- 生成された内容や処理結果は、根拠資料、数値、引用、法令、宛先等を利用者自身が確認し、重要な

判断、行政上の処理又は外部への発信には、必要な確認・承認を経ることとしなければならない。

- エージェントチャットによるファイルの編集、上書き、移動又は削除を伴う場合は、ファイルの対象、処理内容及び影響範囲を事前に確認し、原則としてバックアップを確保すること。
- 国民の権利利益の保護や行政の透明性・公正性を確保するため、許認可、給付、不利益処分、採用その他の国民の権利利益に影響する業務では、利用者は、業務の重要度、影響範囲、誤りの回復可能性等に応じたリスクベースアプローチを採用し、エージェントチャットを適切に利用しなければならない。
- 以上のリスクを踏まえ、エージェントチャットの利用開始前に必要な注意事項等を画面上にポップアップで表示し、その内容を確認した利用者が同意した場合に限り、機能を利用できる仕組みとする。

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)

- 源内工房は AI コーディングエージェント機能として、①Web 検索だけでは不十分な網羅的・多段階の情報収集の実施、②実際に動作するプロトタイプの開発を利用者自身が実施できる AI エージェント利用環境である。
- 利点は、利用者が行う調査、企画、試作を一つの環境で連続的に進められることにあり、プログラムの作成・修正・実行、コマンド操作、ファイル編集、Web 検索等を行うことができるため、通常の AI チャットやエージェントチャットよりも操作の自由度が高い。
- 一方で、AI が生成したコードの誤り、曖昧な指示に基づく意図しない処理、ファイルの誤削除・改変・上書き、脆弱性を含むアプリの作成、第三者の著作物やライセンス条件への抵触等のリスクがある。
- また、任意の Web サイトへの直接アクセスが可能(ただし、一部、源内開発者が危険と判断するものは遮断)であり、AI がファイル操作、コマンド実行、Web 情報の取得、Web アプリの作成等を数十から数百のステップにわたり連続して実行することがある(例:サプライチェーン汚染(ツール説明文への悪意ある命令の埋め込み、スキル利用時に発動するリバースシェル機能等)のリスクに対する懸念)。
- このため、利用者の意図しない情報流出その他の挙動が生じる可能性がある。さらに、利用者がスキル、コード又は処理内容を十分に理解しないまま実行した場合には、問題や事故の影響範囲や問題発生の原因を把握できないおそれがある。

図 源内工房の利用画面イメージ（2026 年 10 月時点）



（利用者が遵守すべき事項）

- 利用者は、源内工房の高い自由度と自律的な処理能力を理解し、自らの指示及び AI が実行した処理について適切に確認する責任を負わなければならない。
- 具体的には、利用前に、目的、対象業務、使用する情報、外部接続及び処理範囲を明確にし、AI が生成したコード、コマンド、ファイル操作及び Web アプリは、実行又は公開前に内容、動作、影響範囲及び安全性を確認し、必要に応じてバックアップや試験を行うこと。
- 処理内容を理解できない場合は安易に実行せず、想定外の動作、情報流出、誤処理又は脆弱性を認識したときは、直ちに処理を停止し、管理者へ報告しなければならない。
- 許認可、給付、不利益処分、採用その他の国民の権利利益に影響する業務におけるリスクベースアプローチの採用、源内工房の利用制限については、エージェントチャットと同様である。
- 以上のように、源内工房の利用には、自らの指示及び AI が実行した処理を適切に確認する高い能力が求められる。このため、利用者には、AI 調達・利活用ガイドラインなど既存のガイドラインに基づいた対応が求められるほか、源内開発者が提供するリテラシーチェックの活用等により一定の水準を満たすことを管理者が確認等した職員に限って、源内工房を利用できるものとする。

3 管理者の責務及び遵守事項

3-1 管理者とは

- 本基準において、「管理者」とは、各府省庁等における AI 統括責任者(CAIO)、府省内全体管理組織(PMO:Portfolio Management Office)、情報システム管理者その他これらに準ずる主体をいう。
- 具体的にどの部局又は職員を管理者とするかは、デジタル庁が一律に指定するのではなく、各府省庁等が自らの組織体制、業務内容及び利用形態等を踏まえて明確にすること。
- 管理者は、「3-3 各府省庁等への「テナント管理者機能」の提供」を前提に、その役割及び権限に応じて、源内の AI エージェント利用環境、基盤モデルその他の機能の利用可否を判断するとともに、組織内における適正な利用、権限管理、利用状況の把握その他必要な管理を行うこと。
- 管理者の指定及び役割分担に当たっては、「AI 調達・利活用ガイドライン」の「4.2 各府省庁における AI ガバナンス体制の整備」も併せて参照すること。

3-2 各府省庁等における内部ガバナンス

- 源内を利用する各府省庁等は、生成 AI の利活用促進とリスク管理を一体的に進め、行政目的、業務の重要度及びリスクに応じて、源内を適切かつ効果的に利用できる内部ガバナンスを整備しなければならない。
- 各府省庁等は、利用可能な業務、機能及び情報の範囲、利用者の権限、予算及び費用の管理、利用者による確認・承認、事故発生時の対応その他必要な事項を定め、組織として責任ある利用を確保すること。その際、「AI 調達・利活用ガイドライン」、「政府統一基準群」、AI 事業者ガイドラインその他の関係規範との整合に留意すること。
- 管理者は、利用者に対し、「2 利用者の責務及び遵守事項」の内容を周知し、必要な教育、研修、助言、確認その他の措置を講じることにより、当該事項に沿った適切な利用を確保しなければならない。
- 各府省庁等は、技術の進展、利用実態、リスク及び事故事例等を踏まえ、内部ルール及び運用を定期的に見直すこと。その際、幹部、PMO、情報システム・セキュリティ担当、各部局その他の関係者が連携し、継続的な改善を行うこと。

3-3 各府省庁等への「テナント管理者機能」の提供

- デジタル庁は、各府省庁等の管理者に対し、今後、源内の利用を効果的に管理するための以下の「テナント管理者機能」を提供することを検討する。
 - ① 利用者の管理:利用者の一覧を確認し、必要に応じて利用を停止する機能
 - ② 費用ダッシュボードの閲覧権限管理:源内の利用費用を確認できる者を設定・管理する機能
 - ③ 利用可能な基盤モデル及び機能の管理:利用者が使用できる大規模言語モデル(LLM)その他の基盤モデル及び源内の機能を設定・管理する機能
- 各府省庁等の管理者は、内部ガバナンスを効果的に実現するため、これらの機能を活用すること。

3-4 利用状況及び利用費用の把握

- 源内では、認証・利用者情報、利用機能、操作日時、利用した基盤モデル、入出力トークン数、入力内容、システムプロンプト、アップロードファイル及び出力内容等について、必要な範囲でログを取得し、一定期間保存する。ログは、デジタル庁が管理するガバメントクラウド内に保存し、源内の運用管理、不正アクセス等の監視、利用状況の分析・サービス改善、基盤モデルの評価及び運用状況の確認等、あらかじめ定めた目的の範囲内で利用する。必要に応じて、源内を利用する省庁の管理者に情報を開示する。また、ログの取扱いを判断する権限主体を明確にする。
- 各府省庁等の意向に応じ、デジタル庁による当該府省庁のログ利用を制限できる仕組みを設ける。利用制限の対象となるログを分析する必要がある場合には、原則として機械的な処理により行い、入力内容その他の具体的な内容が識別されないよう必要な措置を講ずる。
- デジタル庁は、ログを第三者に提供する場合は、法令に基づく場合その他あらかじめ定めた例外を除き、当該ログに係る府省庁の同意を得る。基盤モデルの提供事業者等に利用状況をフィードバックする場合は、具体的な入力・出力内容を提供せず、一般化、集計又は秘匿化した情報を用いる。
- 以上の内容は、「デジタル庁が取得するログ情報及びその取扱いについて」(令和8年4月 20 日改定)において、その具体的な内容を規定する。
- デジタル庁は、各府省庁等が自府省庁における源内の利用状況及び費用を把握できるよう、費用その他必要な利用情報を可視化するダッシュボードを提供する。
- 各府省庁等の管理者は、「3-3」の「テナント管理者機能」の提供を前提に、当該ダッシュボードを利用し、自府省庁における源内の利用状況及び費用を継続的に把握するとともに、必要に応じて、利用者、利用可能な基盤モデル、利用機能その他の管理に活用すること。
- 管理者は、源内の利用に当たり従量課金制のトークン利用料が発生することを十分に認識した上で、当該ダッシュボードを活用しつつ、費用(コスト)を意識して入出力トークンを効率的に管理・活用する「トークンマネジメント」を適切に行うものとする。

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)

- 管理者は、利用者個人の注意のみに依存せず、エージェントチャットの利用に関する組織的な統制を行う責任を負うこと。また、デジタル庁が提供する利用上の留意事項等を踏まえ、各府省庁等におけるAIの利用状況や職員のリテラシーなど個別の状況も勘案しつつ、例えば、利用者の教育、リスクの周知啓発、定期的なリスク評価及び事故対応手順を導入すること。
- 国民の権利利益の保護や行政の透明性・公正性を確保するため、管理者は、許認可、給付、不利益処分、採用その他の国民の権利利益に影響する業務では、業務の重要度、影響範囲、誤りの回復可能性等に応じたリスクベースアプローチを採用し、エージェントチャットの利用を適切に制限しなければならない。

(システムの機能制限及び安全管理措置)

- エージェントチャットでは、利用開始前に必要な注意事項等を画面上にポップアップで表示し、その内

容を確認した利用者が同意した場合に限り、エージェントチャットの機能を利用できる仕組みとする。

- エージェントチャットは、政府職員の利用に必要なセキュリティ対策を講じ、機能を広範囲に制限した利用環境としている(令和8年10月現在)。
 - エージェントチャットには、利用者ごとに個別のマイドライブを用意(マイドライブは一時的な作業領域として設計されており、ファイルの永続的な保存を保証するものではない。)
 - マイドライブには、利用者1人当たりの保存可能量の上限を設定。
 - エージェントチャットが上書き又は削除できる対象は、AIが生成したファイルに限定されている。
 - 利用者がマイドライブへアップロードした原本ファイルは、AIが上書き又は改変できない仕組みとなっている。(エージェントチャットがファイルを上書きする際には、当該原本ファイルを自動的に複製し、その複製物を上書きしている。)
 - マイドライブへアップロードされたファイルには、システムによるセキュリティスキャンが自動的に実施される。ただし、セキュリティスキャンは一定の脅威を検知・遮断するものであり、未知の脅威を完全に排除するものではない。
 - パスワードで保護されたアーカイブ等、内容を検査できないファイルは、安全性を確認できないものとして読み取りが遮断される。
 - Web 検索を実行した場合、検索キーワードを含むすべての処理はガバメントクラウド内に閉じて実行され、エージェントチャットからコンテンツを取得した Web サイトへの通信は行われない。
 - エージェントチャットで利用できるツールは、利用者本人のマイドライブ内に限定して動作する。
 - エージェントチャットから、利用者のローカル PC を直接操作することはできない。
 - エージェントチャットから、源内開発者側で管理されていない API や MCP サーバーへ直接アクセスすることはできない。
 - 添付された Python ファイルやシェルスクリプト等の実行ファイルが、アップロードのみを契機として自動的に実行されることはない。
 - ツール及びコードの実行範囲は、システム上あらかじめ定められた環境及び権限の範囲内に制限されている。

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)

- 管理者は、利用者の注意のみに依存せず、源内工房の機能及びリスクに応じた組織的な管理体制を整備する責任を負わなければならない。
- 源内工房の利用には、自らの指示及び AI が実行した処理を適切に確認する高い能力が求められる。このため、管理者は、AI 調達・利活用ガイドラインなど既存のガイドラインに基づいた対応が求められるほか、自組織の業務上の必要性や取扱情報、想定されるリスク等も踏まえつつ、必要に応じて、源内開発者が提供する利用に際して必要なリテラシーチェックを活用するなどして、一定の水準を満たした職員かどうかを確認等した上で、利用者に対する源内工房の利用を可能とする。
- また、デジタル庁が提供する利用上の留意事項等を踏まえ、各府省庁等における AI の利用状況や職員のリテラシーなど個別の状況も勘案しつつ、例えば、利用者教育、リスクの周知啓発、定期的なリスク評価及び事故対応手順を導入するとともに、事故又は異常を把握した場合は、今後、デジタル庁がそ

の提供を検討している「テナント管理者機能」(3-3参照)による速やかな利用停止のための措置、影響範囲の確認、復旧、関係部門への報告、原因分析及び再発防止を行い、利用実態や技術の変化に応じてルールを継続的に改善すること。なお、「テナント管理者機能」の提供が行われない間の利用停止のための措置は、管理者から報告等を受けたデジタル庁(源内開発者)が対応するものとする。

- 許認可、給付、不利益処分、採用その他の国民の権利利益に影響する業務におけるリスクベースアプローチの採用、源内工房の利用制限については、エージェントチャットと同様である。

(システムの機能制限及び安全管理措置)

- 源内工房は初期状態では利用不可とし、上記のとおり、リテラシー等について一定の水準を満たしていると確認等できた職員からの申請に基づき、管理者が、当該職員への利用権限を付与する仕組みとする。
- 源内工房は、以下のように、政府職員の利用に必要なセキュリティ対策を講じ、機能を一部制限した利用環境としている(令和8年10月現在)。
 - 外部へのデータ送信(POST/PUT 等)はシステム上遮断しており、利用者の意図の有無にかかわらず、外部へのデータ書き出しを防止する仕組みを設けている。
 - Web 検索に必要な GET リクエストは許可するが、通信先、通信量及び通信日時をログとして記録し、事後に確認できる設計としている。
 - 利用者によるカスタマイズは、利用者権限で変更することが許可された範囲に限定されている。
 - セキュリティフックの状態にかかわらず、ネットワーク層に設けられた防御措置を回避できない構成となっている。
 - システム上の操作はログに記録され、事後に追跡・確認できる仕組みとなっている。
 - 共通設定は、サーバーの起動時に最新版が自動的に適用される。共通設定に手動の変更が加えられた場合でも、次のサーバー起動時に所定の設定で上書きされる。
 - 源内工房への入出力テキスト、外部通信その他の操作情報は、改ざん防止措置が講じられた形式でログに記録・保存される。
 - セキュリティ上の問題、規程違反又は不正利用のおそれが確認された場合は、源内開発者が利用権限を停止し、当該利用者のアクセスを遮断できる仕組みとなっている。

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

- 本基準において、「源内開発者」とは、デジタル庁において源内の企画・開発・提供・運用に携わる職員をいう。デジタル庁がこれらの業務の全部又は一部を外部事業者等に委託する場合は、当該委託業務に実際に従事する当該外部事業者等及びその担当者を含む。
- 源内開発者は、源内を利用者が安定的かつ継続的に利用できるよう、源内本体、エージェントチャット、AI アプリ、源内工房その他の関連機能について、開発、改修、保守及び運用を行う。
- 源内開発者は、利用者及び管理者が源内を用いて業務を円滑に遂行できるよう支援する。また、利用状況、業務上の必要性及び技術的な実現可能性等を踏まえ、必要と判断した機能の追加又は改修を行う。
- 源内開発者は、源内のシステムや AI アプリを開発するだけでなく、公開可能なソースコード(プログラムを構成する記述)、開発上の知見及びノウハウを、OSS として公開するとともに、関連情報を公開する役割を担う。

4-2 開発及び運用に用いる環境

- 源内では、開発、検証及び本番運用を安全かつ安定的に行うため、用途に応じて環境を適切に分離するとともに、開発環境であっても本番相当の機密性を有するデータを扱う可能性を踏まえ、必要なセキュリティ水準を確保する。また、各環境に対応する Web、AI アプリ、データ等の接続関係を一貫して管理し、環境間の誤接続や未検証機能の本番影響を防止する。
- このため、源内の開発及び運用に用いるガバメントクラウドの環境は、原則として次の四つに区分する。
 - ① 開発環境
 - ② ステージング環境
 - ③ 本番環境
 - ④ 共通環境
- 開発環境においても、本番環境に近い条件で十分な試験・検証を行うため、本番環境で取り扱うデータと同等の機密性又は構造を有する試験データを使用する場合がある。このため、開発環境を含む各環境についても、本番環境に準じたセキュリティ管理が必要であり、ガバメントクラウド上の分類ではいずれも「本番」として取り扱う点に留意すること。
- 原則として、開発環境の源内 Web には開発環境の AI アプリを登録し、本番環境の源内 Web には本番環境の AI アプリを登録すること。開発環境と本番環境の誤接続や構成の不整合を防止するため、このような環境間の対応関係(システムを構成する要素の接続関係)を維持しなければならない。
- ガバメントクラウドへのアクセス制御及び共通的なセキュリティ対策には、可能な限り GCAS(ガバメントクラウドの利用及び管理に用いる仕組み)の機能を利用すること。この点に留意し、源内側で同様の機能を重複して開発することを避けること。

（開発環境）

- 開発環境は、源内の利用者に影響を与えることなく、源内開発者が新たな技術、機能、AI アプリ又はシステム構成等を試行し、その実現可能性、挙動及び性能を確認するための環境とすること。
- 新たな技術を採用し、又は既存システムを変更する場合は、原則として、まず開発環境において開発及びデプロイを行い、想定した挙動、性能及びシステムへの影響等を確認する。生成 AI、特に AI エージェントは、事前の設計のみでは挙動を十分に予測することが難しいため、実際の試作及び試験を通じて意図した挙動及び性能が得られるかを確認すること。
- 試験の目的及び内容に応じて、本番相当のデータ又は本番データを使用する場合がある。また、先端的な AI 技術については、仕様や公式文書等が十分に整備される前に検証が必要となる場合もあることから、利用者から隔離された開発環境において、安全性を確保しつつ必要な試行錯誤を行えるようにすること。
- 開発環境の源内 Web は、開発中の AI アプリ等について、ユーザビリティの確認や限定的なユーザーテストを行うためにも利用すること。このため、IP アドレス制限、利用登録時のメールアドレスのドメイン制限その他必要なアクセス制御を行うこと。
- 開発環境は、検証目的に応じて、デジタル庁の本番環境を再現した環境、他府省庁の利用環境を再現した環境その他複数の環境を設けることができる。

（ステージング環境）

- ステージング環境は、開発環境で検証を終えたプログラム、AI アプリその他の成果物について、本番環境へのデプロイ前に、本番環境と同等又はこれに近い構成で最終的な確認を行うための環境とすること。具体的には、安全にデプロイできること、特定の環境に依存した設定や意図しない構成が含まれていないこと等を確認すること。
- ステージング環境は、利用者に影響を与えない点では開発環境と共通するが、源内開発者が継続的な試行錯誤を行うことを主な目的とせず、本番環境への移行可否を判断するための検証を行うことを主な目的とすること。
- 脆弱性診断（システムに攻撃につながる弱点がないかを確認する検査）その他、本番環境と同等の構成で実施する必要があるかつ、利用者への影響を避ける必要がある検査については、ステージング環境を活用すること。

（本番環境）

- 本番環境は、利用者が源内を利用するための環境であり、あわせて、源内開発者が運用上必要な作業を行う環境でもある。
- 本番環境では、利用者への影響を最小化し、安定したサービス提供及び運用を最優先としなければならない。このため、特別な事情がある場合を除き、開発環境のような試行錯誤は行わず、クラウドリソースの管理、障害・エラーへの対応、設定変更その他の安定運用に必要な作業に限定すること。
- 政府機関向けの源内では、クラウド上の生成 AI サービスの利用可能容量その他の制約を踏まえ、複数のクラウド環境に源内本体及び AI アプリを展開すること。このため、各環境への自動デプロイ、ログの集約その他、デジタル庁がホストする源内を一体的かつ安定的に運用するための共通的な仕組み

及び本番環境を維持・発展させること。

- 源内の本番環境では、「政府統一基準群」で求められる主体認証及び多要素認証の要件を満たすため、外部の認証基盤との連携（GSS の Entra ID 等）を基本とし、その対応が難しい組織向けには、IP 制限及び登録時のメールアドレスのドメイン制限を採用した源内を展開すること。
- ログの取得・保存・利用、利用規約その他利用者に関する運用ルールについては、原則として本番環境における利用を対象とすること。
- 本番環境では、監査のためのログ取得、障害対応その他の必要な非日常業務を除き、源内開発者が利用者のデータや稼働中の AI アプリ等に直接アクセスしない運用を基本とし、ゼロタッチ運用（自動デプロイ・自動運用による人為的アクセスの原則排除）を徹底しなければならない。

（共通環境）

- 共通環境は、開発環境及び本番環境の双方で利用する必要がある一方、費用対効果又は技術的な制約から、それぞれの環境に同一のリソースを個別に構築・維持することが困難な場合に設ける環境とすること。具体的には、利用に高額な契約又は個別の利用条件を伴う生成 AI サービス等のリソースや、構築に多大な手間を要し、同一状態を再現することが困難な大規模データセット、検索基盤その他の共通的な資源が該当すること。
- 共通環境を利用する場合には、開発環境と本番環境の利用が相互に影響しないよう、アクセス権限、利用範囲、データの取扱い及び変更管理等を適切に管理しなければならない。

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）

- 源内開発者に付与する権限及び責任は、各源内開発者の能力並びに組織上のガバナンス（組織として適切に管理・統制する仕組み）の観点から、必要最小限となるよう適切に分離・設定すること。
- 権限及び責任の分離は、「開発担当」「運用担当」といった固定的な役割や職位のみを基準とするものではなく、源内の企画・開発・提供・運用に必要な業務内容、リスク及び責任範囲に応じて柔軟に設計しなければならない。これにより、源内の価値向上と安定運用を両立するとともに、過大な権限の集中を防止すること。
- 入出力内容を含むログへの人的なアクセスについても、通常時は可能な限り行わないことを基本としなければならない。業務上やむを得ずアクセスする場合には、その必要性及び目的を確認した上で事前承認を行い、アクセス対象及び期間を必要最小限に限定するとともに、権限付与及びアクセスの記録を保存しなければならない。

（本番環境の変更承認）

- 本番環境の源内本体や AI アプリに対する変更は、不特定多数の源内利用者や管理者に影響を及ぼす可能性があるため、変更内容、影響範囲及び必要性を確認した上で、当該変更について決定権限及び責任を有する者（源内班の参事官等）の事前承認を得なければならない。その承認の証跡を記録する運用とすること。

(デプロイ権限及び自動化)

- デプロイ作業とは、成果物をガバメントクラウド上で動かすための一連の作業を指す。デプロイは、可能な限り、継続的デリバリーを実現する機構(GitHub Actions 等)を利用して自動化し、その機構へのアクセスや利用できる源内開発者を制限すること。
- 本番環境及び共通環境へのデプロイは、利用者や他のシステムへの影響が大きいことから、原則として事前のレビュー及び承認を必要とすること。一方、開発環境については、迅速な試行錯誤を可能とするため、リスク及び影響範囲を踏まえ、原則として個別のレビュー及び承認を要しない運用とすること。

(変更内容のレビュー)

- 本番環境へのデプロイは、権限者による承認に加え、技術的観点から変更内容の妥当性を確認するため、他の源内開発者によるレビューを経ること。
- レビューには一定水準以上の技術的知識及び経験が必要であることから、レビューを行う権限は、必要な能力を有する源内開発者に限定すること。
- 本番環境へのデプロイにより利用者又は管理者に影響が生じた場合には、変更を実施した源内開発者及びレビューを行った源内開発者が連携し、影響範囲の確認、必要に応じた切戻し、原因分析及び再発防止その他の復旧対応を行うこと。

(データセットの変更管理)

- データ変更権限とは、源内の AI アプリのうち、行政実務用 RAG(行政資料等を検索し、その内容を参照して回答を生成する仕組み)が参照するデータセットについて、追加、更新、削除その他の変更を行う権限をいう。
- データ変更権限は、デプロイ権限及びレビュー権限とは区別して管理し、業務上の必要性及び担当範囲に応じて付与すること。このため、源内班以外の職員(例えば、SEABIS 担当者)が、所管する行政資料等に係るデータ変更権限を保有する場合がある。
- データを変更した者は、変更内容が行政実務用 RAG に正しく反映され、期待した挙動を示すことを確認し、その結果をレビュー担当者に報告する責任を負うこと。

(GCAS 管理権限の付与及び管理)

- GCAS 責任者は、ガバメントクラウドの利用に係る GCAS 上の権限区分であり、源内固有の役割ではない。GCAS 責任者は、源内が利用するクラウド環境へのアクセス権限の付与、変更及び削除等を管理できる強い権限を有することから、源内では、業務上の必要性を十分に確認した上で、当該権限を必要とする最小限の源内開発者に限定して付与すること。

4-4 開発・運用上の行動原則

本節では、源内の開発及び運用に当たって、源内開発者が従うべき行動原則を定める。ここでは、「デジタル社会推進標準ガイドライン」に定める事項を踏まえつつ、源内の開発及び運用に固有の行動原則を定める。

(確認すべきガイドライン群)

- デジタル社会推進標準ガイドライン、特に 100 番台(システム開発全般)、200 番台(情報セキュリティ関連)、300 番台(クラウド関連)、600 番台(サービスデザイン関連)及び DS-920(AI 調達・利活用ガイドライン)には、源内の開発に直接関係する記述が多く、Normative と位置付けられている文書には遵守すべきルールが定められている。源内開発者はこれらのガイドラインを必ず確認し、日々の開発において遵守しなければならない。

(合理的に説明できる技術選定)

- 源内開発者は、技術を選定する際、コストパフォーマンスなどのビジネス上の観点や、開発・運用のしやすさ・セキュリティなどのシステムの観点を踏まえ、第三者に対しても合理的に説明できる根拠に基づいて判断を行うこと。

(開発及び運用における AI の活用)

- 源内についても、政府情報システムとして、運用経費の継続的な削減及び費用対効果の向上に努める。大規模な AI 利用環境を可能な限り少ない人的資源で安定的に運用できるよう、運用の自動化、省力化及び効率化を前提としたシステム設計・開発を行うこと。
- また、源内開発者は、開発、テスト、レビュー、障害調査、ログ分析、情報収集その他の企画・開発・提供・運用の各工程において AI を徹底活用し、自らの業務プロセスについても継続的な自動化及び効率化を図ること。これにより、限られた人員及び予算の下で、開発速度、品質及び運用品質を高め、最大限の成果を得ることを目指すこと。

(国内外の最新動向の継続的把握)

- 生成 AI に関する技術、サービス、開発手法及び活用ノウハウは、国内外で急速かつ継続的に進展している。このため、源内開発者は、源内が我が国の行政における先進的かつ信頼性の高い AI 利用環境であり続けるよう、国内外の最新動向及び先行事例を継続的に把握すること。
- また、新たな技術、サービス又は手法について、その有効性、安全性、行政実務への適合性及び費用対効果等を検証し、有益と認められるものについては、必要な評価及び統制を行った上で、源内への導入及び活用を積極的に検討すること。

(不確実性の早期低減)

- 源内を取り巻く技術及び利用環境は変化が大きく、将来の不確実性が高い。このため源内開発者は、開発及び運用に伴う不確実性を可能な限り早期に把握し、検証を通じて低減すること。
- 具体的には、想定する利用者に早期にアイデアやプロトタイプを利用してもらい、実際の利用を通じて有用性、実現可能性、使いやすさ及び課題を確認すること。また、利用者にとって高い価値が見込まれる機能を優先し、小さく開発・検証した上で、得られた知見を次の改善に反映すること。
- このような取組では、DevOps(開発と運用を連携させ、継続的に改善する考え方)で用いられる一般的な手法やノウハウを活用し、短いサイクルで検証と改善を繰り返すこと。

(ムリ・ムダ・ムラを抑えた開発・運用)

- 源内開発者は、限られた人員、時間及び予算の下で源内の価値を最大化するため、開発及び運用における過剰な作業、重複、手待ち、属人化、作業量の偏り等を継続的に見直し、ムリ・ムダ・ムラの削減に努めること。あわせて、自動化、標準化及びAIの活用等により、より少ない負担で高い成果を継続的に得られる開発・運用プロセスを構築すること。
- 源内は、AIの活用を通じて行政業務の効率化及び職員の就労環境の改善に寄与することを目的の一つとする。このため、源内開発者自身についても、過度な長時間労働や特定の者への業務集中を前提とした開発・運用を避け、業務の平準化、知識の共有、自動化等を進めることにより、持続可能な開発・運用体制を確保すること。
- ムダを抑制するため、利用者認証にはGSSの仕組みを利用し、クラウド基盤へのアクセス及び管理にはGCASを利用するなど、既存の共通機能を優先的に活用すること。同じ機能を源内で重複して開発する、いわゆる「車輪の再発明」は避けること。
- ムラ(作業品質のばらつき)を抑制するため、繰り返し行う作業をコード化し、コーディングAIエージェント(プログラム作成等を支援するAI)のスキルとして定型化し、又はテストの観点を言語化して生成AIに実施させることが有効である。継続的デリバリーの機構(例:GitHub Actions)にセキュリティ診断の工程を組み込むことにより、成果物の品質を安定させること。
- 複数回発生し、作業ミスが生じやすい定型作業は、可能な限り自動化すること。例えば、各組織の源内本番環境にAIアプリを個別に登録する作業等がこれに当たること。

4-5 源内の開発・運用におけるセキュリティ等の統制

- 源内開発者は、開発・運用全般において、以下のセキュリティ統制を維持しなければならない。
なお、各セキュリティ統制に関しては、源内開発者が別途定めるセキュリティ基準に沿ってセキュリティを確保するものとする。

(1) 脆弱性への対処

源内を構成するソフトウェアに既知の脆弱性(攻撃に悪用される可能性のある弱点)が確認された場合は、その重要度に応じた期限内に対処しなければならない。重要度の高い脆弱性が残った状態で、本番環境へリリースしてはならない。

(2) セキュリティ事象への即応

セキュリティ監視で検知した事象のうち、重要度が高いものには、可能な限り速やかに対処しなければならない。直ちに解決できない場合は、影響範囲の拡大を防ぐ暫定措置を講じた上で、対応方針を決定しなければならない。

(3) 変更の追跡可能性

本番環境に対するすべての変更は、誰が、いつ、何を変えたかを事後的に追跡可能な形で記録しなければならない。手動操作による変更は原則として行わず、コードとして管理された変更に対して承認プロセスを経て適用すること。

(4) 監査証跡の不可変性

セキュリティ上重要なログ(操作履歴、認証記録、通信記録等)は、源内開発者を含むいかなる主体によっても改ざん・削除が行えない構造としなければならない。この不可変性は、技術的手段により強制し、運用規律のみに依存しない。

(5) 権限の最小化と分離

各源内開発者に付与する権限は、業務上の必要性が認められた最小限の範囲にとどめる。日常の運用操作と、インフラ構成を変更する操作は、異なる権限として分離し、より強い権限の行使には他の源内開発者への通知又は承認を伴わせること。付与済み権限は定期的に棚卸しし、不要となった権限を速やかに削除すること。

(6) 秘密情報の管理

認証情報、API キー、暗号鍵等の秘密情報をソースコードやリポジトリに含めてはならない。秘密情報の漏えいを自動的に検知する仕組みを維持し、漏えいが確認された場合は、対象となる秘密情報を速やかに無効化し、又はローテーション(新しい情報に更新)を行うこと。

(7) 本番環境の保護

本番環境への変更は、他の源内開発者によるレビュー及び参事官等の権限者の承認を経なければならない。本番環境への直接的なアクセスは、障害対応等の非日常的な業務に限定すること。

源内開発者は、日々の開発においてこれらの統制を維持する責任を負い、統制を毀損する変更を行ってはならない。統制の変更が必要な場合は、承認権者の承認を得ること。

(8) バックアップ及び復旧

源内開発者は、障害、誤操作、サイバー攻撃その他の事象によりデータ又はシステムが毀損した場合に備え、必要なバックアップを取得し、復旧手順を整備すること。バックアップから実際に復旧可能であることを定期的に確認し、復旧不能な状態を放置してはならない。

(9) 障害対応手順の整備

重大なセキュリティ事故又は障害に備え、連絡先、役割分担、判断権限、初動対応、利用停止、切戻し、復旧及び対外連絡等の手順をあらかじめ定めなければならない。障害発生時に個人の経験や判断のみに依存せず、組織的かつ迅速に対応できる状態を維持すること。

(10) 監視及び異常検知

源内の可用性、性能、エラー、認証、通信その他の重要な状態を継続的に監視し、通常と異なる挙動を早期に検知できる仕組みを維持すること。重大な異常を検知した場合は、自動通知その他の手段により、必要な担当者が速やかに把握できるようにすること。

(11) データの誤消去・誤変更の防止

本番環境のデータについて、誤消去、誤更新その他の人為的な操作による毀損を防止するため、重要な操作には必要に応じて権限制御、確認手順、承認又は復旧手段を設けること。特に大量のデータを変更又は削除する操作については慎重に取り扱うこと。

(12) セキュリティ更新の継続

OS、ライブラリ、コンテナイメージ、ミドルウェアその他の構成要素について、提供元のサポート状況及びセキュリティ更新を継続的に確認すること。サポートが終了した製品や更新されていない構成要素を漫然と本番環境で使用し続けてはならない。

(13) セキュリティ事故等の発生時のエスカレーション

源内開発者は、セキュリティ事故、重大な障害その他利用者又は源内の安定運用に重大な影響を及ぼす事象を把握した場合、自己判断で対応を完結させず、影響範囲及び緊急度を確認の上、速やかに、CAIO などの責任者その他あらかじめ定められた関係者へ報告し、必要な指示及び判断を仰がなければならない。

重大なセキュリティ事故又は障害が発生した場合は、発生日時、影響範囲、原因、対応内容及び復旧経過を記録すること。復旧後は必要に応じて原因分析を行い、技術、運用、体制その他の改善措置を実施し、同種事象の再発防止につなげること。

- セキュリティに関しては、「AI 事業者ガイドライン」及び AI 調達・利活用ガイドラインに加え、ISO/IEC27001 附属書 A、ISO/IEC42001、OWASP Top 10 等の国内外の主要文書も参照しつつ、源内開発者が別途定めるセキュリティ基準に沿ってセキュリティを確保すること。

—以上—

エージェントチャット及び源内工場の概要と比較

1. エージェントチャットの概要

源内が提供するエージェントチャットは、源内 Web 内のチャット画面から利用できる、高機能なチャット（対話型 AI）として位置付けられる AI エージェント利用環境であり、利用者が目的を伝えれば手順の組み立てから成果物の仕上げまでを自律的に進めるものである。

主な利用目的は、職員が日常業務の中で AI を手軽に利用し、質問や壁打ち、情報整理、データ分析、文書作成、資料生成、コード実行、スキルの利用・作成などを効率化することである。

利用シーンとしては、複数文書の読解と要約、資料の作成・整形、調査、定型事務の反復処理等のほか、業務上の疑問への回答、文章や報告資料の作成、CSV 等の集計・分析、PowerPoint・PDF・Excel の生成、既存スキルを用いた定型業務の実行などが想定される。

特徴は、源内 Web から即座に利用できる操作の簡便さ、複数の基盤モデルを選択できること、マイドライブ上でファイルを保存・管理できることである。ファイルはセッション終了後も保持される。外部情報の取得は Web 検索に限定され、利用者が任意の URL を直接読み取る機能は持たない。

利点は、特別な開発環境の操作を必要とせず、多くの日常業務を一つのチャット画面から実行できる点にある。また、利用量に応じた従量課金制であり、個人専用マシン型の源内工場と比べて、一人当たりの運用コストを抑えやすい構成である。通信先を許可されたサービスに限定し、アップロードファイルのマルウェアスキャンや認証・アクセス制御を行うなど、安全性を通信経路の限定によって確保している。

一方、デメリットは、外部情報の取得が Web 検索に限られ、任意サイトを直接読み取りながら多段階で深く調査できる Deep Research (AI が Web 上の複数の情報を自律的に収集・分析し、詳細なレポートを作成する機能) には一部対応していない点である (ウェブ検索索引を使用した調査とレポート生成は利用可能)。

また、コード実行は隔離された環境で行われるため、通常の PC と同様にサーバーを起動し、Web アプリを構築してブラウザで動作確認するようなプロトタイプ開発には向かない。

したがって、日常的な分析・文書作成・スキル実行には適するが、網羅的な調査や動くシステムの試作が必要な場合には、源内工場を利用する必要がある。

2. 源内工房の概要

源内が提供する源内工房は、利用者ごとに割り当てられる個人専用の AI エージェントの開発環境であり、主な利用目的は、

①Web 検索だけでは不十分な網羅的・多段階な情報収集の実施

②実際に動作するプロトタイプの開発を政府職員自身が実施

である。利用シーンとしては、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、Web アプリの構築、サーバー起動、ブラウザでの動作確認、データ分析、文書・スライド・Excel 等の生成、コードやスキルの作成・実行などが想定される。

特徴は、SageMaker Studio と Claude Code を基盤とし、JupyterLab(非営利団体が開発したオープンソースのプログラム実行・開発環境のツール)、専用 WebUI、CLI(Command Line Interface)を通じて、bash、Python、Node 等を実行できる点である。個人専用のファイルシステムにソフトウェア一式や作業ファイルを保持でき、環境が停止してもファイルは保持される。

また、任意サイトへの直接アクセスが可能であり、AI 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することを前提としている。このため、利用可能なモデルには、高精度の Tool Use、長いコンテキストでの安定性、複雑な指示への追従性が求められる。

利点は、政府職員が行う調査、企画、試作を一つの環境で連続的に進められることである。情報システムや Web デザインなどの動くプロトタイプを早期に作成し、利用者に見せて検証した上で、プロトタイプを仕様書の一部として活用することで調達作業につなげることができる。

紙面の仕様書だけを作成して発注する場合と比べ、実現可能性や利用者ニーズを事前に確認しやすくなる。また、外部への書き込みをブロックし、危険コマンドの事前制御、個人情報検出、通信内容の検査、異常時の環境停止等を組み合わせた多層的な対策が採られている。

一方、デメリットは、通常のチャットより操作が複雑で、専用 URL、JupyterLab、WebUI 等を利用する必要がある点である。自動バックアップは提供されず、必要な成果物は利用者自身がダウンロードして保存する必要がある。利用コストも、エージェントチャットと比較して高額になる可能性がある。このため、日常的な質問や文書作成だけの作業であれば、エージェントチャットの方が簡便かつ効率的である。

なお、現時点では利用可能なモデルが限定的であり、将来は他モデルに対応していくことを検討中である。

3. 両者の比較

表 「エージェントチャット」及び「源内工房」の比較表①

大分類	比較観点	エージェントチャット	源内工房	差異・使い分け
位置付け	一言で言うと	高機能なチャット	個人専用の開発マシン	前者は手軽な対話・業務実行、後者は開発・調査向けの専用環境
	基本的な役割	質問、壁打ち、分析、文書作成、スキル実行等	Deep Research、プロトタイプ開発を含む高度な作業	大半の業務は両方で可能だが、源内工房は調査と開発に強み
	選択基準	迷った場合に選ぶ環境	Web を広く調べる、または動くシステムを作る場合に選ぶ環境	Web 情報を幅広く直接読み取ったり、プロトタイプを開発したりする場合は源内工房
調査	Web 検索	利用可能	利用可能。	共通機能
	検索キーワードの取り扱い	ガバメントクラウドに構築済みの Web 検索索引を利用するためガバメントクラウド内で完結	ガバメントクラウド外に露出	機微な検索キーワードはエージェントチャットの利用を検討
	任意 URL の直接取得	不可。外部情報取得は Web 検索のみ	任意サイトへ直接アクセス可能(※リスクがあるとシステム側が判断するものについては遮断等のブロック機能あり)	源内工房のみ、Web 検索で届かないサイトを含む情報収集が可能
	Deep Research	簡易な Deep Research のみ可能(ウェブ検索索引を使用した調査とレポート生成は利用可能)	インターネット全体の任意サイトに多数アクセスし、多段階で情報を収集・比較・統合が可能	網羅的・多段階な調査が必要な場合は源内工房
プロトタイプ開発		Web アプリの開発用途には向かない(隔離された実行環境のため) Excel のマクロや簡単な Python スクリプト等であれば十分に作成可能	Web アプリ構築、サーバー起動、ブラウザ動作確認が可能	動くシステムを企画・設計・試作・動作検証したい場合は源内工房
操作性	利用画面	源内 Web 内のチャット UI	JupyterLab＋専用 WebUI/CLI	前者は統合された簡易操作、後者は開発環境型
	起動の手軽さ	源内 Web で即利用	専用 URL→JupyterLab→WebUI	手軽さはエージェントチャット

大分類	比較観点	エージェントチャット	源内工房	差異・使い分け
実行環境	実行基盤	AWS 管理コンテナ。呼ぶたびに起動	個人専用仮想マシン	サーバレス型と専用環境型の違い
	コード実行	実行基盤とは隔離された実行環境で Python、TypeScript 等を実行	個人専用仮想マシンで Bash、Python、TypeScript 等を直接実行	自由度と安全制約の差
ファイル	保存場所・用途	デジタル庁が調達するガバメントクラウド内のストレージ。ドキュメント保管が主体	デジタル庁が調達するガバメントクラウド内のストレージ。専用インターフェースのソースコード等を管理	文書保管中心と開発作業領域の違い
	自動バックアップ	あり。定期的なスナップショット(ただし、エージェントチャットではマイドライブ(ストレージ)を一時的な作業領域として提供しているため、永続的な保存は保証しない。)	なし。利用者がダウンロードし、ファイルを別途保存することが必要	—
セキュリティ	外部とのネットワーク制御	許可リスト方式(AI 推論サービス等、システムが必要とする宛先のみ通信許可)	送信ブロック方式。GET は全サイト可、POST や PUT は全サイト不可	行き先を絞る方式と、書き込みを止める方式
	プロンプトインジェクション対策	promptAttack スコアリング、警告注入、監視(ただし、完全な機能ではない。)	PreToolUse Hook による危険コマンド事前ブロック、書き込み不可(ただし、完全な機能ではない。)	検出・監視中心と、実行前・通信構造による抑止
	個人情報・認証情報検出	なし。通信先限定で担保	ガバメントクラウド内のガードレール製品を使い、入力テキストを検査し、個人情報や認証情報(API キー等)等が疑われる場合は送信前に利用者に確認(ただし、完全な機能ではない。)	源内工房は開発用途で利用されるため本番データの混入や API キーのアップロードが含まれる可能性が相対的に高い。そのため、完全な機能では無いものの安全性を高めるため、検知システムを導入。なお源内

大分類	比較観点	エージェントチャット	源内工房	差異・使い分け
				工房の中でのクラウドリソースの利用(生成 AI サービス、音声認識、ストレージ等)はクラウドのロール機能を使うため、API キー等の認証情報は源内工房の中に持たない
共通	スキル共有	スキル作成が可能、共有されたスキルを使用可能	同左	同じスキルを両環境で共用可能
	基本成果物	データ分析、文書、PowerPoint、PDF、Excel 等を生成可能	同左	基本的な業務成果物生成は共通

(令和8年 10 月時点)

表 「エージェントチャット」及び「源内工房」の比較表②（セキュリティ対策含む）

新たに源内に導入予定の「AIエージェント」利用環境（2種類）

	エージェントチャット	源内工房
利用者	AIエージェント利用上の注意事項に同意する一般の政府職員向け	AIエージェントに係るリスク等に対するリテラシーを持つ政府職員限定
基盤とした技術	Bedrock AgentCore Runtime	SageMaker Studio + Claude Code
AIエージェントの実行基盤	サーバーレス環境	個人専用の仮想マシン
AIエージェントの動作形態	Pythonプログラムとして動作	Claude Code（対話型CLI）として動作 又は、Claude Codeを実行する源内工房内の専用Webアプリ（源内工房WebUI）のPythonプログラムとして動作
AIエージェントのコード実行	隔離された実行環境（安全だが制約あり）	個人専用の仮想マシンで直接実行（Bash/Python/TypeScript）
セッション長（最大）	最大8時間	12時間無操作で停止、再起動で継続
トークンのコスト制御	利用料金を1時間ごとに集計し、前日から当日までの累計額で判定（上限の75%でお知らせを表示、上限到達で新規の質問不可。従来の源内チャットは引き続き利用可）	月額予算制 + 1時間100回まで、管理者側で監視
操作画面	源内Web内のチャットUI	JupyterLab + 専用WebUI/CLI
セッションの永続性	セッション終了後も履歴として保持（364日保管）	環境停止後も永続保持
利用者向けのストレージ	デジタル庁が調達するガバメントクラウド内のストレージ。保管量に上限あり（目安：数万ファイル以上、1ファイル数百MBまで）	デジタル庁が調達するガバメントクラウド内のストレージ
バックアップ	自動（定期的にスナップショット） ただし、一時的な作業領域のため、永続的な保存は保証しない	なし（利用者が自分でダウンロード保存）
セキュリティ（ネットワーク制御）	許可リスト方式（AI推論サービス等、システムが必要とする宛先のみ通信許可）	送信ブロック方式（GETはブロックリスト方式で危険性の高いもののみ不可、POSTやPUTは全サイト不可）
セキュリティ（外部情報の取得）	AWS管理のweb検索のみ（任意URLの直接取得は不可） 検索キーワードを含む処理はガバメントクラウド内で完結	直接アクセス。ブロックリスト方式で危険性の高いもののみ不可
セキュリティ（プロンプトインジェクション対策）	promptAttackスコアリング（検出時に警告注入+監視）	PreToolUse HookでAIの危険コマンドを事前ブロック
セキュリティ（個人情報検出）	なし（通信先の限定で担保）	ガバメントクラウド内のガードレール製品を使い、入力テキストを検査し、個人情報疑われる場合は送信前に利用者に確認
セキュリティ（ユーザー分離）	ログイン認証+属性ベースのアクセス制御	OSレベルのユーザー分離（利用者ごとに独立領域）
セキュリティ（ファイル分離）	利用者IDで保存先を分離	OSのユーザー権限で分離
セキュリティ（マルウェア対策）	アップロードファイルの自動スキャン	同左（アップロードファイルの自動スキャン）
セキュリティ（異常検知）	認証失敗・権限拒否の閾値超過で警報	自動検知+利用者環境の自動停止
セキュリティ（通信検査）	暗号化（標準）	暗号を一度復号して中身を検査してから再暗号化

源内において提供中の AI アプリ一覧
(2026 年 8 月時点におけるデジタル庁内提供アプリ(順不同))

- チャット(着想や整理のための壁打ち)
- 国会答弁を調査・分析(国会質問に関連する過去の国会答弁からレポート作成(エージェント型))
- 法制度に関する調査(法制度(法令)に関する情報や知識を検索(通称:Lawsy))
- SEABIS 質問応答-RAG(デジタル庁)(出張経費精算や謝金諸手当などに関する疑問を解決)
- パブコメ分類 AI(パブリック・コメントを AI で自動分類・ラベル付け)
- 補助金制度調査(jGrants)(補助金情報を検索し、条件に合った制度を分析してレポートを自動生成(エージェント型))
- 国会審議中継の速記データ検索(国会審議中継から機械的に文字起こししたものを検索)
- 画像の ALT(代替テキスト)作成(画像の内容を説明する ALT(代替テキスト)を作成(エージェント型))
- 手元の表データを分析・作図(簡易版)(表形式のデータから簡単な分析と作図を実施)
- 目標設定チェッカー(人事評価の目標設定をブラッシュアップ)
- アナログ規制点検ツール(アナログ規制に関係しうる条項の点検対象の確認)
- Excel 関数の提案(Excel 計算のための数式・関数作成)
- Excel ファイルのテキストを 1 行ずつ処理(最大で 10 万行の Excel ファイルのテキストを生成 AI で処理)
- 文体を指定して要約(箇条書きなど、指定の文体でテキストを要約)
- EASY 質問応答-RAG(デジタル庁)(マニュアル・FAQ から、「行政文書の電子的管理システム(EASY)」に関する疑問を解決)
- IT サポート質問応答(デジタル庁)(GSS 環境の利用等に関する照会を庁内情シス FAQ から回答)
- 文章を生成(手元の情報をもとに文章を作成)
- 翻訳(手元の文章を他の言語に翻訳)
- 画像を生成(文章や単語から画像を生成)
- ダイアグラムを生成(テキストからフローチャートやマインドマップを作成)
- 音声ファイルから文字起こし(音声ファイルを元に文字起こし)
- ウェブ情報や手元資料を要約(ウェブ検索や指定したリンク先、手元資料を Gemini でまとめ(通称:AI+ウェブ検索(Gemini)))
- ファイルの内容を徹底的に調査・分析(複数ファイルを読み込み、内省を繰り返しながら品質確認・修正を行い最終報告書を作成(エージェント型))
- 物品管理システム質問応答-RAG(デジタル庁)(マニュアル・FAQ をもとに、SEABIS 物品管理システムに関する疑問を解決)
- 不開示情報を判定(情報公開請求対象の文書から不開示情報の該当箇所を抽出(エージェント型))
- 行政文書の用例を調査・検索(特定の単語や表現が行政文書(国会答弁・法令・答申)での使われ方を調査(エージェント型))
- サービス名称のガイドライン適合性を確認(デジタル庁専用)(サービス・プロダクト名称のガイドライン適

合を確認(エージェント型))

- 文章を校正・添削(ファイル添付版)(Office ファイルや PDF ファイル内の文章を校正(エージェント型))
- 政策・法令の過去経緯を調査し根拠付きレポート作成(国会答弁・法令改正・答申から経緯を調査し、出典と確信度を明記したレポートを作成(エージェント型))
- 大量のテキストの自動分類(数百から数万件の自然文を、かんたんに集計や集約)
- ネ申エクセルを最低限の機械処理が可能な状態に変換(ネ申エクセル(結合セル・複雑なヘッダー)を構造化された表形式に変換)
- 会議記録を作成(Teams の会話データから会議の記録を作成)
- 行政手続きの電子化状況(R7)を調査・分析(R7 行政手続き悉皆調査(約 75,000 件)のオンライン化状況を分析(エージェント型))
- 人事評価制度質問応答-RAG(デジタル庁)(目標設定・評価に関する疑問を解決)
- 手元の表データを調査・分析(アンケート等の表形式のデータの集計・分析(エージェント型))
- 公用文を校正(文化庁ガイドラインに準拠した公用文の作成支援)
- 日⇄英翻訳(日本語の文章を英語に、または英語の文章を日本語に翻訳(通称:PLaMo 翻訳))
- マークダウン形式の表に変換(表のテキストデータをマークダウンのテーブルに整形)

「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」

(令和7年 12 月 19 日人工知能戦略本部決定)に基づきガバメント AI 源内として取り組むべき事項

源内の企画・開発・提供・運用に当たっては、「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、信頼できる AI の実現に向けた取組を推進する。

具体的には、人間中心、公平性、安全性、透明性、アカウントビリティ、セキュリティ、プライバシー・個人情報保護等を基本として、利用分野、目的及び影響度に応じたリスクの特定・評価・対策を行う。

1 適正性確保に必要となる主要要素

指針の事項	事項の概要	源内基準の対応
人間中心	人間の尊厳・基本的人権及び法令を尊重し、AI を活用する範囲や条件について人間が最終判断を行う。	○
公平性	AI によって不当な偏見や差別を生じさせ、又は助長しない。	○
安全性	AI の活用によって生命、身体、財産、自由、名誉等に危害を及ぼさないようにする。	○
透明性	必要かつ技術的に可能な範囲で情報を開示し、事後的な検証可能性を確保する。	○
アカウントビリティ	責任の所在を明確にし、責任を果たす仕組みを構築する。	○
セキュリティ	不正操作による意図しない動作や停止等のリスクを低減し、AI のセキュリティを確保する。	○
プライバシー・個人情報	データの重要性に応じてプライバシーを保護し、個人情報保護法等を遵守する。	○
公正競争	AI 資源の集中による不公正な取引等を防ぎ、公正な競争の促進に貢献する。	—
AI リテラシー	AI の便益とリスクを理解するための知識・能力及び倫理観を身に付ける。	○
イノベーション	持続可能性を確保しつつ、社会課題の解決に資する AI の開発・活用を促進する。	○

2 適正性確保のための基本方針

指針の事項	事項の概要	源内基準の対応
リスクベースでのアプローチ	AI のリスクを特定・評価し、利用分野、目的及び影響度に応じた対策を講じる。	○
ステークホルダーの積極的な関与	AI の便益又はリスクの影響を受ける主体が、ガバナンスに関与し、協働して課題を解決する。	○

指針の事項	事項の概要	源内基準の対応
一気通貫での AI ガバナンス	研究開発から社会実装までの各段階を一体として管理し、便益の最大化とリスク管理を図る。	○
アジャイルな対応	PDCA サイクルを通じ、変化する技術及びリスクに柔軟かつ迅速に対応する。	○

3 AI を開発・提供する主体として取り組む事項

指針の事項	事項の概要	源内基準の対応
AI ガバナンスによる俯瞰的な適正性の確保	ライフサイクル全体でリスクの特定・評価・対処を行う組織的プロセスを構築し、運用・継続改善する。	○
経営層の関与したモニタリング・評価	経営層の関与を含むモニタリング及び評価の仕組みを設ける。	○
情報の適切な開示	AI ガバナンス及びリスク管理に関する情報を適切に開示する。	○
教育・研修	AI の適正な開発・提供・利用に必要な教育及び研修を行う。	○
説明可能性の確保	学習データの出所や生成物について、知的財産・プライバシー等に配慮しつつ、合理的な範囲で説明可能性を確保する。	○
利用者への情報提供	AI の仕組み・限界、禁止事項、データ収集方針、出力の信頼性に関する注意等を利用者に提供する。	○
違法行為への悪用防止	サイバー攻撃、詐欺その他の犯罪・違法行為に利用されるリスクを評価し、対策を講じる。	○
不適切な出力等の抑制	ハルシネーション、偏見・差別、偽・誤情報、意図しない動作及び誤作動の抑制・改善に取り組む。	○
AI 生成物であることを判断できる技術	必要に応じ、電子透かし、来歴管理、API 等の技術の開発・実装に努める。	×
事業継続性の確保	障害時の損害を抑え、早期復旧するための事業継続計画をあらかじめ策定する。	○
質の高いデータの確保・適正活用	質の高いデータを確保し、適正に活用する。	○
データ保有者等とのコミュニケーション	データの適正な活用について、知的財産等の保有者その他の関係者と継続的に意思疎通を図る。	—
データ保有者等への利益還元等	社会的影響力の大きい AI の提供者は、利益還元のエコシステム等の検討・実施に努める。	—

4 国として特に取り組む事項

指針の事項	事項の概要	源内基準の対応
AI の積極的かつ先導的な活用	国自らが AI を積極的かつ先導的に活用し、活用事例や留意点を広く周知する。	○

指針の事項	事項の概要	源内基準の対応
公共調達における開発 実証機会	公共調達を通じた AI の開発・実証機会の提供を進める。	○
AI リテラシーの向上	職員を含む全ての主体が、倫理、法令、人権、安全等を理解し、責任ある利用者として行動できるようにする。	○
技術動向・利用実態の 把握	最新の技術動向及び活用実態を把握し、リスクと対応策を検討する。	○
教育・ガイダンスの推進	AI の基本的な使い方や注意点を学ぶコンテンツ及び技能習得支援を提供する。	○
AI ガバナンスの継続的 検討	国内外の動向及び技術進歩を踏まえ、ガバナンスやガイドラインを継続的に点検・見直しする。	○
責任の所在等の明確 化	AI 利用に伴う課題について、責任の所在等に関する論点・考え方を整理する。	○
相互運用性への配慮	国際的なガバナンスの必要性を踏まえ、相互運用性にも配慮する。	○
行政としてのアカウント ビリティ	行政の信頼性を確保するため、適切なリスク対策を行い、判断根拠が不明瞭にならないよう国民への説明責任を果たす。	○
AI ガバナンス責任者の 任命	各府省庁等において AI ガバナンスの責任者を任命する。	○

○:源内基準において対応済

×:未対応

－:対応の対象外

■改訂履歴

版数	発行日	改訂履歴
第 1 版	2026 年 10 月 20 日	初版発行